



HAL
open science

Charting the equine miRNA landscape: An integrated pipeline and browser for annotating, quantifying, and visualizing expression

Jonah N Cullen, Jakub Cieslak, Jessica L Petersen, Rebecca R Bellone, Carrie J Finno, Ted S Kalbfleisch, Kirstine Calloe, Stefano Capomaccio, Katia Cappelli, Stephen J Coleman, et al.

► **To cite this version:**

Jonah N Cullen, Jakub Cieslak, Jessica L Petersen, Rebecca R Bellone, Carrie J Finno, et al.. Charting the equine miRNA landscape: An integrated pipeline and browser for annotating, quantifying, and visualizing expression. PLoS Genetics, 2025, 21 (9), pp.e1011835. <10.1371/journal.pgen.1011835>. <hal-05343164>

HAL Id: hal-05343164

<https://hal.science/hal-05343164v1>

Submitted on 3 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

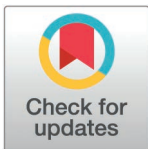


Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

RESEARCH ARTICLE

Charting the equine miRNA landscape: An integrated pipeline and browser for annotating, quantifying, and visualizing expression

Jonah N. Cullen^{1*}, Jakub Cieslak², Jessica L. Petersen³, Rebecca R. Bellone^{4,5}, Carrie J. Finno⁴, Ted S. Kalbfleisch⁶, Kirstine Calloe⁷, Stefano Capomaccio⁸, Katia Cappelli⁸, Stephen J. Coleman⁹, Ottmar Distl¹⁰, Sian A. Durward-Akhurst¹¹, Elena Giulotto¹², Natasha A. Hamilton¹³, Emmeline W. Hill¹⁴, Lisa M. Katz¹⁵, Dan A. Klaerke⁷, Gabriella Lindgren^{16,17}, David E. MacHugh^{14,18}, Mariusz Mackowski², James N. MacLeod¹⁹, Julia Metzger¹⁰, Barbara A. Murphy¹⁴, Ludovic Orlando²⁰, Terje Raudsepp²¹, Maurizio Silvestrelli⁸, Eric Strand²², Teruaki Tozaki²³, Dagmar S. Trachsel⁷, Laura S. Valderrama Figueroa¹, Brandon D. Velie²⁴, Claire M. Wade²⁵, Bianca Waud²⁶, James R. Mickelson²⁷, Molly E. McCue¹



1 Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota, United States of America, **2** Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Poznan, Poland, **3** Department of Animal Science, University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America, **4** Department of Population Health and Reproduction, School of Veterinary Medicine, University of California - Davis, Davis, California, United States of America, **5** Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California - Davis, Davis, California, United States of America, **6** Department of Veterinary Science, Martin-Gatton College of Agriculture, Food, and Environment, University of Kentucky, Lexington, Kentucky, United States of America, **7** Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg, Denmark, **8** Sport Horse Research Centre, Department of Veterinary Medicine, University of Perugia, Perugia, Italy, **9** Department of Animal Sciences, Colorado State University, Fort Collins, Colorado, United States of America, **10** Institute of Animal Genomics, University of Veterinary Medicine Hannover, Hannover, Germany, **11** Department of Veterinary Clinical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota, United States of America, **12** Department of Biology and Biotechnology, University of Pavia, Pavia, Italy, **13** Equine Genetics Research Centre, Racing Australia, Scone, New South Wales, Australia, **14** UCD School of Agriculture and Food Science, University College Dublin, Belfield, Dublin, Ireland, **15** UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin, Ireland, **16** Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden, **17** Department of Biosystems, Center for Animal Breeding and Genetics, KU Leuven, Leuven, Belgium, **18** UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin, Ireland, **19** Department of Veterinary Science, Gluck Equine Research Center, University of Kentucky, Lexington, Kentucky, United States of America, **20** Centre for Anthropobiology and Genomics of Toulouse, CNRS, Université de Toulouse, Toulouse, France, **21** College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, United States of America, **22** Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Aas, Norway, **23** Genetic Analysis Department, Laboratory of Racing Chemistry, Utsunomiya, Tochigi, Japan, **24** Equine Genetics & Genomics Group, School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia, **25** School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia, **26** Sydney School of Veterinary Science, University of Sydney, Sydney, New South Wales, Australia, **27** Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St Paul, Minnesota, United States of America

* cull0084@umn.edu

OPEN ACCESS

Citation: Cullen JN, Cieslak J, Petersen JL, Bellone RR, Finno CJ, Kalbfleisch TS, et al. (2025) Charting the equine miRNA landscape: An integrated pipeline and browser for annotating, quantifying, and visualizing expression. *PLoS Genet* 21(9): e1011835. <https://doi.org/10.1371/journal.pgen.1011835>

Editor: Androniki Psifidi, Royal Veterinary College, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: February 17, 2025

Accepted: August 6, 2025

Published: September 5, 2025

Copyright: © 2025 Cullen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All presented data are either already publicly available through NCBI's Sequence Read Archive (SRA), the FAANG data portal, or EMBL-EBI's European Nucleotide Archive. No unpublished results

Abstract

MicroRNAs (miRNAs) are essential regulators of gene expression, yet few comprehensive databases exist for miRNA expression in non-model species, limiting our ability to characterize their roles in gene regulation, development, and disease.

from other individuals are included. Primary: PRJNA1224819 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1224819?reviewer=on1n5qkg09j03k37uc8783480h3>) FAANG: <https://data.faang.org/dataset?sortTerm=accession&sortDirection=desc> PRJEB14485: <https://www.ebi.ac.uk/ena/browser/view/PRJEB14485> PRJEB20494: <https://www.ebi.ac.uk/ena/browser/view/PRJEB20494> PRJEB27174: <https://www.ebi.ac.uk/ena/browser/view/PRJEB27174> PRJNA573500: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA573500> PRJNA593513: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA593513> PRJNA637872: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA637872> PRJNA681404: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA681404> PRJNA705083: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA705083> PRJNA726388: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA726388> PRJNA306540: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA306540> PRJNA352827: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA352827> PRJNA169102: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA169102> PRJNA393269: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA393269> PRJNA553103: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA553103>.

Funding: This project was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Grant 016-10133 (M.E.M., J.R.M.), University of Minnesota College of Veterinary Medicine NIFA Multistate competitive funds 2017-2020, 2020-2022 (M.E.M., J.R.M.), USDA NIFA Agriculture and Food Research Initiative (AFRI) Predoctoral Fellowship 2020-67034-31739 (J.N.C.), Grayson-Jockey Club Research Foundation (C.J.F., J.L.P., R.R.B.), USDA NIFA Animal Breeding and Functional Annotation of Genomes (A1201) Grant 2019-67015-29340 (J.L.P., C.J.F., R.B., T.K., E.G.) as well as NRSP-8 Species Coordinator Funds from the USDA NIFA (C.J.F., J.L.P., R.R.B.), and the the UC Davis Center for Equine Health (C.J.F., J.L.P., R.R.B.). SDA salary was supported by National Institutes of Health's National Center for Advancing Translational Sciences, grants K12TR002492 and UL1TR002494. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health's National Center for Advancing

Similarly, isomiRs - length and sequence isoforms of canonical miRNAs with potentially altered regulatory targets and functions - have received even less attention in non-model species, including the horse, leaving a critical gap in our understanding of their biological significance. To address these challenges, we developed an open-source, containerized pipeline for identifying and quantifying miRNAs and isomiRs (FARmiR: Framework for Analysis and Refinement of miRNAs), and an associated interactive browser (AIMEE: Animal IsomiR and MiRNA Expression Explorer). AIMEE was developed to make miRNA expression data more accessible and user-friendly, a feature often lacking from other expression atlases. These tools were developed using equine data but can be readily extended to other species. Using these tools, we aggregated 461 small RNA-seq datasets, spanning 61 distinct tissues, integrating data from public repositories, an American Quarter Horse cohort, and the Functional Annotation of ANimal Genome (FAANG) consortium Thoroughbred samples, predicting 5,781 miRNAs and isomiRs. This work represents the largest systematically curated atlas of equine miRNA expression to date, providing a valuable resource that will enhance our understanding of miRNA and isomiR functions in tissue-specific regulation and ultimately improve biomarker discovery, functional genomics, and precision veterinary medicine.

Author summary

MicroRNAs (miRNAs) play a critical role in gene regulation, influencing many biological processes including development, metabolism, and disease. Despite their importance, miRNA research in non-model species like the horse has remained limited. Moreover, what research has been done has been more narrowly focused on canonical miRNA expression in specific tissues under various disease states. To address this gap, we developed a pipeline and interactive browser for reproducibly quantifying and visualizing miRNA expression. Using these tools, we aggregated and analyzed 461 small RNA sequencing datasets from 61 equine tissue types, generating the most comprehensive equine miRNA expression atlas to date. Our findings highlight the complexity of miRNA and isomiR expression across tissues and delivers a user-friendly resource for researchers studying gene regulation. Our work enhances the ability to investigate miRNA-regulated biological processes in the horse and provides a framework for similar studies in other species.

Introduction

MicroRNAs (miRNAs) are a class of endogenous small (~22 nucleotides) noncoding RNAs that primarily regulate gene expression through binding messenger RNA (mRNA), triggering mRNA degradation or the inhibition of translation [1]. Over 60%

Translational Sciences. J.N.C was funded by USDA NIFA Fellowships, 2020-67034-31739 and 2024-67012-41761. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

of human protein-coding genes contain at least one miRNA binding site, highlighting the importance of miRNA-mRNA pairing and miRNA-based regulation [2]. MicroRNAs play critical roles in numerous biological processes, including development [3], stem-cell maintenance [4], metabolism [5], apoptosis [6], and cell-cell communication [7]. Given their large regulatory footprint, miRNA dysregulation has been implicated in many diseases [8–12], and miRNA profiling has demonstrated both ubiquitous and tissue-specific expression across species [13–16].

High-throughput sequencing (HTS) for profiling of human small RNA expression has identified miRNA isoforms (isomiRs) that differ from the reference miRNA (Ref-miR) [17]. Originally believed to be sequencing or alignment errors, isomiRs are now thought to be the result of imprecise cleavage or post-transcriptional modification [18–21], resulting in differences in sequence composition and/or length from the reference form [22–24]. IsomiRs can be categorized into five classes based on compositional differences from their Ref-miR counterparts: reference or canonical (identical to the miRNA database sequence [25,26]), 5' isomiR (length changes to the 5' end), 3' isomiR (length changes to the 3' end), polymorphic (sequence modifications with identical length), and mixed (length and sequence changes) [27,28]. The region of the miRNA critical for mRNA target recognition is referred to as the “seed sequence”. IsomiRs with edited or shifted seed regions (e.g., 5'-isomiRs or polymorphisms within residues 2–8) are of particular interest [29,30], as they may have altered target sets that impact similar or different biological pathways as compared to their Ref-miR counterpart [31–36]. Similar to Ref-miRs, isomiRs can be highly expressed [28] with tissue-dependent expression patterns [37,38] and have been associated with various disease states, highlighting their potential as biomarkers [39–41]. Despite the availability of many isomiR-capable processing tools [42], “gold standards” do not exist for isomiR profiling. As a result, isomiRs are often excluded from miRNA studies and databases, introducing profiling biases and missing biological significance [43]. While several miRNA databases exist, only a few include isomiR expression [38,44], and, with a few exceptions, these are limited to humans and mice [45].

Characterization of the tissue-specific miRNA transcriptome in companion animals and livestock has lagged behind that of humans and model organisms for several reasons. First, miRNA profiling is often based on the available Ref-miR sequences from databases such as miRBase (v22) [25], and the more sparse annotation of miRNAs in these species may have a significant impact on expression profiling. For example, the current version of miRBase contains 2,656 mature miRNA sequences in humans and 1,978 in mice, compared to only 690 in horses. Of the miRBase horse miRNAs, over half (359) are based on a single *in silico* study [46], and the remainder (331) are from small RNA-sequencing (RNA-seq) of testes from one horse [47]. Moreover, while earlier versions of MirGeneDB (2.1) [26] did not include horse, the latest release (3.0) [48] provides a database of manually curated, high-confidence miRNAs for EquCab3. While this represents a major improvement in annotation quality over previous resources, it is more conservative in scope, containing 438 equine mature miRNA sequences from 417 precursors (69 of which are not in miRBase). As the goal of this study was to catalog all potentially expressed miRNAs, we included

the broader set of miRBase sequences. Second, few large-scale miRNA expression profiles of normal equine tissues exist. Most of the equine miRNA efforts have focused on profiling select tissues from horses with or without disease [49–53], during exercise [54–56], or at different developmental stages [50]. MiRNA expression profiling across healthy tissues has been previously conducted with only limited tissue sets [57]. The largest study targeted nine tissues, including gluteus medius muscle tissues (the largest muscle in horses, involved in hip extension) from normal horses and horses with polysaccharide storage myopathy type I (PSSM1) [58]. While these studies identified tissue-specific reference and putative novel miRNAs, isomiRs were not thoroughly characterized. To the best of our knowledge, only two equine miRNA analyses have included isomiRs [59,60]. Third, miRNA data can be difficult to access for animal researchers intending to explore and utilize previous findings, and analyses of miRNA data requires bioinformatics proficiency. Finally, the majority of available computational tools and pipelines for characterizing isomiRs were developed for processing human data, many of which require additional informatic manipulations to configure for non-model organisms [42].

To address these limitations, we developed the Framework for Analysis and Refinement of miRNAs (FARmiR) pipeline and Animal IsomiR and MiRNA Expression Explorer (AIMEE) browser. We used these tools to catalog the most comprehensive collection of miRNA expression at the isoform resolution in normal equine tissues, resulting in the identification of 5,463 isomiRs. These resources allow for the reproducible processing and analysis of high-throughput small RNA-seq datasets, and the export of raw and normalized data for additional downstream analyses. The aggregated expression data, together with the pipeline and browser, will be a significant resource for researchers investigating the composition and functional roles of miRNAs in species that lack expression databases like those available for humans and mice. Finally, our pipeline-browser framework will enable researchers to process their own RNA-seq samples in a consistent and reproducible manner, facilitating comparisons across different disease states, phenotypes of interest, and developmental processes with our comprehensive catalog.

Results

Equine miRNA expression atlas

Overview and QC of the primary, FAANG, and public small RNA-seq datasets. A total of 5.9 billion (B) raw reads were obtained from 230 *primary* samples sequenced as part of this study, 86 *FAANG* samples [61,62], and 185 *public* samples (S1 Fig). The 185 public samples were sourced from 14 distinct BioProjects. Of these, 11 projects included a single tissue type, 1 project included two tissues (uterus and uterine fluid), and two projects contained samples from four tissue types.

Following pre-processing, 28 samples (7 *primary*, 1 *FAANG*, and 20 *public*) were excluded due to post-filtered \log_2 read counts less than three median absolute deviations (MAD) from the median of all post-filtered \log_2 read counts (15 serum, 3 uterine fluid, 2 uterus, 2 longissimus muscle, and one each of lamina, left atrium, pituitary, retroperitoneal adipose, sesamoid bone, and subchondral bone). An additional 4 samples (1 *primary* and 3 *FAANG*) passed the MAD filter but consisted of <10% post-filtered reads assigned to miRNA biotypes (2 third metacarpal bone, 1 sesamoid bone, and 1 omental adipose). Subsequently, 3 tissue types were excluded as they had only one sample remaining (third metacarpal bone, sesamoid bone, and uterine fluid). This filtering reduced total tissue types from 64 to 61, with 3 tissue types having samples from all three sources (lung, gluteal and longissimus muscles), 14 from two, and 44 from a single source. Overall, the samples that were retained for quantification consisted of 222 (47%) *primary*, 164 (35%) *public*, and 82 (18%) *FAANG* (Fig 1). Most tissue types from *primary* data were musculoskeletal (59/222, 26.6%), integumentary (54/222, 24.3%), and cardiopulmonary (33/222, 14.9%). Musculoskeletal was also the predominant tissue type from *FAANG* (18/82, 22.0%), followed by nervous system (16/82, 19.5%) and urogenital (14/82, 17.1%). The *public* data were predominantly from blood and related tissues (119/164, 72.6%), followed by urogenital (16/164, 9.8%) and the remaining, constituting <8% per system.

