An Improved microRNA Annotation of the Canine Genome

Luca Penso-Dolfin[1], Ross Swofford[3], Jeremy Johnson[3], Jessica Alföldi[3], Kerstin Lindblad-Toh[3,4], David Swarbreck[2], Simon Moxon[2] *, Federica Di Palma[1] *

1 Vertebrate and Health Genomics, The Genome Analysis Centre, Norwich, United Kingdom, **2** Regulatory Genomics, The Genome Analysis Centre, Norwich, United Kingdom, **3** Vertebrate Genome Biology, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **4** Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

* Simon.Moxon@tgac.ac.uk (SM); Federica.Di-Palma@tgac.ac.uk (DFP)

Abstract

The domestic dog, *Canis familiaris*, is a valuable model for studying human diseases. The publication of the latest Canine genome build and annotation, *CanFam3.1* provides an opportunity to enhance our understanding of gene regulation across tissues in the dog model system. In this study, we used the latest dog genome assembly and small RNA sequencing data from 9 different dog tissues to predict novel miRNAs in the dog genome, as well as to annotate conserved miRNAs from the *miRBase* database that were missing from the current dog annotation. We used both *miRCat* and *miRDeep2* algorithms to computationally predict miRNA loci. The resulting, putative hairpin sequences were analysed in order to discard false positives, based on predicted secondary structures and patterns of small RNA read alignments. Results were further divided into high and low confidence miRNAs, using the same criteria. We generated tissue specific expression profiles for the resulting set of 811 loci: 720 conserved miRNAs, (207 of which had not been previously annotated in the dog genome) and 91 novel miRNA loci. Comparative analyses revealed 8 putative homologues of some novel miRNA in ferret, and one in microbat. All miRNAs were also classified into the genic and intergenic categories, based on the Ensembl RefSeq gene annotation for *CanFam3.1*. This additionally allowed us to identify four previously undescribed MiRtrons among our total set of miRNAs. We additionally annotated piRNAs, using *proTRAC* on the same input data. We thus identified 263 putative clusters, most of which (211 clusters) were found to be expressed in testis. Our results represent an important improvement of the dog genome annotation, paving the way to further research on the evolution of gene regulation, as well as on the contribution of post-transcriptional regulation to pathological conditions.

Introduction

MicroRNAs (miRNAs) are small (~22 nt) non-coding RNAs found in both plants and animals. By binding to complementary RNA molecules, they can lead to translational repression or

mRNA degradation [1–4] and have also been proposed as regulators of long non-coding RNA (lncRNA) expression [5]. MiRNAs typically derive from a primary transcript, the *pri-miRNA*, which is first transcribed inside the nucleus of a cell. The pri-miRNA contains one or more stem-loop structures which are processed by the enzyme Drosha, cleaving the double-stranded stem region. The resulting molecule, called the *pre-miRNA*, is then transported out of the nucleus, where a second enzyme, Dicer, excises the loop region and generates a ~22 bp, double stranded RNA molecule [6, 7]. One of these two strands (referred as 5p- and 3p-miRNA) will typically be degraded [7], while the other will be loaded into the miRNA-induced silencing complex and guide the targeting of mRNA molecules, by partial base-pairing [6].

First described in *C. elegans* two decades ago [8], miRNAs have increasingly attracted the interest of the scientific community, as their crucial implications in embryonic development, pathology [9, 10], and functional evolution [11] became clearer. Alteration of miRNA expression occurs in cancer cells [12–15] and represents a valid tool for the distinction of different tumor types [10]. Mutations in the miRNA seed region (typically positions 2–7 of the mature miRNA, crucial for target recognition) [16] are associated with diseases such as hereditary progressive hearing loss [17], or skeletal and growth defects [18]. Because of their wide implication in gene regulation, there is a great interest in identifying novel miRNAs in human and model animal systems, both for elucidating the biology of disease conditions and the evolution of phenotypic variation.

Recent advances in sequencing technologies and compute performance in the last decade means we can generate and analyse much larger amounts of data, with greater sensitivity and exponentially lowered costs. In particular, the advent of RNA-Seq technology [19] gives us a unique opportunity to study miRNA function and evolution. Parallel to such technological innovations came the development of computational tools for the prediction of novel miRNA loci [20–22]. Such tools utilise small RNA data and a genome assembly from a particular organism as the input information, returning a list of putative miRNA loci in that genome. These tools have allowed for an integrated experimental and *in silico* approach in the discovery of new miRNAs [21, 23, 24].

Other studies focused on the computational identification of conserved miRNAs between species, using a publicly available catalogue of annotated miRNAs, miRBase [25]. Buza *et al.* [26], for instance, identified a set of homologous counterparts of human disease-associated miRNAs in domestic animals. A similar approach was used by Sunkar *et al.* [27] for the identification of miRNA homologues in a variety of plant genomes. However, homology-based annotation will not find new species-specific miRNAs, as opposed to the tools described in [20–22], or provide information on tissue or stage specific expression.

Our study focuses on the domestic dog, *Canis familiaris*, the result of wolf (*