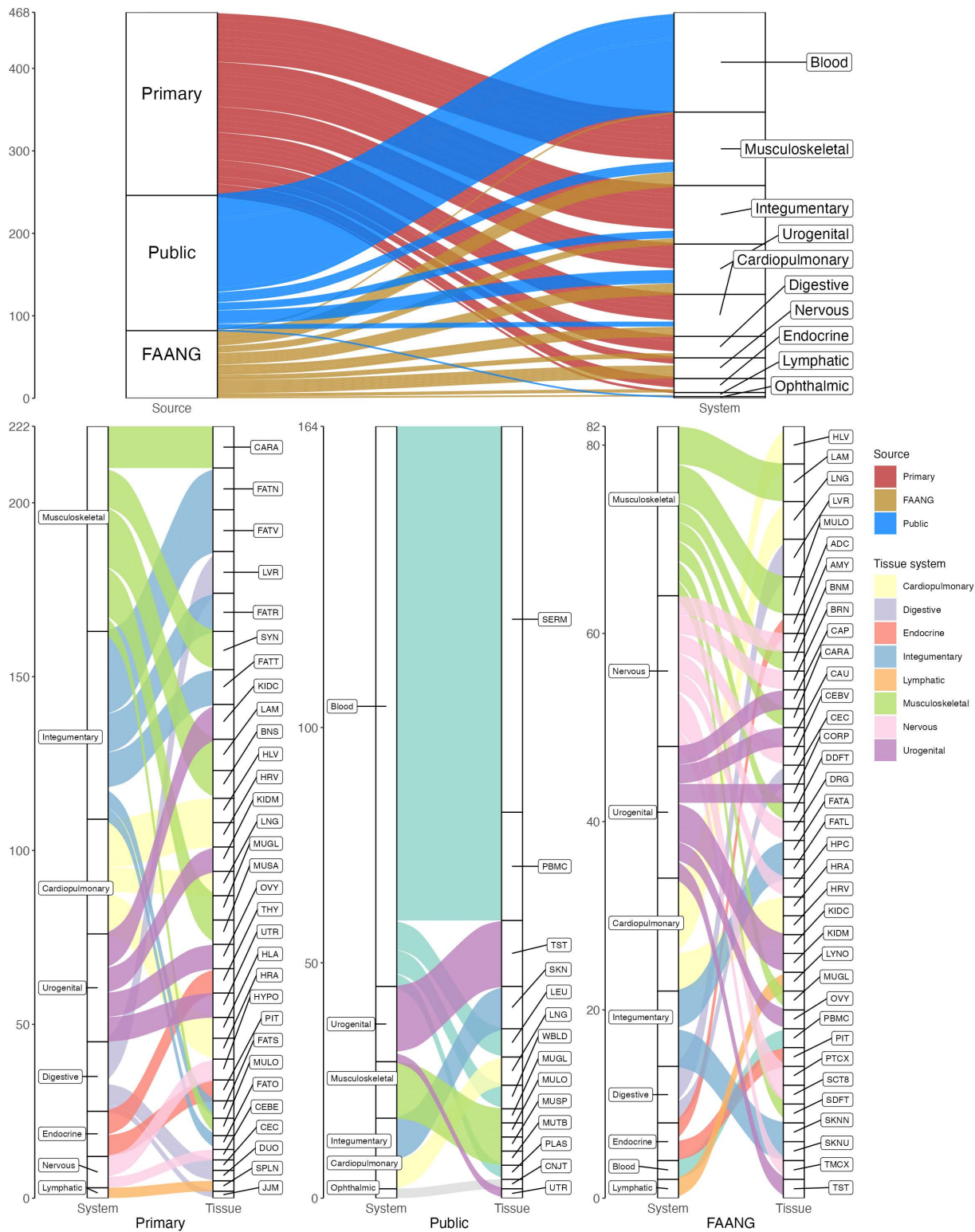


Fig 1. Tissue samples included for quantification by data source and system (top) and tissue type (bottom).

<https://doi.org/10.1371/journal.pgen.1011835.g001>

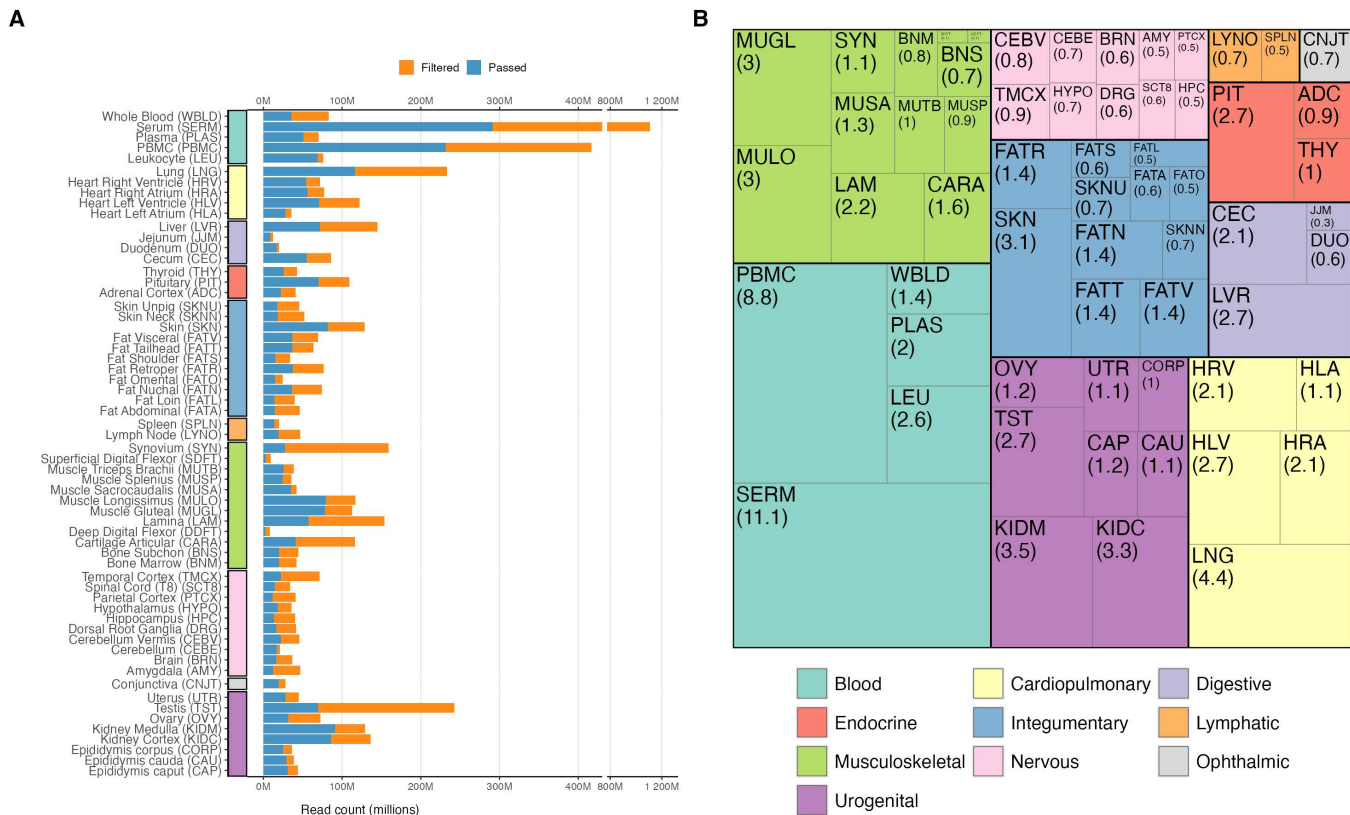


Fig 2. Total read counts of samples included for quantification. (A) Total passed and filtered read counts (in millions) across tissues. **(B)** Passed reads nested within the tissue system (percent of total reads in parentheses). (Note both deep digital flexor tendon (DDFT) and superficial digital flexor tendon (SDFT) have passed read percentages of 0.1).

<https://doi.org/10.1371/journal.pgen.1011835.g002>

After quality and length filtering, 2.6B post-filtered reads were retained with a median of 2.8 million (M) reads per tissue. Serum had the largest number of post-filtered reads (290.9M reads from 82 samples), followed by PBMC (231.5M reads from 25 samples) and lung (116.3M reads from 17 samples), with the smallest being deep digital flexor tendon (2.2M reads from 2 samples) (Fig 2A). Over half of all post-filtered reads are from only three tissue types: blood (25.9%), musculoskeletal (15.8%), and urogenital (15.0%) (Fig 2B).

Among the 468 tissue samples included for quantification, the median number of post-filtered reads was 4.4M (range 428,998–32,121,885) (S2 Fig). Sample-level post-filtered reads (Passed vs Filtered), and transcript type percentages for pre-process (i.e., adapter removed, simple quality filtered, and joined reads) and post-filtered (i.e., pre-quantification reads) are depicted in Fig 3A and 3B. Y RNA and rRNA were the predominant transcript types in 14.3% (67/468) and 5.6% (26/468) of pre-processed samples, respectively. Following processing, only 0.85% (4/468) of samples were dominated by rRNA (2 articular cartilage, 1 serum, and 1 skin) and 0 by Y RNA. Overall, the processing and removal of non-miRNA reads resulted in increased proportions of miRNA reads with a sample mean of 90.4% (SD = 11.9%) post-process compared to the pre-process mean of 66.5%, SD = 29.3%. These results suggest FARmiR processing was reliable in enriching for miRNAs.

Seven samples (6 serum and 1 kidney cortex) with a high proportion of missingness (i.e., at least 90% of miRNAs had RPMs of zero) were removed. The excluded kidney cortex exhibited a much higher percent reduction in passed reads (91.2%) compared to other kidney cortex samples (mean = 35.8%, SD = 10.1). All 6 of the excluded serum samples

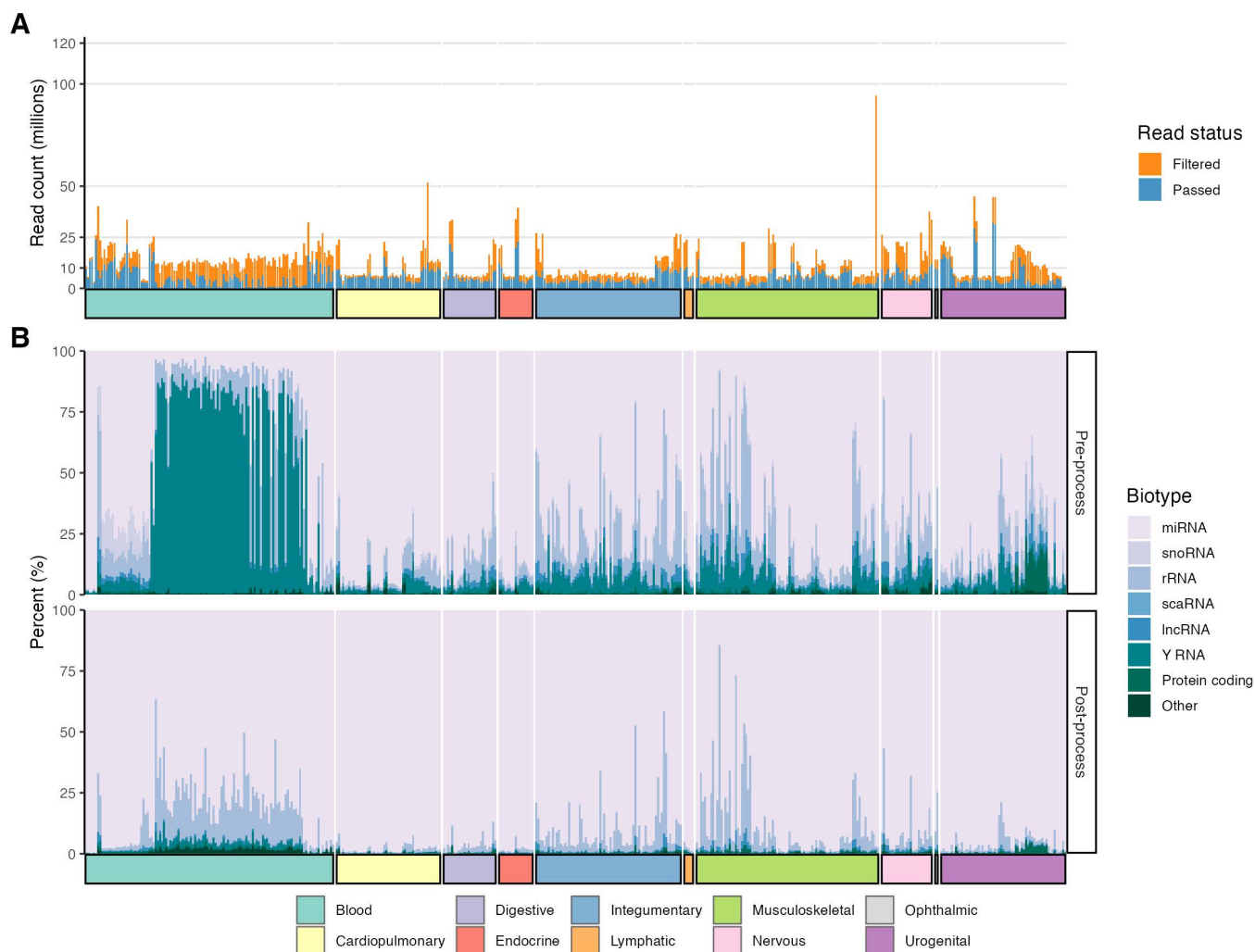


Fig 3. Per-sample passed and filtered read counts and biotype percentages pre- and post-processing. (A) Passed and filtered read counts (in millions) for each sample. **(B)** Percentage of biotypes per *featureCounts* pre- and post-processing by FARmiR.

<https://doi.org/10.1371/journal.pgen.1011835.g003>

originated from the same BioProject and may reflect technical differences in library preparation or sequencing. The final expression set contained 461 samples from 61 tissue types. There were 299 tissues of female origin (7 fillies and 292 mares) and 162 of male (9 colts, 42 geldings, 70 stallions, and 41 unknown). Consequently, the total number of passed reads from female samples (1.54B) was 49.4% more compared male samples (1.02B) (S4 Table). However, when averaged per sample, we observed no statistical difference in passed read depth between female and male samples (mean difference = 40,046; 95% confidence interval (CI): -1.3M – 1.4M). The 461 samples span 15 breeds (plus 4 mixed parentage breeds) with 32.1% (823.3M) of the total passed read depth from Thoroughbreds (n=83; excluding 2 mixed-breed samples), followed by Quarter Horse (n=209; excluding 3 mixed-breed samples) with 31.8% (817.5M), and 13.0% (334.2M) from Warmbloods (n=82) (S5 Table). Although there are imbalances in sample representation between sex and breeds, the consistent per sample read depth between sexes and breadth of tissues across multiple breeds provide a reasonable foundation for cataloging tissue-specific expression.

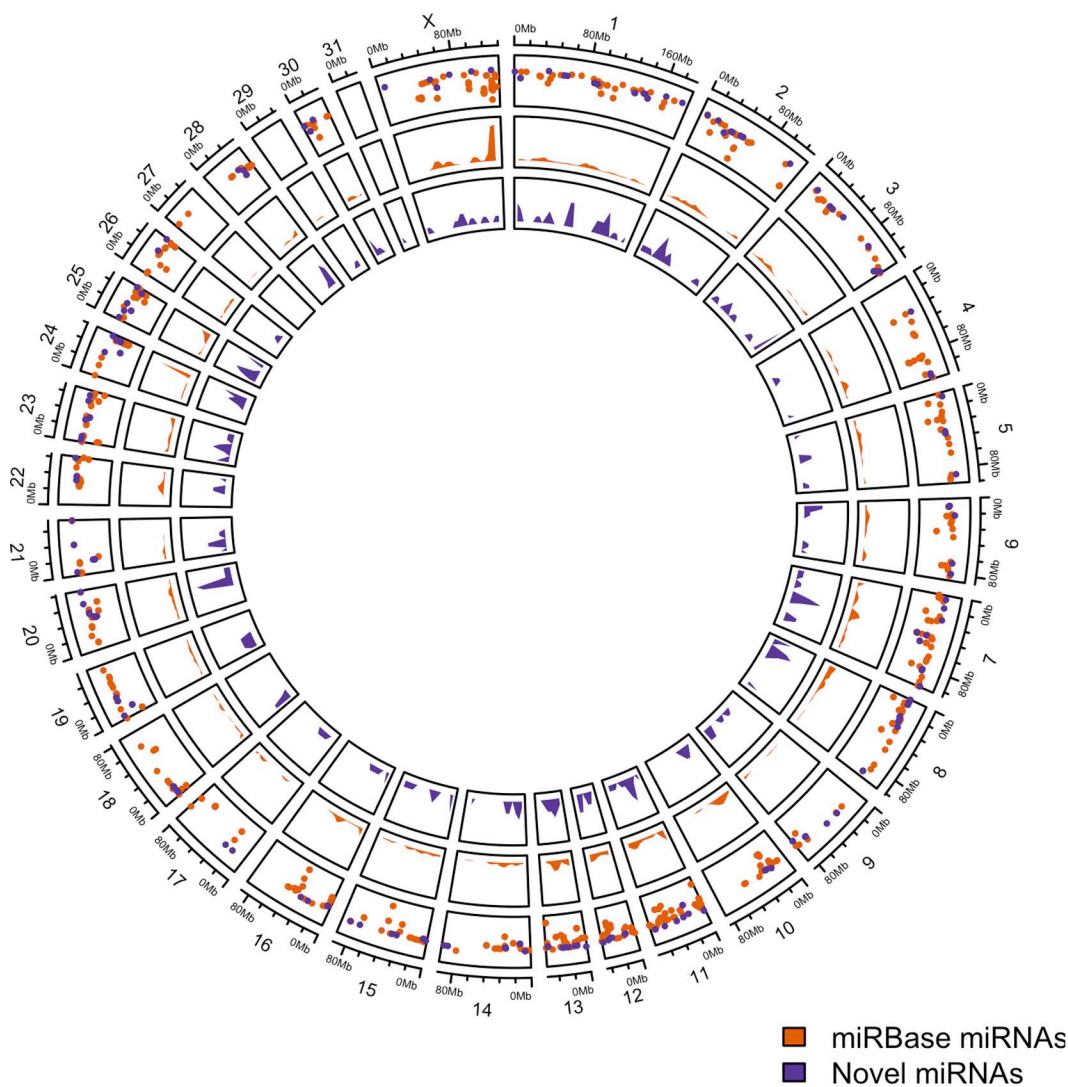


Fig 4. Genomic map of miRNA hairpin loci. Distribution of known (orange) and predicted (purple) miRNA hairpin loci on EquCab3. The outermost ring indicates loci density, with the Y-axis (log scale) representing the distance between neighboring loci.

<https://doi.org/10.1371/journal.pgen.1011835.g004>

EquCab3 miRNA annotation and predicted novel miRNA loci. Of the 715 hairpin loci lifted from EquCab2 to EquCab3, 24 were on unplaced contigs. To evaluate the updated miRNA annotation, only hairpins successfully lifted between the two reference genomes and located on chromosomes were considered ($n = 673$). The majority of hairpins were on the same chromosomes in both EquCab2 and the updated EquCab3 annotation (S3 Fig). From the 461 included samples, 217 mature miRNAs from 195 hairpin loci were predicted (S6 Table). Only 14 (7.2%) of the predicted hairpins were found in another species, compared to 150 (69.1%) of the mature miRNAs with seed matches to other species. A genomic map of canonical and predicted hairpin locations on EquCab3 is depicted in Fig 4.

Quantification of Ref-miRs/isomiRs. A total of 83,057 miRNAs were detected from the included samples. Importantly, the same miRNA sequence can arise from different loci or precursors. For example, *eca-miR-26a* may be processed from two distinct hairpin loci *eca-mir-26a* (chr6:76,295,576–76,295,660(-)) or *eca-mir-26a-2* (chr16:48,848,855–48,848,932(-)). Similarly, the sequence of an *eca-miR-378* isomiR is identical to four novel isomiRs:

eca-novel-miR-2043 (orthologous to the *Bos taurus* (bta) precursor *bta-mir-378d*), *eca-novel-miR-10819*, *eca-novel-miR-13829*, and *eca-novel-miR-15239*. As a result, it is not always possible to confidently assign each mature miRNA sequence to a specific source. Collapsing identical mature sequences across multiple loci resulted in 73,131 distinct miRNA sequences. However, all mature miRNA names and sequence identifiers were retained and are accessible through AIMEE, allowing users to explore potential precursor origins and genomic loci for each sequence.

To reduce the likelihood of false positives, total miRNAs were filtered to include only those with RPMs > 20 in at least 2 samples. The final expression atlas contained 6,951 miRNAs with 5,781 unique mature sequences, of which 318 represent Ref-miRs (248 canonical miRBase entries and 70 novel), 5,448 isomiRs (4,593 canonical and 855 novel), and 15 that are both canonical Ref-miR and novel isomiR. Four of the 248 canonical miRBase entries have identical mature sequences, thus of the 690 distinct miRBase entries, 252 were identified. Of these 252, only 5.2% (13/252) were originally discovered experimentally [47], and the remaining 94.8% (239/252) were predicted *in silico* [46]. Compared to the 387 unique (438 total) miRNA sequences from MirGeneDB v3.0 [48], there were 259 (66.9%) exact matches. Additionally, 2,150 sequences showed full-length containment (1,117 atlas miRNA sequences wholly contained within MirGeneDB and 1,033 MirGeneDB sequences wholly contained within the atlas) bringing the total overlap to 2,409 (41.7%).

Quantification using *isoMIRmap* enables the ability to tag each identified miRNA sequence according to whether it is located exclusively within the “miRNA-space” (the union of all known and predicted hairpin sequences) [63]. In this way, miRNAs are tagged in reference to the miRNA-space as exclusive (found only within the miRNA-space) or ambiguous (found outside of the miRNA-space). Of the 6,951 miRNAs, 5,498 (79.1%) were exclusive to the miRNA-space, with an additional 116 (1.7%) also located within repeat islands (e.g., DNA, LINE, or SINE elements) (Fig 5A). The majority of identified miRNAs were isomiRs of “mixed” type (4,121/6,951 - 59.3%) (i.e., sequence changes at both the 5' and 3' ends compared to the canonical sequence), followed by strictly 3' (1,469/6,951 - 21.1%), and 3' non-templated (683/6,951 - 9.8%). IsomiRs containing only 5' templated modifications were observed at the lowest frequency (302/6,951 - 4.3%) (Fig 5B). All identified isomiRs were derived from a total of 493 canonical and predicted Ref-miRs of which nearly half (237/493 - 48.1%) existed with at most five different seed regions (i.e., isomiRs with changes to the 5' end) (Fig 5C). The top five miRNAs with the highest number of seed shift variants were *eca-miR-7* (n = 174), a novel miRNA *eca-novel-miR-6634* (n = 126), *eca-miR-140-3p* (n = 125), *eca-miR-1* (n = 121), and *eca-miR-133a* (n = 108).

Robust rank aggregation was used to identify the top five miRNAs per tissue across the entire atlas (Fig 6). The canonical form of *eca-miR-26a* was the most abundant, occurring in the top five for 50 of the 61 tissues (82.0%), followed by *eca-miR-143* in both canonical (24/61 - 39.3%) and 3' templated (20/61 - 32.8%) forms, canonical *eca-let-7a* (18/61 - 29.5%), and *eca-miR-148a* (16/61 - 26.2%). Besides the canonical Ref-miR and isomiRs, two predicted miRNAs (named

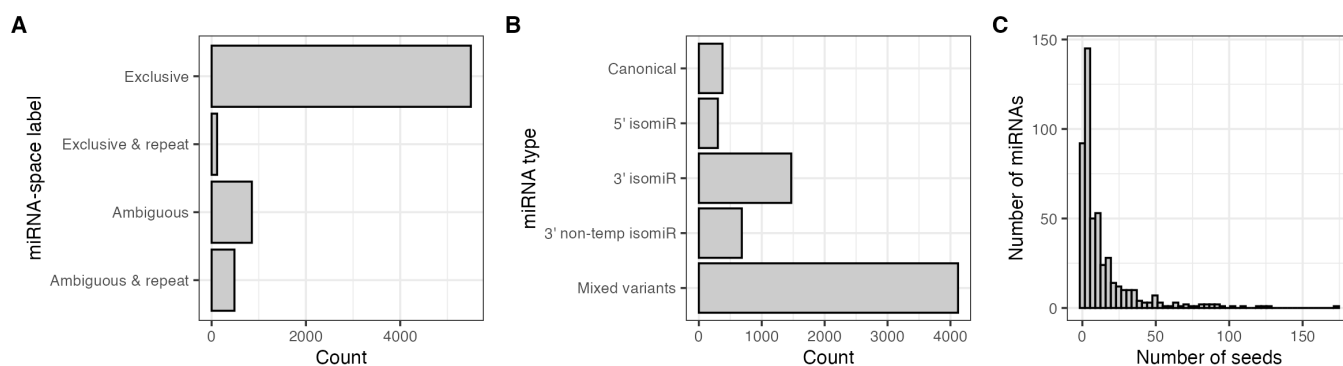


Fig 5. MiRNA location tags, variant types, and seed distribution. Counts of miRNA location labels (A), canonical and isomiR forms (B), and the distribution of seed shifting variant forms for each identified canonical and predicted Ref-miR (C).

<https://doi.org/10.1371/journal.pgen.1011835.g005>

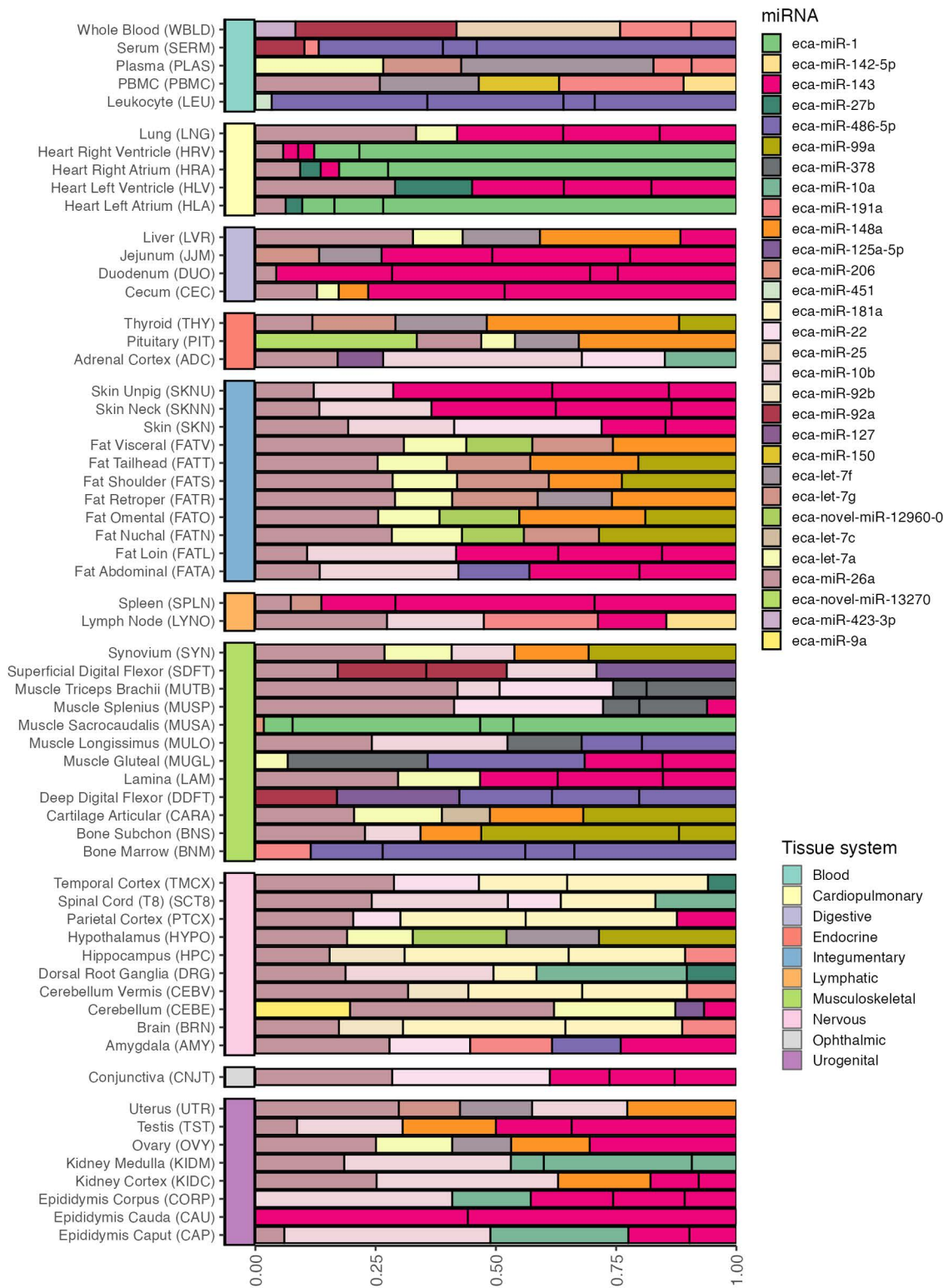


Fig 6. Proportions of the top 5 miRNAs per tissue based on robust rank aggregation.

<https://doi.org/10.1371/journal.pgen.1011835.g006>

via the default *mirPro* naming convention) were also ranked highly in select tissue types. Notably, the canonical form of the predicted miRNA *eca-novel-miR-12960-0* (chr6:80,187,888–80,187,910(+)) was in the top five for the hypothalamus and three adipose tissues (nuchal, omental, and visceral). Furthermore, a 3' templated isomiR of *eca-novel-miR-13270* (chr6:8,343,131–8,343,150(-)) was in the top five across pituitary samples. Amongst the included tissues of the digestive tract (cecum, duodenum, and jejunum), either canonical or isomiR variants of *eca-miR-143* were top ranked. Similarly, in the epididymis cauda, *eca-miR-143* was overwhelmingly the most dominant miRNA. The canonical *eca-miR-1* was top ranked in the right ventricle of the heart and the left and right atria but not the left ventricle.

Tissue specificity. Among the 5,781 unique mature miRNA sequences, 4,032 had a mean RPM > 20 in at least one organ across all included samples. Tissue specificity was determined using OEI values where OEI < 0.15 was considered ubiquitous, OEI between 0.15 and 0.85 intermediate, and OEI > 0.85 tissue-specific. As described in the methods, mean RPMs were calculated for each miRNA per organ. Tissue types from the same organ (skin sections, chambers of the heart, kidney regions, and brain subregions) were averaged together, while other tissues were retained separately. Only 9 (0.22%) miRNAs were detected in a single organ, 6 (0.15%) in two, 10 (0.25%) in three, while the remaining were observed in all (762/4,032 - 18.9%) or found in some but absent in others. Due to the unbalanced nature of the included projects and the resulting inability to control for technical variation between samples of the same tissue, OEIs were calculated separately for the *primary* (n = 2,570 miRNAs with at least one organ RPM mean > 20) and *FAANG* (n = 2,831) sources. Over half (1,560/2,570 - 60.7%) of miRNAs from primary were expressed within the intermediate OEI range (0.15 - 0.85), and 39.3% (1,010/2,570) were classified as tissue-specific (OEI > 0.85) (Fig 7A). Conversely, mean organ expression from FAANG samples yielded 34.5% (977/2,831) intermediate and 65.4% (1,852/2,831) tissue-specific miRNAs (Fig 7B). Moreover, two miRNAs were uniquely enriched (OEI = 1) in single tissues, a 3'-isomiR of *eca-miR-7* in the pituitary and a 5'-isomiR of *eca-miR-215* in the cecum.

AIMEE usage examples

AIMEE delivers a range of analyses and visualizations for the processed miRNA expression data (Fig 8). We present two brief application scenarios an end user may undertake.

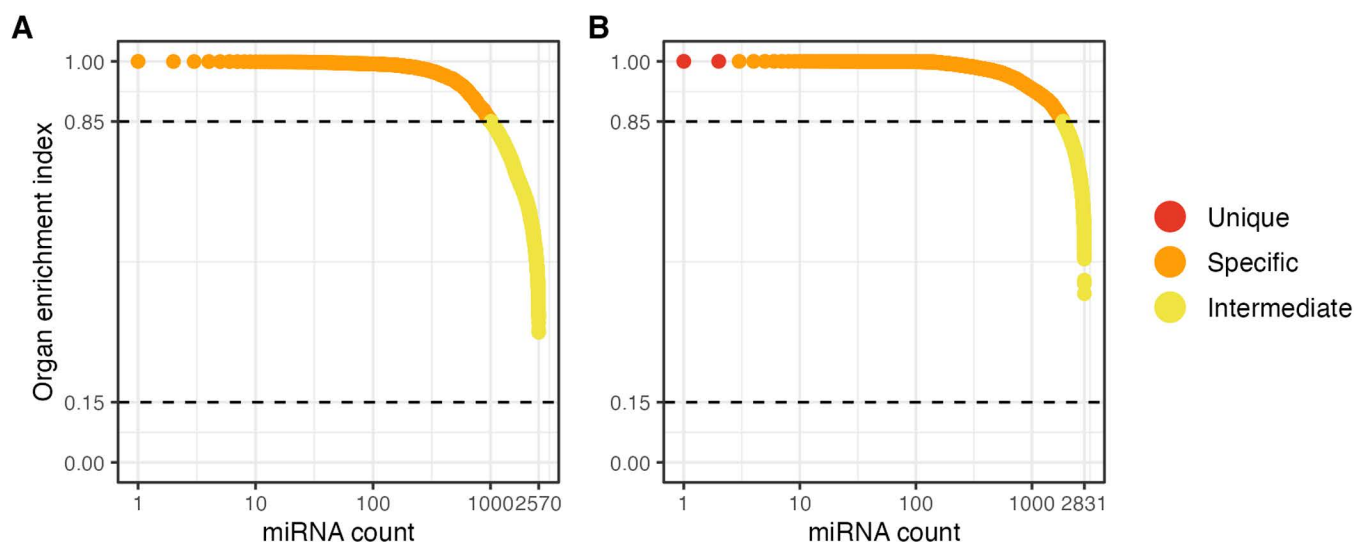
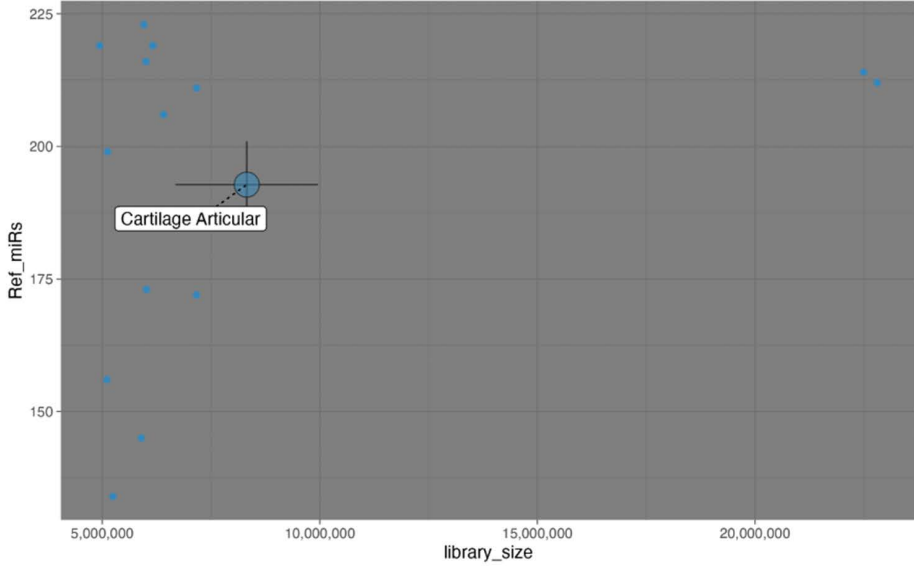


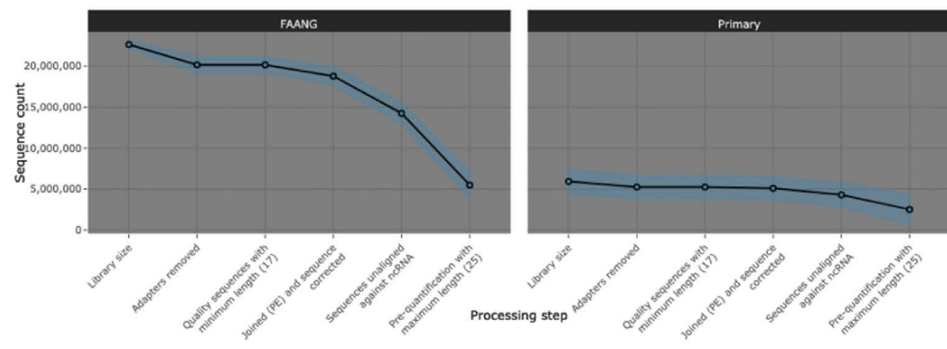
Fig 7. Organ enrichment index for miRNAs from *primary* (A) and *FAANG* (B) data.

<https://doi.org/10.1371/journal.pgen.1011835.g007>

Process explorer



Stepwise read loss



Top miRs

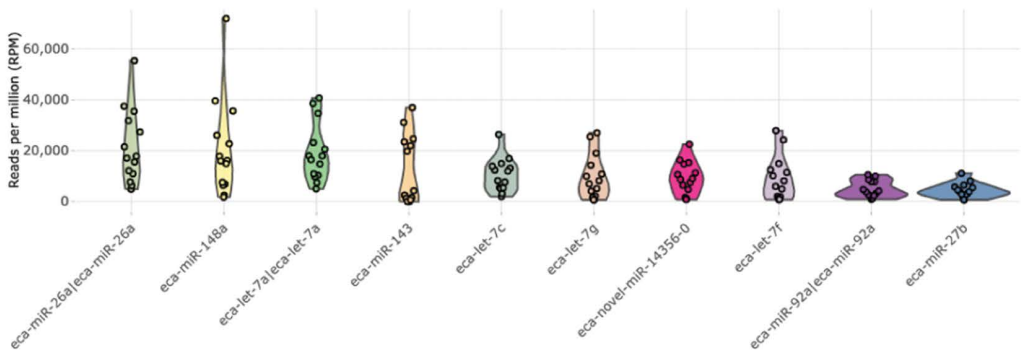


Fig 8. Select exploration module examples of the equine miRNA expression atlas supported by AIMEE. *Process explorer* (top) enables quick access to the counts of each processing step. In this example, the number of unique Ref-miRs is plotted as a function of library size for articular cartilage samples. Each point represents a single sample, and the larger points are the centroids of the highlighted tissue. *Stepwise read loss* (middle) charts the read attrition from library size through pre-quantification, which can be summarized at the tissue level (as shown) or displayed for each sample. *Top miRs* (bottom) allows visualization of the top N Ref-miRs or isomiRs based on mean RPMs from a selected source and tissue.

<https://doi.org/10.1371/journal.pgen.1011835.g008>

Use case: select a tissue of interest, determine the top-expressed miRNAs, and explore their expression across all available tissues

Researchers are often interested in characterizing the miRNA profile from a specific tissue and identifying miRNAs that are consistently highly expressed within that tissue (e.g., across samples or time points). Without access to healthy tissue samples from which to prepare libraries and sequences, especially in the case of tissues requiring euthanasia, a researcher may need to either identify relevant data from public repositories or parse expression matrices and metadata. In both cases, advanced bioinformatics expertise is often required, and for the latter scenario, assumes the methodology used to generate the expression data was precise. AIMEE bypasses these hurdles, providing access without requiring exhaustive searches of public repositories for appropriate datasets or extensive bioinformatics experience. For example, the subchondral bone plays an important role in the development and progression of osteoarthritis (OA) in both horses and humans [64]. While there have been studies investigating miRNAs as potential biomarkers for OA in horses [53,65,66], subchondral bone has not yet been reported on in this context. Using the AIMEE *Rank aggregation* view, all available samples with subchondral bone data ($n=8$; *primary* only) can be selected, and the top 10 miRNAs are calculated using RRA [67,68] and visualized using bump (Fig 9A) and box plots (Fig 9B). The top ranked miRNA, *eca-miR-99a*, can then be used as input to examine where else it is expressed (Fig 9C).

Use case: exporting raw counts and differential expression

Formatted raw (or normalized) expression data and metadata may be extracted. With careful consideration of the data source and potential batch effects, these exports are suitable for downstream analyses, including differential expression (DE) between tissues or samples of interest. To demonstrate this utility, un-normalized count and metadata data from kidney cortex and medulla samples were exported for DE analysis with DESeq2 [69]. Briefly, pre-filtering was applied to the un-normalized counts retaining only miRNAs with counts of at least 10 in a minimum of 6 samples (the smallest group size). Un-normalized (raw) counts of expressed miRNAs were analyzed using DESeq2 with a negative binomial generalized linear model that included individual horse and tissue type. Including horse or sample in the model accounts for the differences between individuals while estimating the effect due to tissue type. The Benjamini-Hochberg method was used to control the FDR with an adjusted p -value < 0.01 set for miRNAs to be considered differentially expressed. This more stringent threshold was chosen to avoid identifying false positives at the expense of false negatives. Kidney was chosen as an example as there are six paired samples per group or structure, the recommended minimum for DESeq2 DE [70], from the same source (*primary*) to minimize batch effects. Of the 2,673 pre-filtered non-zero miRNAs, 81 (3.0%) were significantly up-regulated (FDR p -value < 0.01), and 154 (5.8%) down-regulated in the medulla compared to the cortex (Fig 10A). Of the 235 DE miRNAs, 10.6% (10/235) were classified as canonical or predicted Ref-miRs, while 89.4% (210/235) were isomiRs. Moreover, when considering mixed type with 5' modifications and 5' isomiRs, 32.0% (57/235) of DE miRNAs had seed-shifting alterations (Fig 10B). The Ref-miR and IsomiRs of *eca-miR-192* were the most abundant DE miRNAs at 12.8% (30/235), all of which were down-regulated in the cortex compared to the medulla. A 3' non-templated isomiR of *eca-miR-205* had the highest \log_2 fold change at 8.78 (FDR p -value < 0.001), which can be visualized as user-selected input with AIMEE (Fig 10C). Throughout the atlas, *eca-miR-205* expression was not restricted to the kidneys but was also observed in other tissues (e.g., lamina).

Discussion

From 462 samples across 61 tissue types, AIMEE includes the largest collection of uniformly processed equine miRNA expression data to date. This expression atlas and explorer contains nearly 6,000 putative and canonical miRNA sequences, ~95% of which are isomiRs. The inclusion of isomiRs is of particular value as the importance of these molecules has rapidly grown in the last decade [28]. Consistent with previous studies, the equine atlas contains far fewer 5'-isomiRs compared to 3'-isomiRs [32]. Due to the well-known batch effects inherent in RNA-seq data [71,72], robust

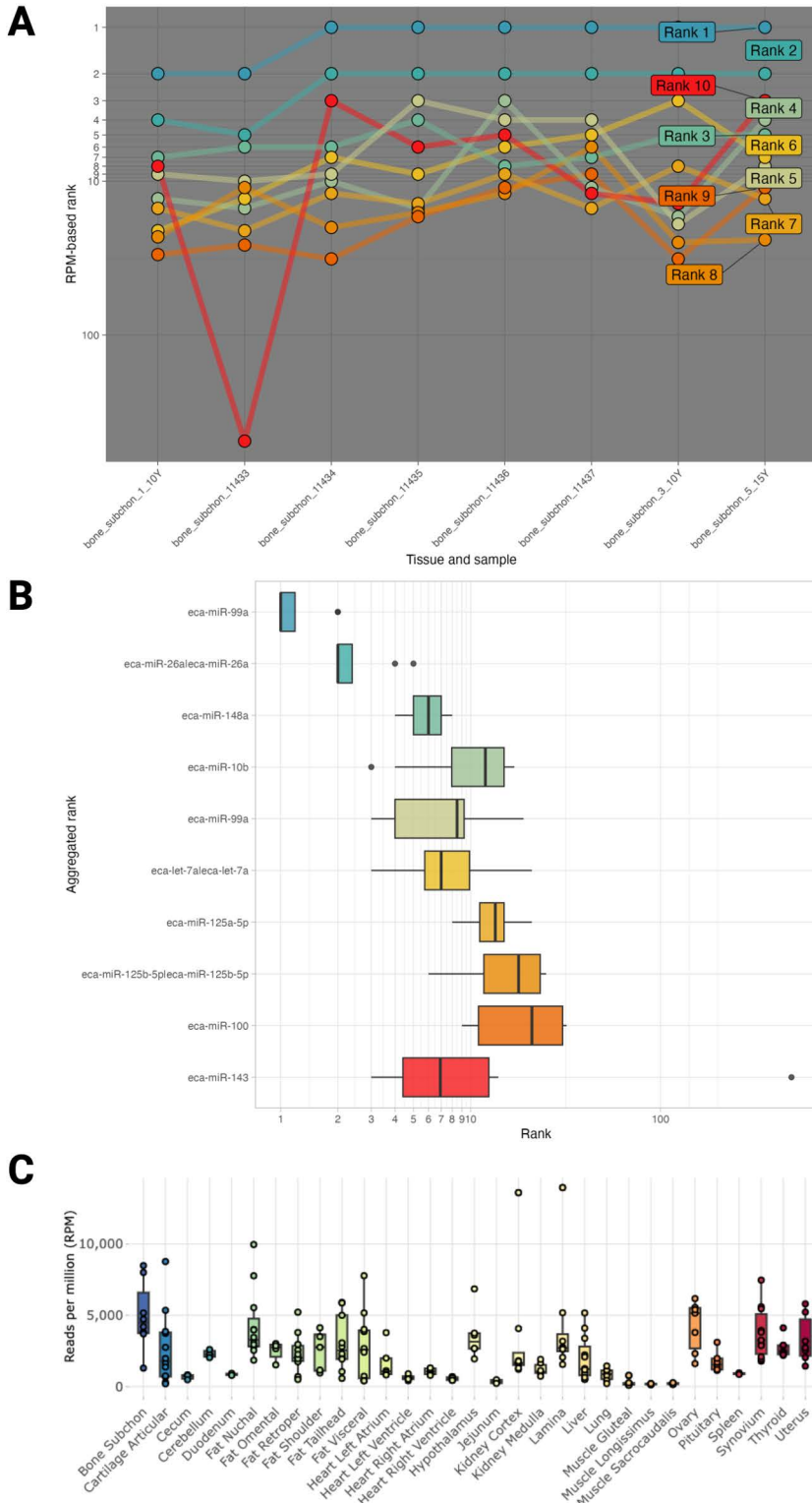


Fig 9. Application of AIMEE to identify top-ranked miRNAs in a tissue of interest. (A) Rank bump chart following robust rank aggregation from subchondral bone samples ($n=8$). **(B)** Aggregated ranks of the top 10 expressed miRNAs. **(C)** Expression of the subchondral bone top ranked miRNA (*eca-miR-99a*) across all primary tissues.

<https://doi.org/10.1371/journal.pgen.1011835.g009>

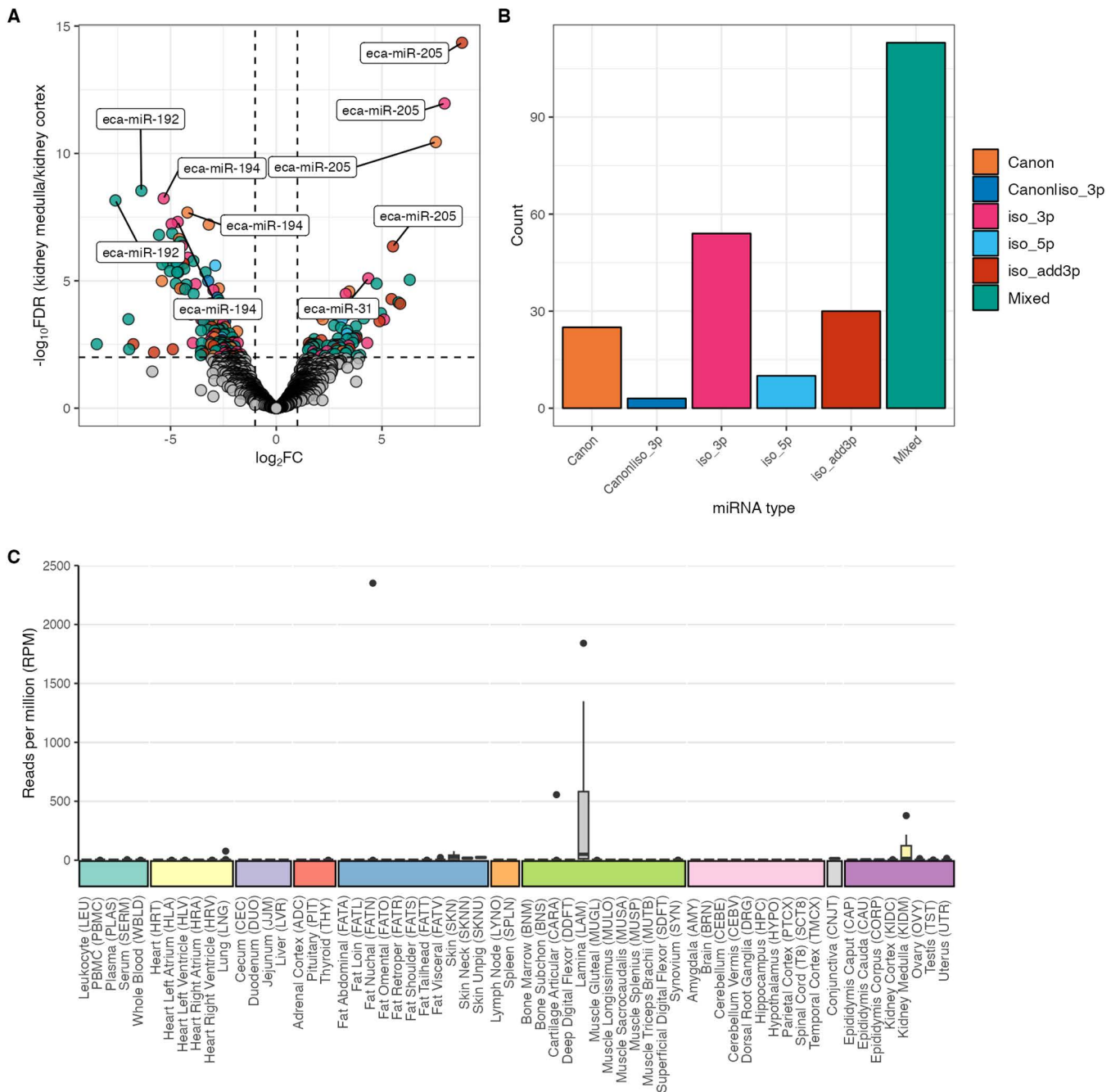


Fig 10. Application of AIMEE to export normalized expression data for downstream analysis. (A) Volcano plot of differentially expressed miRNAs between kidney medulla and cortex with \log_2 fold (X-axis) and $-\log_{10}$ FDR adjusted p -values. Grey circles represent non-significant miRNAs (FDR p -value > 0.01). The top five miRNAs with increased expression and top five miRNAs with decreased expression are labeled with the canonical name and colored by form. **(B)** Counts of the significant miRNA types from **A**. **(C)** Tissue-level RPM expression of *eca-miR-205* across the entire atlas.

<https://doi.org/10.1371/journal.pgen.1011835.g010>

rank aggregation was used to circumvent the challenges associated with batch correction methods and integrate samples from heterogeneous sources. Also, in the context of minimizing unwanted technical variation, the tissue specificity index (TSI) was calculated for each unique miRNA that was sufficiently expressed (i.e., mean RPM > 20 in at least one organ) independently for the included *primary* and *FAANG* samples. From the *FAANG* samples, for example, a 3'-isomiR of *eca-miR-7* and a 5'-isomiR of *eca-miR-215* were exclusively expressed in the pituitary and cecum, respectively. The *miR-7* miRNA family has been associated with regulating gene expression in the pituitary of mice [73], rats [74], and cattle [75]. Similarly, the *miR-215* family has been repeatedly identified as a critical tumor suppressor in gastric and colorectal cancer [76,77], as well as associated with ulcerative colitis [78].

The scope of this miRNA expression atlas has implications for both basic research and future clinical applications. The tissue-specific miRNA and isomiR profiles generated as part of this work advances our understanding of gene regulatory mechanisms across equine tissues. In particular, the isomiR cataloging of various muscle types, cartilage, and blood, creates opportunities for biomarker discovery, improving precision in the diagnosis and monitoring of various equine conditions. Data from previous studies investigating the role of miRNAs in equine osteochondritis and osteoarthritis [49,79] can be analyzed along with the uniformly processed tissue data from this study, possibly contributing to the identification of additional biomarkers for early osteoarthritis detection. Furthermore, understanding the distribution of specific miRNAs across tissues is essential for developing targeted therapeutic interventions, particularly for metabolic and inflammatory conditions where miRNA-mediated regulation plays a significant role [80]. The characterization of isomiRs in individual tissues highlights new possibilities for personalized medicine approaches. For example, the isomiR catalog we developed for miRNAs like *eca-miR-7* (174 seed variants) and *eca-miR-140-3p* (125 seed variants) underscores the potential for individual variation in isomiR expression. This type of fine-tuned control of the regulatory machinery could have significant functional consequences that may warrant careful consideration in therapeutic strategies.

This equine miRNA expression resource does have some limitations. First, publicly available samples that met the inclusion criteria were based entirely on the descriptions reported by the authors in the original published main text, supplemental materials, or data repositories. Consequently, we cannot be as confident in the health assessments of these *public* samples compared to the *primary* or *FAANG* sets which included extensive ante and postmortem phenotyping. Second, we only identified 252 of the 690 miRBase-defined canonical miRNAs [25]. This is in contrast to previous studies that identified greater proportions, for example, in the testes [50,81]. This could be for a couple of reasons. Identifying miRNAs in earlier studies, especially in studies that were not using tools designed for miRNA identification but instead repurposing general aligners like BWA [82], can over-inflate miRNA counts. Even when using tools designed for miRNA profiling, there are significant differences in sensitivity and accuracy and poorly overlapping miRNA sets between tools [83]. Additionally, misannotated or spurious miRNAs in the reference set may hinder accurate identification and quantification. Many profiling tools utilize miRBase [25] or miRCarta [84] to define the miRNA space of a given species. For the horse, this may be suboptimal, as the 690 miRBase reference is based on two studies, with over half (359) being based on *in silico* prediction [46], and the remainder (331) from small RNA of testes of one horse of unknown breed and health status [47]. Both of these studies relied on the previous horse assembly [85]. In contrast, MirGeneDB v3.0 provides a curated set of 438 mature miRNAs (387 unique sequences) based on the current EquCab3 assembly [86]. We observed 259 of the 387 unique sequences (66.9%) exactly matched the mature sequences identified in this atlas, with an additional 2,150 exhibiting full-length containment. These findings underscore the potential biological complexity introduced by isomiRs and the challenges of maintaining comprehensive miRNA annotations. Importantly, FARmiR was not designed for this task but instead to catalog expression using the known miRNA space and predict additional miRNA loci beyond that space. As a result, it is possible that the parental miRNA precursors (and locations) have not been filtered against the current repeat element annotations, leading to hairpin-shaped precursors that are not true miRNAs [63]. Third, we did not include any *in silico* miRNA target prediction nor enrichment analyses based on predicted targets. This decision was made for two reasons: 1) it has been previously reported that computational target prediction (without experimental

validation) in non-model organisms is not reliable [87]. 2) A follow-up to this study will couple prediction and validation to predict the RNA targets more accurately. Finally, this atlas profiled miRNA expression at the tissue-level, using bulk small RNA-seq. Characterization at the cellular level has demonstrated cell-specific expression of both human Ref-miRs and isomiRs [38,88,89]. Although still in its infancy and technically challenging, recent advances in HTS technologies have demonstrated the feasibility and potential for single-cell miRNA-seq protocols [90,91]. To date, there have been no rigorous characterizations of the equine miRNA landscape using single-cell RNA-seq methodology, thus representing a natural extension of this current work.

Using FARmiR and AIMEE represents a unified approach to miRNA data processing, analysis, and visualization. As an open-source, containerized pipeline for miRNA identification and quantification, FARmiR aligns with the principles of sustainable data analysis, promoting transparency, adaptability to other species, and *in silico* reproducibility [92,93]. AIMEE provides researchers a user-friendly and accessible option for exploring miRNA tissue profiles without the substantial effort and cost associated with preparing small RNA-seq data. AIMEE also includes a filtering option to consider only the mature sequences from this atlas that overlap with the curated set provided by MirGeneDB v3.0 [48]. The *primary*, *FAANG*, and *public* data, uniformly processed with FARmiR and accessible via AIMEE, contribute valuable information about the expression of equine miRNAs and, ultimately, the tissue-specific regulatory functions contributing to equine health and performance.

Materials and methods

Ethics statement. Tissues were collected from horses donated for use in this study at the time of euthanasia. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Minnesota (1712-35369A). All methods were performed in accordance with the IACUC guidelines and regulations.

Three small RNA-seq data sources were utilized for this atlas, one *primary* (i.e., the 12 horses enrolled as part of this study) and two public sources. The latter two sources consisted of four Thoroughbreds from the publicly available Functional Annotation of Animal Genomes (FAANG) consortium (referred to as *FAANG*), and systematic identification of compatible datasets from public databases.

Primary data source

Horses. A total of 12 healthy horses (11 American Quarter Horses and 1 American Paint Horse) donated to the University of Minnesota provided the *primary* source of data for this study. Physical examinations were performed by a boarded large animal internal medicine specialist to ensure horses were clinically normal. Horses were housed on a dry lot as a group and were fed free choice hay and water. Isolation and twice daily temperature checks were performed for two weeks to safeguard against communicable diseases. The *primary* source consisted of two cohorts with the first in the fall of 2013 (3 mares, one gelding, and one stallion 5–15 years of age) and the second in the fall of 2020 (7 mares 12–14 years of age).

Tissue collection and RNA isolation. Horses were sedated with 0.05 mg/kg intravenous (IV) injection of xylazine, an alpha-2 adrenergic agonist, and humanely euthanized with ≥ 100 mg/kg injection of pentobarbital sodium (Fatal-Plus, Vortech Pharmaceutical) following the American Association of Equine Practitioners Euthanasia Guidelines (<https://aaep.org/resource/euthanasia-guidelines/>). Immediately following euthanasia, all tissues were collected in triplicate and flash frozen in liquid nitrogen or placed in RNAlater Stabilization Solution (ThermoFisher Scientific). Both flash-frozen and RNAlater aliquots were stored at -80°C until further processing. A total of 31 tissues were collected from the first cohort and up to 70 tissues from the second cohort (S1 Table). All tissues were collected within 45 min of euthanasia. The following tissues (n = 31) were selected for further processing: articular cartilage, lamina, subchondral bone, synovium, multiple muscles (gluteal, longissimus, and sacrocaudalis dorsalis), duodenum, jejunum, cecum, liver, spleen, thyroid, kidney (cortex and medulla), pituitary, cerebellum, hypothalamus, lung, heart (left and right atrium and ventricle), multiple

fat depots (nuchal, omental, retroperitoneal, shoulder, tailhead, and visceral), ovary, and uterus. Only tissues with at least two biological replicates were included for library preparation and sequencing. Tissues were selected to capture a range of potential expression diversity across tissue types, as well as to represent those actively being investigated by our laboratory and others. The remaining tissues are biobanked in the University of Minnesota Equine Genetics and Genomics Laboratory tissue repository. RNA was isolated using tissue-type specific protocols to maximize RNA quality and yield. Due to wide diversity of tissue types and composition, no single extraction protocol was appropriate for all sampled tissues. Tissue type specific protocols were developed and refined based on composition and assessment of RNA quality via NanoDrop 8000 spectrophotometer (Thermo Scientific, cat. no ND-8000-GL, Waltham, MA, USA) and the Agilent TapeStation system (Agilent Technologies, Santa Clara, CA, USA). See Supplemental Materials for complete descriptions of isolation protocols. Tissues isolated and sequenced as part of this study are referred to as *primary* throughout the remainder of this manuscript.

Small RNA library preparation and sequencing. Sample QC, library preparations, and sequencing reactions were conducted at GENEWIZ, LLC/Azenta US, Inc (South Plainfield, NJ, USA). The RNA samples were quantified using a Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA), and RNA integrity (RIN) was checked using the Agilent TapeStation (Agilent Technologies, Santa Clara, CA, USA). The Small RNA sequencing libraries were prepared using the NEBNext Small RNA Library Prep Kit for Illumina using the manufacturer's instructions (New England Biolabs, Ipswich, MA, USA). Briefly, Illumina 3' and 5' adapters were added to RNA molecules with a 5'-phosphate and a 3'-hydroxyl group sequentially. A reverse transcription reaction was used to create single-stranded cDNA. The cDNA was then PCR amplified using a common primer and a primer containing an index sequence. The amplified cDNA construct was purified by polyacrylamide gel electrophoresis, and the correct band (~145–160 bp) was excised from the gel and eluted with water. The eluted cDNA was concentrated by EtOH precipitation generating the sequencing libraries. The sequencing libraries were multiplexed and clustered onto a flowcell. After clustering, the flowcell was loaded onto the Illumina HiSeq 4000 or NovaSeq 6000 instrument according to the manufacturer's instructions. The samples were sequenced using a 2 × 150 bp paired-end (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) were converted into FASTQ files and de-multiplexed using Illumina bcl2fastq 2.20 software. One mismatch was allowed for index sequence identification.

Equine Functional Annotation of ANimal Genomes (FAANG) data

Ethics statement. All procedures were approved by the Institutional Animal Care and Use Committee at the University of California - Davis (protocols #19037 and 21033). All methods were performed in accordance with the IACUC guidelines and regulations.

Horses. As part of the equine FAANG project, small RNA-seq libraries were constructed and sequenced using tissue samples collected from two healthy Thoroughbred mares (aged 4 and 5 years) [61] and two healthy Thoroughbred stallions (aged 3 and 4 years) [62].

Tissue collection and RNA isolation. Horses were sedated and humanely euthanized prior to sample collection and preservation as described [61,62]. The following tissues (n = 39) were selected for RNA isolation and miRNA sequencing: adrenal cortex, amygdala, articular cartilage (fetlock and stifle), bone marrow, brain, cecum, cerebellum vermis, deep digital flexor tendon, dorsal root ganglia, epididymis (caput, cauda, and corpus), fat depots (abdominal and loin), heart (left and right ventricle, right atrium), hippocampus, kidney (cortex and medulla), lamina, liver, lung, lymph node, muscle types (gluteal and longissimus dorsi), ovary, parietal cortex, peripheral blood mononuclear cells (PBMC), pituitary, sesamoid bone, two skin sections (neck and unpigmented), superficial digital flexor tendon, T8 thoracic vertebrae, temporal cortex, testis, and third metacarpal bone (S2 Table).

Small RNA library preparation and sequencing. Sample QC, library preparations, and sequencing reactions were conducted at the University of Minnesota Genomics Center (UMGC) (n = 40) or Admera Health (n = 46). At UMGC,

total RNA isolates were quantified using a fluorimetric RiboGreen assay on a BioTek Synergy 2 (BioTek Instruments, Winooski, VT, USA). Total RNA integrity was assessed using capillary electrophoresis (Agilent BioAnalyzer 2100, Agilent Technologies, Santa Clara, CA, USA) to generate RNA Integrity Numbers (RIN). While UMGC recommends total RNA yields of at least 500ng and RINs > 8, TruSeq Small RNA libraries were prepared for samples (n = 36) with a range of RINs (5.4 – 9.5). Libraries for four bone-derived samples (two sesamoid and two third metacarpals) with the lowest RINs (3.7 – 4.0) were prepared using the Clontech smRNA-Seq kit (Takara Bio USA, San Jose, CA, USA) for Illumina following the manufacturer's instructions, which is more accommodating for low-input RNA. The sequencing libraries were multiplexed and clustered onto a flowcell. After clustering, the flowcell was loaded onto the Illumina HiSeq 2500 and sequenced using a 2 × 50 bp PE configuration. Raw sequence data (.bcl files) were de-multiplexed and converted into FASTQ files using CASAVA v1.8.2 software. At Admera Health, isolated RNA sample quality was assessed by BioAnalyzer RNA Eukaryotic Nano Assay (Agilent Technologies, Santa Clara, CA, USA) and quantified by Invitrogen Broad Range RNA Qubit Assay (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Library construction was performed with TruSeq Small RNA kit (Illumina, San Diego, CA, USA). Final libraries quantity was assessed by Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and quality was assessed by TapeStation HSD1000 ScreenTape (Agilent Technologies, Santa Clara, CA, USA). Mean final library size was about 200 bp with an insert size of about 50 bp. Illumina 6-nt single-indices were used. Equimolar pooling of libraries was performed based on QC values and sequenced on an Illumina NovaSeq 6000 with a read length configuration of 150 PE for 20M PE reads per sample (10M in each direction) and data was trimmed to 1x50 sequencing configuration. Tissue samples sequenced across different runs or lanes were combined prior to processing.

In the *primary dataset*, both fragment ends were sequenced for all libraries. However, only the first read of the *FAANG* data was available and the included *public* samples were sequenced from single-end (SE) libraries. Therefore, only the first read from the *primary* dataset was included to mitigate library configuration as a source of unwanted technical variation or batch effect, an unavoidable and inevitable challenge with aggregating multi-center RNA-seq data [71,94].

Public data sources

Database search and validation. The National Center for Biotechnology Information (NCBI) BioProject and PubMed databases were queried using “(equine OR horse) AND (miRNA OR microRNA)” on April 4, 2023, and May 3, 2023, respectively, to identify potentially relevant studies and datasets. The search was limited to studies and datasets describing Illumina-based small RNA-seq of equine tissue samples with publicly available sequencing data. Studies that profiled non-blood derived fluids/cell types (e.g., semen, milk, synovial fluid), extracellular vesicles, or other culture-expanded cell types (e.g., skeletal muscle satellite cells, chondrocytes) were excluded to focus on native tissue-derived expression. These sample types were omitted due to the potential for increased technical variability associated with the additional, type-specific handling and preparation steps, the evaluation of which was beyond the scope of this study. In contrast, blood derived sample types were included due to their widespread usage and established relevance in miRNA expression studies despite variability in processing. From each study, samples were considered eligible if described as healthy without experimental treatment. For example, data from the pre-challenge samples in a challenge study [95] were included, while data from horses described as “non-performers” in an exercise or endurance trial were excluded [56]. Similarly, if horses were not described as “non-performers,” both pre- and post-race samples were considered eligible [55]. Samples were excluded if health status was not described or could not be inferred from available metadata. Since it has been suggested that sex and breed-specific miRNA profiles may exist in the horse [58] and other species [96,97], samples lacking sex or breed information were also omitted. Mixed breeds were only included if at least one parental breed was described (i.e., horses labeled “cross” were excluded, whereas horses labeled “Thoroughbred-cross” were included). Samples described as male were classified as male-unknown unless expressly indicated as stallion or gelding. In the case of females, if age was reported, a distinction was made between mare (>3 years) and filly. Similarly, males

under 3 years were labelled colt. All sample data passing inclusion criteria were downloaded from the NCBI Sequence Read Archive (SRA) with *parallel-fastq-dump* v0.6.6 (<https://github.com/rvalieris/parallel-fastq-dump>) (a *fastq-dump* wrapper from the SRA toolkit v2.11.0 [98]) or directly from the European Nucleotide Archive (ENA). Study details and sample data for included *public* samples are reported in [S3 Table](#).

Small RNA processing and analysis

We developed the Framework for Analysis and Refinement of miRNAs (FARmiR; <https://github.com/jonahcullen/FARmiR>) pipeline to catalog miRNA expression in a reproducible, transparent, and adaptable manner, adhering to the principles of sustainable data analysis [92]. FARmiR is an open-source, Snakemake-based [92] containerized pipeline [93] developed for quality assessment, pre-processing, and miRNA profiling (Fig 11). Although designed for use with any animal species, we illustrate its application and inputs using the current equine reference genome EquCab3 [86].

Read quality assessment and pre-processing. Raw sequence reads from all three sources were first assessed for quality without any filtering using *fastp* v1.3.1 [99]. Adapters and low-quality reads (> 40% of bases with phred quality scores < 15 or at most 5 missing base calls) were removed. For SE samples, the initial approach employs *adapt_find* to identify and remove adapters [100]. If adapters were not detected, *DNApi* is used for adapter prediction followed by *cutadapt* [101,102]. Although not included here, adapter detection and removal with *fastp* is used for samples with PE layouts [99]. The adapter prediction functionality of *fastp* was not included for SE samples due to observed inconsistencies in adapter prediction across datasets. We instead implemented this sequential approach to improve robustness without requiring prior knowledge of the adapter sequence, which is often not readily available in the publication. If included, PE reads will be joined (no mismatches allowed and a minimum overlap of six bases required) [103]. By convention, miRNA profiling typically only considers SE-sequenced libraries (or the first read of PE layouts). We designed FARmiR to handle PE layouts as there is recent evidence that PE sequencing of small RNA-seq libraries may improve isomiR identification and accuracy [104]. Sequence error correction, a critically important consideration for isomiR discovery and tissue-specific quantification, was conducted with *miREC* [105], followed by length filtering (minimum length of 17) [99]. To remove likely non-miRNA molecules, cleaned reads were first aligned against the EquCab3 [86] non-coding reference (excluding miRNA sequences) from Ensembl (release 103). Unmapped reads were then aligned against non-coding RNA (excluding miRNA sequences) from the Rfam v14.8 database [106]. Both alignments use *bowtie* v1.3.1 [107] with zero mismatches permitted (“-n 0”), only one reportable alignment allowed (“-m 1”), and without alignment against the reverse-complement (“--norc”). Reportable alignments were limited to one as there are existing miRNA sequences that have significant overlap with other non-miRNA regions, such as piRNAs [108]. Reads unmapped against Rfam with lengths greater than 25 bases were removed (referred to as *post-filtered reads*) [99].

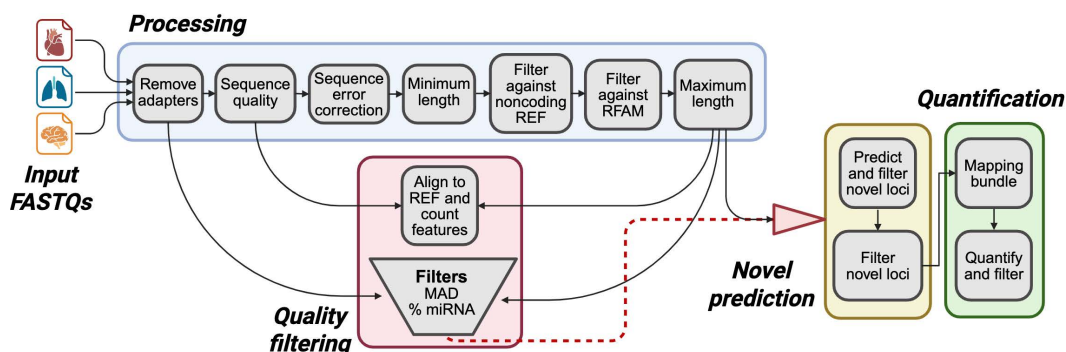


Fig 11. Overview of the FARmiR pipeline. Created with BioRender.com.

<https://doi.org/10.1371/journal.pgen.1011835.g011>

Pre- and post-filtered reads were aligned against EquCab3 using *bowtie* v.1.3.1 [107]. Default parameters were used except allowing for no mismatches (“-n 0”) within the first 8 bases (“-l 8”), up to 6 multiple alignments (“-m 6”) and reporting only the best stratum alignment (“-a --best --strata”). A maximum of 6 multiple alignments were permitted as there are at most 6 multimaps between the miRBase mature and hairpin sequences. We note however there exist up to 11 multimaps between human mature and hairpin sequences. Mapped reads and the Ensembl (release 103) EquCab3 annotation were input to *featureCounts* v2.0.1 [109], allowing for multi-mapped reads (“-M”) with per-exon counting (“-t exon”). MultiQC was used to aggregate pre-processing steps [110].

Sample filtering. Samples with \log_2 post-filtering counts less than three times the unbiased median absolute deviation were excluded [111]. Following this, samples with reads consisting of less than 10% miRNA biotypes (as estimated by *featureCounts*) were also discarded. Finally, only tissues with at least two independent samples (i.e., regardless of age, sex, or breed) were retained for quantification. This approach was taken to improve overall data quality and reduce the influence of batch effects.

Updated miRNA annotation. The current version of the equine miRNA annotation from miRBase is based on EquCab2 [85], and to the best of our knowledge, an EquCab3 annotation is not publicly available. To address this, we used a combination of *Liftoff* [112] and *LiftOver* [113] to map hairpins and mature sequence loci from EquCab2 to EquCab3. Of the 715 non-redundant miRNA hairpin genes (4 hairpins are encoded at more than one locus: *eca-mir-703* at 5 loci, *eca-mir-9128-2* at 3, and both *eca-mir-1842* and *eca-mir-9150* at 2 loci each), 7 hairpins were unmapped during conversion to EquCab3. All 7 unmapped hairpin loci (*eca-mir-8921-1*, *eca-mir-8921-2*, *eca-mir-8923-4*, *eca-mir-8923-5*, *eca-mir-8934-1*, *eca-mir-8935-6*, and *eca-mir-8946*) were predicted to be on unplaced EquCab2 contigs and are not included in either the Ensembl 103 or most recent 109 annotations. However, of the 690 mature miRNA sequences in EquCab2 (59 miRNAs are encoded by more than one hairpin), only one (*eca-miR-8946*) was not identified in EquCab3.

Discovery and quantification of miRNAs

Novel miRNA loci. Samples that passed filtering were used for novel miRNA prediction with *mirPro* v.1.1.4 [114]. *MirPro* uses *Novoalign* v3.02.12 [115] and a prediction algorithm similar to *miRDeep2* [116] with modifications to reduce false positives. Novel prediction was performed using default parameters with mature and hairpin miRNA sequences of human, mouse, dog, cow, and pig from miRBase v22 as related species [25]. miRBase v21 family information was used as miRBase v22 does not contain miRNA family information. There is no apparent difference between miRBase v21 and v22 families for equine miRNAs. *BLAST+* v2.13.0 [117] was used to query predicted hairpins against equine RNA sequences from RNAcentral [118] to remove other possible hairpin-forming RNAs (e.g., tRNA and rRNA). Genomic loci of predicted hairpins with a maximum overlap of 17 (based on a minimum mature sequence length of 17 as above) were merged with *BEDtools* [119]. Predicted hairpin sequences were then queried against the EquCab3 reference genome, and sequences with at most 15 alignments were retained. Predicted hairpin counts were then normalized using the post-filter counts from above to reads per million (RPM) and filtered to include only hairpins with RPMs ≥ 20 in at least 2 samples. In the cases of overlapping and merged loci, the hairpin with the highest total counts across all samples was retained. If total counts were equal between two or more hairpins, one was randomly selected to define the genomic locus. Expressed candidate loci that overlap (maximum of 17) with the updated miRNA annotation hairpin loci were removed. Retained expressed candidate loci were then used to select the associated mature representative. Novel miRNAs were named following a modified strategy previously established for sheep miRNA profiling [120]. Briefly, retained candidate loci were queried (*BLAST+* v2.13.0 with default settings and *-task blastn*) against the miRBase precursor sequences for cattle, sheep, pig, mouse, and human. Alignments with a q-value (minimum false discovery rate (FDR)) below 0.01 and sufficient query coverage (> 80%) were retained and the subject(s) with the highest percent identity was selected. For queries with valid alignments to more than one species, the evolutionarily closest species was chosen for naming. Note the original *mirPro* naming convention (i.e., *eca-novel-mir-NNNN*) was preserved for downstream analyses and within AIMEE to

avoid potential confusion. Orthologs of predicted miRNAs are noted where applicable and available in the supplemental data (S4 Table). The candidate hairpins and mature miRNAs were converted to BED format and merged with the updated miRNA annotation to be utilized for quantification.

Ref-miRs and isomiRs. Quantification was conducted with *isoMiRmap* [63]. *IsoMiRmap* requires species-specific “mapping bundles.” Currently, the developers only provide a human miRNA mapping bundle. To address this, we generated a species mapping bundle prior to quantification, starting with the putative miRNA annotation (see Loher et al. (2021) [63] for details and <https://github.com/jonahcullen/FARmiR> for implementation). The post-filtered and sized reads from each sample, along with the mapping bundle, were then quantified by *isoMiRmap*. This procedure returns (among other outputs) a Ref-miR/isomiR miRGFF3 annotation [121], a proposed standard from the miRNA TRanscriptomic Open Project (miRTOP), which utilizes a previously described naming convention [122] and nucleotide-based unique identifier per isomiR [123]. RPM (normalized by the final post-filtered and sized read counts) and count data were extracted from these annotations and assembled into an expression matrix. Ref-miR/isomiRs were retained only where at least 2 samples have RPMs greater than 20. Following this, samples consisting of less than 10% non-zero RPMs were excluded, and ultimately, tissues were dropped if less than 2 independent samples remained. While this filtering improves interpretability, it may impact detection of lowly expressed miRNAs, particularly in sample types with high compositional variability across RNA biotypes. Various metrics were then calculated for each Ref-miR (e.g., numbers of hairpins, isomiRs, and seed-shifted isomiRs).

Shiny app and browser. To promote accessibility and maximize insights derived from the collected data, we developed a Shiny web application built with the *golem* framework [124]. This app, Animal IsomiR and MiRNA Expression Explorer (AIMEE; <https://github.com/jonahcullen/AIMEE>), was designed to provide user-friendly accessibility and point-and-click functionality for data exploration and visualization, simple analyses, and exporting quantification data for subsequent investigations. AIMEE utilizes expression matrices generated by FARmiR and associated sample metadata as input for multiple visualization and analysis modules. These include simple quality control, read loss through FARmiR processing, isolating Ref-miRs/isomiRs of interest, and upset analysis with *ComplexUpset* [125]. Assessing inter- or intra-tissue Ref-miR/isomiR rank aggregation and tissue-specificity via the organ enrichment index (OEI) is also possible with AIMEE.

Robust rank aggregation. Unbiased integration of small RNA-seq data is critical when attempting to make inferences amidst unknown or unquantified sources of technical variation. One such approach to reduce these biases is to represent each sample as an independent prioritized miRNA gene list (i.e., miRNAs ranked by expression at a user-defined threshold). These gene lists can then be aggregated using order statistics, with *p*-values assigned to each miRNA denoting how its rank in each sample list surpasses what would be expected by chance. AIMEE implements the R package *RobustRankAggreg* (RRA) [67,68], a probabilistic aggregation model that is robust to the variations associated with high-throughput data from disparate origins, to assign significance probabilities to each miRNA from user-selected tissue or sample sets. As RRA operates on within-sample rankings as opposed to raw expression values, it helps reduce – but does not eliminate – the impact of technical variation across samples. A more traditional batch effect correction (e.g., ComBat from the *sva* R package [126]) was not feasible given the diverse and unbalanced nature of the included samples.

Organ-enrichment index. The OEI is a measurement of miRNA specificity [14,127], analogous to the mRNA tissue specificity index (TSI) ‘tau’ [128]. OEI values range from 0 to 1, with scores close to 0 indicating a miRNA that is expressed at similar levels across many or most organs (i.e., ‘housekeepers’), and scores close to 1 corresponding to miRNAs expressed in only one specific organ (i.e., organ-enriched). To calculate OEI, tissue types originating from the same organ were grouped together. Specifically, heart chambers were collapsed into “heart”, kidney regions into “kidney”, and skin samples into “skin”. Brain regions (parietal and temporal cortex, hippocampus, hypothalamus, amygdala, cerebellum and cerebellum vermis, and dorsal root ganglia) were combined into a single “brain” grouping. All other tissues were retained as separate groupings to preserve potential site-specific expression and enrichment. The mean RPM was

calculated for each miRNA per organ or combined organ grouping. Only miRNAs with a mean RPM > 20 in at least one organ or group were retained. Then, the OEI was calculated for each miRNA j as

$$oei_j = \frac{\sum_{i=1}^N (1 - x_{j,i})}{N - 1}$$

where $x_{j,i}$ represents miRNA j in organ i normalized by the maximum RPM across all organs, and N is the total number of organs considered. Importantly, due to the likely presence of batch effects, AIMEE contains OEI values calculated separately using the *primary* or *FAANG* collection.

Supporting information

S1 Text. The pdf file contains RNA extraction protocols based on biochemical composition (e.g., lipid-rich, protein-rich).

(PDF)

S1 Fig. Flow diagram of the search, screening, and selection of publicly available small RNA-seq datasets. Created with BioRender.com.

(TIFF)

S2 Fig. Counts (in millions) of filtered and passed reads following processing with FARmiR.

(TIFF)

S3 Fig. Mapping hairpin loci between EquCab2 (blue) and EquCab3 (yellow).

(TIFF)

S1 Table. Sample information and RNA-seq metadata for *primary* data.

(XLSX)

S2 Table. Sample information and RNA-seq metadata for *FAANG* data.

(XLSX)

S3 Table. Sample information and RNA-seq metadata for *public* data.

(XLSX)

S4 Table. Post-processed total, mean, and standard deviation of passed read counts by sex per tissue for the final expression set.

(XLSX)

S5 Table. Post-processed total, mean, and standard deviation of passed read counts by breed for the final expression set.

(XLSX)

S6 Table. Description of novel equine miRNAs.

(XLSX)

Acknowledgments

We would like to thank the following individuals for providing assistance preparing tissue samples and RNA extractions: Em Adam, Samantha Beeson, Kendall Blanchard, Lauren Hughes, Elaine Norton, Kelsey Springer, and Matthew Sturos. We would also like to thank Ernest F Bailey, Vinzenz Gerber, Vidhya Jagannathan, Tosso Leeb, Leslie A Lyons, and Cecilia M Penedo.

Author contributions

Conceptualization: Jakub Cieslak, James R Mickelson, Molly E McCue.

Data curation: Jonah N. Cullen.

Formal analysis: Jonah N. Cullen.

Funding acquisition: Jonah N. Cullen.

Investigation: Jonah N. Cullen.

Methodology: Jonah N. Cullen.

Project administration: Laura S Valderrama Figueroa, James R Mickelson, Molly E McCue.

Resources: Jakub Cieslak, Jessica L Petersen, Rebecca R Bellone, Carrie J Finno, Ted S Kalbfleisch, Kirstine Calloe, Stefano Capomaccio, Katia Cappelli, Stephen J Coleman, Ottmar Distl, Sian A Durward-Akhurst, Elena Giulotto, Natasha A Hamilton, Emmeline W Hill, Lisa M Katz, Dan A Klaerke, Gabriella Lindgren, David E MacHugh, Mariusz Mackowski, James N MacLeod, Julia Metzger, Barbara A Murphy, Ludovic Orlando, Terje Raudsepp, Maurizio Silvestrelli, Eric Strand, Teruaki Tozaki, Dagmar S Trachsel, Laura S Valderrama Figueroa, Brandon D Velie, Claire M Wade, Bianca Waud, James R Mickelson.

Software: Jonah N. Cullen.

Supervision: Molly E McCue.

Visualization: Jonah N. Cullen.

Writing – original draft: Jonah N. Cullen.

Writing – review & editing: Jakub Cieslak, Jessica L Petersen, Rebecca R Bellone, Carrie J Finno, Ted S Kalbfleisch, Kirstine Calloe, Stefano Capomaccio, Katia Cappelli, Stephen J Coleman, Ottmar Distl, Sian A Durward-Akhurst, Elena Giulotto, Natasha A Hamilton, Emmeline W Hill, Lisa M Katz, Dan A Klaerke, Gabriella Lindgren, David E MacHugh, Mariusz Mackowski, James N MacLeod, Julia Metzger, Barbara A Murphy, Ludovic Orlando, Terje Raudsepp, Maurizio Silvestrelli, Eric Strand, Teruaki Tozaki, Dagmar S Trachsel, Laura S Valderrama Figueroa, Brandon D Velie, Claire M Wade, Bianca Waud, James R Mickelson, Molly E McCue.

References

- Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet.* 2011;12(2):99–110. <https://doi.org/10.1038/nrg2936> PMID: [21245828](https://pubmed.ncbi.nlm.nih.gov/21245828/)
- Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 2009;19(1):92–105. <https://doi.org/10.1101/gr.082701.108> PMID: [18955434](https://pubmed.ncbi.nlm.nih.gov/18955434/)
- Fu G, Brkić J, Hayder H, Peng C. MicroRNAs in Human Placental Development and Pregnancy Complications. *Int J Mol Sci.* 2013;14(3):5519–44. <https://doi.org/10.3390/ijms14035519> PMID: [23528856](https://pubmed.ncbi.nlm.nih.gov/23528856/)
- Li N, Long B, Han W, Yuan S, Wang K. microRNAs: important regulators of stem cells. *Stem Cell Res Ther.* 2017;8(1):110. <https://doi.org/10.1186/s13287-017-0551-0> PMID: [28494789](https://pubmed.ncbi.nlm.nih.gov/28494789/)
- Rottiers V, Näär AM. MicroRNAs in metabolism and metabolic disorders. *Nat Rev Mol Cell Biol.* 2012;13(4):239–50. <https://doi.org/10.1038/nrm3313> PMID: [22436747](https://pubmed.ncbi.nlm.nih.gov/22436747/)
- Hwang H-W, Mendell JT. MicroRNAs in cell proliferation, cell death, and tumorigenesis. *Br J Cancer.* 2006;94(6):776–80. <https://doi.org/10.1038/sj.bjc.6603023> PMID: [16495913](https://pubmed.ncbi.nlm.nih.gov/16495913/)
- Bayraktar R, Van Roosbroeck K, Calin GA. Cell-to-cell communication: microRNAs as hormones. *Mol Oncol.* 2017;11(12):1673–86. <https://doi.org/10.1002/1878-0261.12144> PMID: [29024380](https://pubmed.ncbi.nlm.nih.gov/29024380/)
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature.* 2005;435(7043):834–8. <https://doi.org/10.1038/nature03702> PMID: [15944708](https://pubmed.ncbi.nlm.nih.gov/15944708/)
- Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet.* 2009;10(10):704–14. <https://doi.org/10.1038/nrg2634> PMID: [19763153](https://pubmed.ncbi.nlm.nih.gov/19763153/)

10. Im H-I, Kenny PJ. MicroRNAs in neuronal function and dysfunction. *Trends Neurosci.* 2012;35(5):325–34. <https://doi.org/10.1016/j.tins.2012.01.004> PMID: [22436491](https://pubmed.ncbi.nlm.nih.gov/22436491/)
11. Quinlan S, Kenny A, Medina M, Engel T, Jimenez-Mateos EM. MicroRNAs in Neurodegenerative Diseases. *Int Rev Cell Mol Biol.* 2017;334:309–43. <https://doi.org/10.1016/bs.ircmb.2017.04.002> PMID: [28838542](https://pubmed.ncbi.nlm.nih.gov/28838542/)
12. Wojciechowska A, Braniewska A, Kozar-Kamińska K. MicroRNA in cardiovascular biology and disease. *Adv Clin Exp Med.* 2017;26(5):865–74. <https://doi.org/10.17219/acem/62915> PMID: [29068585](https://pubmed.ncbi.nlm.nih.gov/29068585/)
13. de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol.* 2017;35(9):872–8. <https://doi.org/10.1038/nbt.3947> PMID: [28829439](https://pubmed.ncbi.nlm.nih.gov/28829439/)
14. Desvignes T, Batzel P, Sydes J, Eames BF, Postlethwait JH. miRNA analysis with Prost! reveals evolutionary conservation of organ-enriched expression and post-transcriptional modifications in three-spined stickleback and zebrafish. *Sci Rep.* 2019;9(1):3913. <https://doi.org/10.1038/s41598-019-40361-8> PMID: [30850632](https://pubmed.ncbi.nlm.nih.gov/30850632/)
15. Keller A, Gröger L, Tschernig T, Solomon J, Laham O, Schaum N, et al. miRNATissueAtlas2: an update to the human miRNA tissue atlas. *Nucleic Acids Res.* 2022;50(D1):D211–21. <https://doi.org/10.1093/nar/gkab808> PMID: [34570238](https://pubmed.ncbi.nlm.nih.gov/34570238/)
16. Yao X, Sun S, Zi Y, Liu Y, Yang J, Ren L, et al. Comprehensive microRNA-seq transcriptomic profiling across 11 organs, 4 ages, and 2 sexes of Fischer 344 rats. *Sci Data.* 2022;9(1):201. <https://doi.org/10.1038/s41597-022-01285-7> PMID: [35551205](https://pubmed.ncbi.nlm.nih.gov/35551205/)
17. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-L, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 2008;18(4):610–21. <https://doi.org/10.1101/gr.7179508> PMID: [18285502](https://pubmed.ncbi.nlm.nih.gov/18285502/)
18. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science.* 2007;315(5815):1137–40. <https://doi.org/10.1126/science.1138050> PMID: [17322061](https://pubmed.ncbi.nlm.nih.gov/17322061/)
19. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs—the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* 2012;28(11):544–9. <https://doi.org/10.1016/j.tig.2012.07.005> PMID: [22883467](https://pubmed.ncbi.nlm.nih.gov/22883467/)
20. Menezes MR, Balzeau J, Hagan JP. 3' RNA Uridylation in Epitranscriptomics, Gene Regulation, and Disease. *Front Mol Biosci.* 2018;5:61. <https://doi.org/10.3389/fmolb.2018.00061> PMID: [30057901](https://pubmed.ncbi.nlm.nih.gov/30057901/)
21. Glogovitis I, Yahubyan G, Würdinger T, Koppers-Lalic D, Baev V. isomiRs—Hidden Soldiers in the miRNA Regulatory Army, and How to Find Them?. *Biomolecules.* 2020;11(1):41. <https://doi.org/10.3390/biom11010041> PMID: [33396892](https://pubmed.ncbi.nlm.nih.gov/33396892/)
22. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell.* 2007;129(7):1401–14. <https://doi.org/10.1016/j.cell.2007.04.040> PMID: [17604727](https://pubmed.ncbi.nlm.nih.gov/17604727/)
23. Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, et al. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annu Rev Genet.* 2015;49:213–42. <https://doi.org/10.1146/annurev-genet-120213-092023> PMID: [26473382](https://pubmed.ncbi.nlm.nih.gov/26473382/)
24. Desvignes T, Batzel P, Berezikov E, Eilbeck K, Eppig JT, McAndrews MS, et al. miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. *Trends Genet.* 2015;31(11):613–26. <https://doi.org/10.1016/j.tig.2015.09.002> PMID: [26453491](https://pubmed.ncbi.nlm.nih.gov/26453491/)
25. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47(D1):D155–62. <https://doi.org/10.1093/nar/gky1141> PMID: [30423142](https://pubmed.ncbi.nlm.nih.gov/30423142/)
26. Fromm B, Høye E, Domanska D, Zhong X, Aparicio-Puerta E, Ovchinnikov V, et al. MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res.* 2022;50(D1):D204–10. <https://doi.org/10.1093/nar/gkab1101> PMID: [34850127](https://pubmed.ncbi.nlm.nih.gov/34850127/)
27. Wu CW, Evans JM, Huang S, Mahoney DW, Dukek BA, Taylor WR, et al. A Comprehensive Approach to Sequence-oriented IsomiR annotation (CASMIIR): demonstration with IsomiR profiling in colorectal neoplasia. *BMC Genomics.* 2018;19(1):401. <https://doi.org/10.1186/s12864-018-4794-7> PMID: [29801434](https://pubmed.ncbi.nlm.nih.gov/29801434/)
28. Tomasello L, Distefano R, Nigita G, Croce CM. The MicroRNA Family Gets Wider: The IsomiRs Classification and Role. *Front Cell Dev Biol.* 2021;9:668648. <https://doi.org/10.3389/fcell.2021.668648> PMID: [34178993](https://pubmed.ncbi.nlm.nih.gov/34178993/)
29. Bartel DP. Metazoan MicroRNAs. *Cell.* 2018;173(1):20–51. <https://doi.org/10.1016/j.cell.2018.03.006> PMID: [29570994](https://pubmed.ncbi.nlm.nih.gov/29570994/)
30. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009;136(2):215–33. <https://doi.org/10.1016/j.cell.2009.01.002> PMID: [19167326](https://pubmed.ncbi.nlm.nih.gov/19167326/)
31. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* 2011;12(12):R126. <https://doi.org/10.1186/gb-2011-12-12-r126> PMID: [22208850](https://pubmed.ncbi.nlm.nih.gov/22208850/)
32. Tan GC, Chan E, Molnar A, Sarkar R, Alexieva D, Isa IM, et al. 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* 2014;42(14):9424–35. <https://doi.org/10.1093/nar/gku656> PMID: [25056318](https://pubmed.ncbi.nlm.nih.gov/25056318/)
33. Telonis AG, Loher P, Jing Y, Londin E, Rigoutsos I. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res.* 2015;43(19):9158–75. <https://doi.org/10.1093/nar/gkv922> PMID: [26400174](https://pubmed.ncbi.nlm.nih.gov/26400174/)
34. Salem O, Erdem N, Jung J, Münstermann E, Wörner A, Wilhelm H, et al. The highly expressed 5'isomiR of hsa-miR-140-3p contributes to the tumor-suppressive effects of miR-140 by reducing breast cancer proliferation and migration. *BMC Genomics.* 2016;17:566. <https://doi.org/10.1186/s12864-016-2869-x> PMID: [27502506](https://pubmed.ncbi.nlm.nih.gov/27502506/)

35. van der Kwast RVCT, Woudenberg T, Quax PHA, Nossent AY. MicroRNA-411 and Its 5'-IsomiR Have Distinct Targets and Functions and Are Differentially Regulated in the Vasculature under Ischemia. *Mol Ther*. 2020;28(1):157–70. <https://doi.org/10.1016/j.ymthe.2019.10.002> PMID: [31636041](https://pubmed.ncbi.nlm.nih.gov/31636041/)
36. Manzano M, Forte E, Raja AN, Schipma MJ, Gottwein E. Divergent target recognition by coexpressed 5'-isomiRs of miR-142-3p and selective viral mimicry. *RNA*. 2015;21(9):1606–20. <https://doi.org/10.1261/ma.048876.114> PMID: [26137849](https://pubmed.ncbi.nlm.nih.gov/26137849/)
37. Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc Natl Acad Sci U S A*. 2015;112(10):E1106–15. <https://doi.org/10.1073/pnas.1420955112> PMID: [25713380](https://pubmed.ncbi.nlm.nih.gov/25713380/)
38. Aparicio-Puerta E, Hirsch P, Schmartz GP, Fehlmann T, Keller V, Engel A, et al. isomiRdb: microRNA expression at isoform resolution. *Nucleic Acids Res*. 2023;51(D1):D179–85. <https://doi.org/10.1093/nar/gkac884> PMID: [36243964](https://pubmed.ncbi.nlm.nih.gov/36243964/)
39. Wang S, Xu Y, Li M, Tu J, Lu Z. Dysregulation of miRNA isoform level at 5' end in Alzheimer's disease. *Gene*. 2016;584(2):167–72. <https://doi.org/10.1016/j.gene.2016.02.020> PMID: [26899870](https://pubmed.ncbi.nlm.nih.gov/26899870/)
40. Telonis AG, Magee R, Loher P, Chervoneva I, Londin E, Rigoutsos I. Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res*. 2017;45(6):2973–85. <https://doi.org/10.1093/nar/gkx082> PMID: [28206648](https://pubmed.ncbi.nlm.nih.gov/28206648/)
41. Ma M, Yin Z, Zhong H, Liang T, Guo L. Analysis of the expression, function, and evolution of miR-27 isoforms and their responses in metabolic processes. *Genomics*. 2019;111(6):1249–57. <https://doi.org/10.1016/j.ygeno.2018.08.004> PMID: [30145283](https://pubmed.ncbi.nlm.nih.gov/30145283/)
42. Schmartz GP, Kern F, Fehlmann T, Wagner V, Fromm B, Keller A. Encyclopedia of tools for the analysis of miRNA isoforms. *Brief Bioinform*. 2021;22(4):bbaa346. <https://doi.org/10.1093/bib/bbaa346> PMID: [33313643](https://pubmed.ncbi.nlm.nih.gov/33313643/)
43. Schmauch E, Laitinen P, Turunen TA, Väänänen M-A, Malm T, Kellis M, et al. isomiRs-specific differential expression is the rule, not the exception: Are we missing hundreds of species in microRNA analysis?. *Cold Spring Harbor Laboratory*. 2021. <https://doi.org/10.1101/2021.12.15.472814>
44. Boffill-De Ros X, Luke B, Guthridge R, Mudunuri U, Loss M, Gu S. Tumor IsomiR Encyclopedia (TIE): a pan-cancer database of miRNA isoforms. *Bioinformatics*. 2021;37(18):3023–5. <https://doi.org/10.1093/bioinformatics/btab172> PMID: [33729437](https://pubmed.ncbi.nlm.nih.gov/33729437/)
45. Zhang Y, Zang Q, Xu B, Zheng W, Ban R, Zhang H, et al. IsomiR Bank: a research resource for tracking IsomiRs. *Bioinformatics*. 2016;32(13):2069–71. <https://doi.org/10.1093/bioinformatics/btw070> PMID: [27153728](https://pubmed.ncbi.nlm.nih.gov/27153728/)
46. Zhou M, Wang Q, Sun J, Li X, Xu L, Yang H, et al. In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. *Genomics*. 2009;94(2):125–31. <https://doi.org/10.1016/j.ygeno.2009.04.006> PMID: [19406225](https://pubmed.ncbi.nlm.nih.gov/19406225/)
47. Platt RN II, Vandewege MW, Kern C, Schmidt CJ, Hoffmann FG, Ray DA. Large Numbers of Novel miRNAs Originate from DNA Transposons and Are Coincident with a Large Species Radiation in Bats. *Molecular Biology and Evolution*. 2014;31(6):1536–45. <https://doi.org/10.1093/molbev/msu112>
48. Clarke AW, Høye E, Hembrom AA, Paynter VM, Vinther J, Wyrożemski Ł, et al. MirGeneDB 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microRNA families and updated covariance models. *Nucleic Acids Res*. 2025;53(D1):D116–28. <https://doi.org/10.1093/nar/gkae1094> PMID: [39673268](https://pubmed.ncbi.nlm.nih.gov/39673268/)
49. Desjardin C, Vaiman A, Mata X, Legendre R, Laubier J, Kennedy SP, et al. Next-generation sequencing identifies equine cartilage and subchondral bone miRNAs and suggests their involvement in osteochondrosis physiopathology. *BMC Genomics*. 2014;15(1):798. <https://doi.org/10.1186/1471-2164-15-798> PMID: [25227120](https://pubmed.ncbi.nlm.nih.gov/25227120/)
50. Li B, He X, Zhao Y, Bai D, Li D, Zhou Z, et al. Analysis of the miRNA transcriptome during testicular development and spermatogenesis of the Mongolian horse. *Reprod Fertil Dev*. 2020;32(6):582–93. <https://doi.org/10.1071/RD19133> PMID: [32209208](https://pubmed.ncbi.nlm.nih.gov/32209208/)
51. Smits K, Gansemans Y, Tilleman L, Van Nieuwerburgh F, Van De Velde M, Gerits I, et al. Maternal Recognition of Pregnancy in the Horse: Are MicroRNAs the Secret Messengers?. *Int J Mol Sci*. 2020;21(2):419. <https://doi.org/10.3390/ijms21020419> PMID: [31936511](https://pubmed.ncbi.nlm.nih.gov/31936511/)
52. Bao T, Han H, Li B, Zhao Y, Bou G, Zhang X, et al. The distinct transcriptomes of fast-twitch and slow-twitch muscles in Mongolian horses. *Comp Biochem Physiol Part D Genomics Proteomics*. 2020;33:100649. <https://doi.org/10.1016/j.cbd.2019.100649> PMID: [31869634](https://pubmed.ncbi.nlm.nih.gov/31869634/)
53. Anderson JR, Jacobsen S, Walters M, Bundgaard L, Diendorfer A, Hackl M, et al. Small non-coding RNA landscape of extracellular vesicles from a post-traumatic model of equine osteoarthritis. *Front Vet Sci*. 2022;9:901269. <https://doi.org/10.3389/fvets.2022.901269> PMID: [36003409](https://pubmed.ncbi.nlm.nih.gov/36003409/)
54. Mach N, Plancade S, Pacholewska A, Lecardonnell J, Rivière J, Moroldo M, et al. Integrated mRNA and miRNA expression profiling in blood reveals candidate biomarkers associated with endurance exercise in the horse. *Sci Rep*. 2016;6:22932. <https://doi.org/10.1038/srep22932> PMID: [26960911](https://pubmed.ncbi.nlm.nih.gov/26960911/)
55. Kim H-A, Kim M-C, Kim N-Y, Ryu D-Y, Lee H-S, Kim Y. Integrated analysis of microRNA and mRNA expressions in peripheral blood leukocytes of Warmblood horses before and after exercise. *J Vet Sci*. 2018;19(1):99–106. <https://doi.org/10.4142/jvs.2018.19.1.99> PMID: [28927254](https://pubmed.ncbi.nlm.nih.gov/28927254/)
56. Cappelli K, Mecocci S, Capomaccio S, Beccati F, Palumbo AR, Tognoloni A, et al. Circulating Transcriptional Profile Modulation in Response to Metabolic Unbalance Due to Long-Term Exercise in Equine Athletes: A Pilot Study. *Genes (Basel)*. 2021;12(12):1965. <https://doi.org/10.3390/genes12121965> PMID: [34946914](https://pubmed.ncbi.nlm.nih.gov/34946914/)
57. Kim M-C, Lee S-W, Ryu D-Y, Cui F-J, Bhak J, Kim Y. Identification and characterization of microRNAs in normal equine tissues by Next Generation Sequencing. *PLoS One*. 2014;9(4):e93662. <https://doi.org/10.1371/journal.pone.0093662> PMID: [24695583](https://pubmed.ncbi.nlm.nih.gov/24695583/)

58. Pacholewska A, Mach N, Mata X, Vaiman A, Schibler L, Barrey E, et al. Novel equine tissue miRNAs and breed-related miRNA expressed in serum. *BMC Genomics*. 2016;17(1):831. <https://doi.org/10.1186/s12864-016-3168-2> PMID: [27782799](https://pubmed.ncbi.nlm.nih.gov/27782799/)
59. Pawlina K, Gurgul A, Szmatola T, Koch C, Mählmann K, Witkowski M, et al. Comprehensive characteristics of microRNA expression profile of equine sarcoids. *Biochimie*. 2017;137:20–8. <https://doi.org/10.1016/j.biochi.2017.02.017> PMID: [28259757](https://pubmed.ncbi.nlm.nih.gov/28259757/)
60. Cowled C, Foo C-H, Deffrasnes C, Rootes CL, Williams DT, Middleton D, et al. Circulating microRNA profiles of Hendra virus infection in horses. *Sci Rep*. 2017;7(1):7431. <https://doi.org/10.1038/s41598-017-06939-w> PMID: [28785041](https://pubmed.ncbi.nlm.nih.gov/28785041/)
61. Burns EN, Bordbari MH, Mienaltowski MJ, Affolter VK, Barro MV, Gianino F, et al. Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project. *Anim Genet*. 2018;49(6):564–70. <https://doi.org/10.1111/age.12717> PMID: [30311254](https://pubmed.ncbi.nlm.nih.gov/30311254/)
62. Donnelly CG, Bellone RR, Hales EN, Nguyen A, Katzman SA, Dujovne GA, et al. Generation of a Biobank From Two Adult Thoroughbred Stallions for the Functional Annotation of Animal Genomes Initiative. *Front Genet*. 2021;12:650305. <https://doi.org/10.3389/fgene.2021.650305> PMID: [33763124](https://pubmed.ncbi.nlm.nih.gov/33763124/)
63. Loher P, Karathanasis N, Londin E, F Bray P, Pliatsika V, Telonis AG, et al. IsoMiRmap: fast, deterministic and exhaustive mining of isomiRs from short RNA-seq datasets. *Bioinformatics*. 2021;37(13):1828–38. <https://doi.org/10.1093/bioinformatics/btab016> PMID: [33471076](https://pubmed.ncbi.nlm.nih.gov/33471076/)
64. Kawcak CE, McIlwraith CW, Norrdin RW, Park RD, James SP. The role of subchondral bone in joint disease: a review. *Equine Vet J*. 2001;33(2):120–6. <https://doi.org/10.1111/j.2042-3306.2001.tb00589.x> PMID: [11266060](https://pubmed.ncbi.nlm.nih.gov/11266060/)
65. Baker ME, Lee S, Clinton M, Hackl M, Castanheira C, Peffers MJ, et al. Investigation of MicroRNA Biomarkers in Equine Distal Interphalangeal Joint Osteoarthritis. *Int J Mol Sci*. 2022;23(24):15526. <https://doi.org/10.3390/ijms232415526> PMID: [36555166](https://pubmed.ncbi.nlm.nih.gov/36555166/)
66. Castanheira C, Balaskas P, Falls C, Ashraf-Kharaz Y, Clegg P, Burke K, et al. Equine synovial fluid small non-coding RNA signatures in early osteoarthritis. *BMC Vet Res*. 2021;17(1):26. <https://doi.org/10.1186/s12917-020-02707-7> PMID: [33422071](https://pubmed.ncbi.nlm.nih.gov/33422071/)
67. Kolde R, Laur S. RobustRankAggreg: Methods for Robust Rank Aggregation. <https://cran.r-project.org/web/packages/RobustRankAggreg/index.html>. 2022.
68. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28(4):573–80. <https://doi.org/10.1093/bioinformatics/btr709> PMID: [22247279](https://pubmed.ncbi.nlm.nih.gov/22247279/)
69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
70. Li D, Zand MS, Dye TD, Goniewicz ML, Rahman I, Xie Z. An evaluation of RNA-seq differential analysis methods. *PLoS One*. 2022;17(9):e0264246. <https://doi.org/10.1371/journal.pone.0264246> PMID: [36112652](https://pubmed.ncbi.nlm.nih.gov/36112652/)
71. Ibing S, Michels BE, Mosdzien M, Meyer HR, Feuerbach L, Körner C. On the impact of batch effect correction in TCGA isomiR expression data. *NAR Cancer*. 2021;3(1):zcab007. <https://doi.org/10.1093/narcan/zcab007> PMID: [34316700](https://pubmed.ncbi.nlm.nih.gov/34316700/)
72. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, et al. Unifying cancer and normal RNA sequencing data from different sources. *Sci Data*. 2018;5:180061. <https://doi.org/10.1038/sdata.2018.61> PMID: [29664468](https://pubmed.ncbi.nlm.nih.gov/29664468/)
73. Ahmed K, LaPierre MP, Gasser E, Denzler R, Yang Y, Rüllicke T, et al. Loss of microRNA-7a2 induces hypogonadotropic hypogonadism and infertility. *J Clin Invest*. 2017;127(3):1061–74. <https://doi.org/10.1172/JCI90031> PMID: [28218624](https://pubmed.ncbi.nlm.nih.gov/28218624/)
74. Wang C-J, Gao F, Huang Y-J, Han D-X, Zheng Y, Wang W-H, et al. circAkap17b acts as a miR-7 family molecular sponge to regulate FSH secretion in rat pituitary cells. *J Mol Endocrinol*. 2020;65(4):135–48. <https://doi.org/10.1530/JME-20-0036> PMID: [33048061](https://pubmed.ncbi.nlm.nih.gov/33048061/)
75. Yuan B, Sun GJ, Zhang GL, Wu J, Xu C, Dai LS, et al. Identification of target genes for adenohipophysis-prefer miR-7 and miR-375 in cattle. *Genet Mol Res*. 2015;14(3):9753–63. <https://doi.org/10.4238/2015.August.19.8> PMID: [26345908](https://pubmed.ncbi.nlm.nih.gov/26345908/)
76. Zhang Y, Huang F, Xu N, Wang J, Li D, Yin L. Overexpression of serum extracellular vesicle microRNA-215-5p is associated with early tumor recurrence and poor prognosis of gastric cancer. *Clinics (Sao Paulo)*. 2021;76:e2081. <https://doi.org/10.6061/clinics/2021/e2081> PMID: [33978071](https://pubmed.ncbi.nlm.nih.gov/33978071/)
77. Vychytilova-Faltejskova P, Merhautova J, Machackova T, Gutierrez-Garcia I, Garcia-Solano J, Radova L, et al. MiR-215-5p is a tumor suppressor in colorectal cancer targeting EGFR ligand ephremin and its transcriptional inducer HOXB9. *Oncogenesis*. 2017;6(11):399. <https://doi.org/10.1038/s41389-017-0006-6> PMID: [29199273](https://pubmed.ncbi.nlm.nih.gov/29199273/)
78. Sun B, Xing K, Qi C, Yan K, Xu Y. Down-regulation of miR-215 attenuates lipopolysaccharide-induced inflammatory injury in CCD-18co cells by targeting GDF11 through the TLR4/NF- κ B and JNK/p38 signaling pathways. *Histol Histopathol*. 2020;35(12):1473–81. <https://doi.org/10.14670/HH-18-278> PMID: [33146403](https://pubmed.ncbi.nlm.nih.gov/33146403/)
79. Antunes J, Salcedo-Jiménez R, Lively S, Potla P, Coté N, Dubois M-S, et al. microRNAs are differentially expressed in equine plasma of horses with osteoarthritis and osteochondritis dissecans versus control horses. *PLoS One*. 2024;19(2):e0297303. <https://doi.org/10.1371/journal.pone.0297303> PMID: [38394252](https://pubmed.ncbi.nlm.nih.gov/38394252/)
80. Deilluis JA. MicroRNAs as regulators of metabolic disease: pathophysiologic significance and emerging role as biomarkers and therapeutics. *Int J Obes (Lond)*. 2016;40(1):88–101. <https://doi.org/10.1038/ijo.2015.170> PMID: [26311337](https://pubmed.ncbi.nlm.nih.gov/26311337/)
81. Han H, Chen Q, Gao Y, Li J, Li W, Dang R, et al. Comparative Transcriptomics Analysis of Testicular miRNA from Cryptorchid and Normal Horses. *Animals (Basel)*. 2020;10(2):338. <https://doi.org/10.3390/ani10020338> PMID: [32098036](https://pubmed.ncbi.nlm.nih.gov/32098036/)
82. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)

83. Li Y, Zhang Z, Liu F, Vongsangnak W, Jing Q, Shen B. Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. *Nucleic Acids Res.* 2012;40(10):4298–305. <https://doi.org/10.1093/nar/gks043> PMID: [22287634](https://pubmed.ncbi.nlm.nih.gov/22287634/)
84. Backes C, Fehlmann T, Kern F, Kehl T, Lenhof H-P, Meese E, et al. miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Res.* 2018;46(D1):D160–7. <https://doi.org/10.1093/nar/gkx851> PMID: [29036653](https://pubmed.ncbi.nlm.nih.gov/29036653/)
85. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science.* 2009;326(5954):865–7. <https://doi.org/10.1126/science.1178158> PMID: [19892987](https://pubmed.ncbi.nlm.nih.gov/19892987/)
86. Kalbfleisch TS, Rice ES, DePriest MS Jr, Walenz BP, Hestand MS, Vermeesch JR, et al. EquCab3, an Updated Reference Genome for the Domestic Horse. Cold Spring Harbor Laboratory. 2018. <https://doi.org/10.1101/306928>
87. Fridrich A, Hazan Y, Moran Y. Too Many False Targets for MicroRNAs: Challenges and Pitfalls in Prediction of miRNA Targets and Their Gene Ontology in Model and Non-model Organisms. *Bioessays.* 2019;41(4):e1800169. <https://doi.org/10.1002/bies.201800169> PMID: [30919506](https://pubmed.ncbi.nlm.nih.gov/30919506/)
88. Smith CM, Hutvagner G. A comparative analysis of single cell small RNA sequencing data reveals heterogeneous isomiR expression and regulation. *Sci Rep.* 2022;12(1):2834. <https://doi.org/10.1038/s41598-022-06876-3> PMID: [35181712](https://pubmed.ncbi.nlm.nih.gov/35181712/)
89. McCall MN, Kim M-S, Adil M, Patil AH, Lu Y, Mitchell CJ, et al. Toward the human cellular microRNAome. *Genome Res.* 2017;27(10):1769–81. <https://doi.org/10.1101/gr.222067.117> PMID: [28877962](https://pubmed.ncbi.nlm.nih.gov/28877962/)
90. Hückler SM, Fehlmann T, Werno C, Weidele K, Lücke F, Schlenka-Lange A, et al. Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat Commun.* 2021;12(1):4316. <https://doi.org/10.1038/s41467-021-24611-w> PMID: [34262050](https://pubmed.ncbi.nlm.nih.gov/34262050/)
91. Hagemann-Jensen M, Abdullayev I, Sandberg R, Faridani OR. Small-seq for single-cell small-RNA sequencing. *Nat Protoc.* 2018;13(10):2407–24. <https://doi.org/10.1038/s41596-018-0049-y> PMID: [30250291](https://pubmed.ncbi.nlm.nih.gov/30250291/)
92. Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable data analysis with Snakemake. *F1000Res.* 2021;10:33. <https://doi.org/10.12688/f1000research.29032.2> PMID: [34035898](https://pubmed.ncbi.nlm.nih.gov/34035898/)
93. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. *PLoS One.* 2017;12(5):e0177459. <https://doi.org/10.1371/journal.pone.0177459> PMID: [28494014](https://pubmed.ncbi.nlm.nih.gov/28494014/)
94. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet.* 2010;11(10):733–9. <https://doi.org/10.1038/nrg2825> PMID: [20838408](https://pubmed.ncbi.nlm.nih.gov/20838408/)
95. Zarski LM, Weber PSD, Lee Y, Soboll Hussey G. Transcriptomic Profiling of Equine and Viral Genes in Peripheral Blood Mononuclear Cells in Horses during Equine Herpesvirus 1 Infection. *Pathogens.* 2021;10(1):43. <https://doi.org/10.3390/pathogens10010043> PMID: [33430330](https://pubmed.ncbi.nlm.nih.gov/33430330/)
96. Mukiibi R, Johnston D, Vinsky M, Fitzsimmons C, Stothard P, Waters SM, et al. Bovine hepatic miRNAome profiling and differential miRNA expression analyses between beef steers with divergent feed efficiency phenotypes. *Sci Rep.* 2020;10(1):19309. <https://doi.org/10.1038/s41598-020-73885-5> PMID: [33168877](https://pubmed.ncbi.nlm.nih.gov/33168877/)
97. Pawlina-Tyszko K, Oczkowitz M, Gurgul A, Szmatoła T, Bugno-Poniewierska M. MicroRNA profiling of the pig periaqueductal grey (PAG) region reveals candidates potentially related to sex-dependent differences. *Biol Sex Differ.* 2020;11(1):67. <https://doi.org/10.1186/s13293-020-00343-2> PMID: [33451362](https://pubmed.ncbi.nlm.nih.gov/33451362/)
98. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* 2011;39(Database issue):D19–21. <https://doi.org/10.1093/nar/gkq1019> PMID: [21062823](https://pubmed.ncbi.nlm.nih.gov/21062823/)
99. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560> PMID: [30423086](https://pubmed.ncbi.nlm.nih.gov/30423086/)
100. Mohideen AMSH, Johansen SD, Babiak I. High-Throughput Identification of Adapters in Single-Read Sequencing Data. *Biomolecules.* 2020;10(6):878. <https://doi.org/10.3390/biom10060878> PMID: [32521604](https://pubmed.ncbi.nlm.nih.gov/32521604/)
101. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 2011;17(1):10. <https://doi.org/10.14806/ej.17.1.200>
102. Tsuji J, Weng Z. DNApi: A De Novo Adapter Prediction Algorithm for Small RNA Sequencing Data. *PLoS One.* 2016;11(10):e0164228. <https://doi.org/10.1371/journal.pone.0164228> PMID: [27736901](https://pubmed.ncbi.nlm.nih.gov/27736901/)
103. Aronesty E. Comparison of Sequencing Utility Programs. *TOBIOJ.* 2013;7(1):1–8. <https://doi.org/10.2174/1875036201307010001>
104. Sanchez Herrero JF, Pluvinet R, Luna de Haro A, Sumoy L. Paired-end small RNA sequencing reveals a possible overestimation in the isomiR sequence repertoire previously reported from conventional single read data analysis. *BMC Bioinformatics.* 2021;22(1):215. <https://doi.org/10.1186/s12859-021-04128-1> PMID: [33902448](https://pubmed.ncbi.nlm.nih.gov/33902448/)
105. Zhang X, Ping P, Hutvagner G, Blumenstein M, Li J. Aberration-corrected ultrafine analysis of miRNA reads at single-base resolution: a k-mer lattice approach. *Nucleic Acids Res.* 2021;49(18):e106. <https://doi.org/10.1093/nar/gkab610> PMID: [34291293](https://pubmed.ncbi.nlm.nih.gov/34291293/)
106. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, et al. Non-Coding RNA Analysis Using the Rfam Database. *Curr Protoc Bioinformatics.* 2018;62(1):e51. <https://doi.org/10.1002/cpbi.51> PMID: [29927072](https://pubmed.ncbi.nlm.nih.gov/29927072/)
107. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/)

108. Li J, Kho AT, Chase RP, Pantano L, Farnam L, Amr SS, et al. COMPSRA: a COMprehensive Platform for Small RNA-Seq data Analysis. *Sci Rep*. 2020;10(1):4552. <https://doi.org/10.1038/s41598-020-61495-0> PMID: [32165660](https://pubmed.ncbi.nlm.nih.gov/32165660/)
109. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30. <https://doi.org/10.1093/bioinformatics/btt656> PMID: [24227677](https://pubmed.ncbi.nlm.nih.gov/24227677/)
110. Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8. <https://doi.org/10.1093/bioinformatics/btw354> PMID: [27312411](https://pubmed.ncbi.nlm.nih.gov/27312411/)
111. Park C, Kim H, Wang M. Investigation of finite-sample properties of robust location and scale estimators. *Communications in Statistics - Simulation and Computation*. 2020;51(5):2619–45. <https://doi.org/10.1080/03610918.2019.1699114>
112. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021;37(12):1639–43. <https://doi.org/10.1093/bioinformatics/btaa1016> PMID: [33320174](https://pubmed.ncbi.nlm.nih.gov/33320174/)
113. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34(Database issue):D590–8. <https://doi.org/10.1093/nar/gkj144> PMID: [16381938](https://pubmed.ncbi.nlm.nih.gov/16381938/)
114. Shi J, Dong M, Li L, Liu L, Luz-Madrigal A, Tsonis PA, et al. mirPRO—a novel standalone program for differential expression and variation analysis of miRNAs. *Sci Rep*. 2015;5:14617. <https://doi.org/10.1038/srep14617> PMID: [26434581](https://pubmed.ncbi.nlm.nih.gov/26434581/)
115. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform*. 2010;11(5):473–83. <https://doi.org/10.1093/bib/bbq015> PMID: [20460430](https://pubmed.ncbi.nlm.nih.gov/20460430/)
116. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40(1):37–52. <https://doi.org/10.1093/nar/gkr688> PMID: [21911355](https://pubmed.ncbi.nlm.nih.gov/21911355/)
117. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421> PMID: [20003500](https://pubmed.ncbi.nlm.nih.gov/20003500/)
118. RNAcentral Consortium. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res*. 2021;49(D1):D212–20. <https://doi.org/10.1093/nar/gkaa921> PMID: [33106848](https://pubmed.ncbi.nlm.nih.gov/33106848/)
119. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
120. Bilbao-Arribas M, Guisasola-Serrano A, Varela-Martínez E, Jugo BM. The sheep miRNAome: Characterization and distribution of miRNAs in 21 tissues. *Gene*. 2023;851:146998. <https://doi.org/10.1016/j.gene.2022.146998> PMID: [36283604](https://pubmed.ncbi.nlm.nih.gov/36283604/)
121. Desvignes T, Loher P, Eilbeck K, Ma J, Urgese G, Fromm B, et al. Unification of miRNA and isomiR research: the mirGFF3 format and the mirtop API. *Bioinformatics*. 2020;36(3):698–703. <https://doi.org/10.1093/bioinformatics/btz675> PMID: [31504201](https://pubmed.ncbi.nlm.nih.gov/31504201/)
122. Loher P, Londin ER, Rigoutsos I. IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. *Oncotarget*. 2014;5(18):8790–802. <https://doi.org/10.18632/oncotarget.2405> PMID: [25229428](https://pubmed.ncbi.nlm.nih.gov/25229428/)
123. Pliatsika V, Loher P, Telonis AG, Rigoutsos I. MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. *Bioinformatics*. 2016;32(16):2481–9. <https://doi.org/10.1093/bioinformatics/btw194> PMID: [27153631](https://pubmed.ncbi.nlm.nih.gov/27153631/)
124. Fay C, Guyader V, Rochette S, Girard C. Golem: A framework for robust shiny applications. <https://github.com/ThinkR-open/golem>. 2023.
125. Krassowski M. ComplexUpset: Create Complex UpSet Plots Using “ggplot2” Components. <https://cran.r-project.org/web/packages/ComplexUpset/index.html>. 2021.
126. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3. <https://doi.org/10.1093/bioinformatics/bts034> PMID: [22257669](https://pubmed.ncbi.nlm.nih.gov/22257669/)
127. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res*. 2016;44(8):3865–77. <https://doi.org/10.1093/nar/gkw116> PMID: [26921406](https://pubmed.ncbi.nlm.nih.gov/26921406/)
128. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*. 2005;21(5):650–9. <https://doi.org/10.1093/bioinformatics/bti042> PMID: [15388519](https://pubmed.ncbi.nlm.nih.gov/15388519/)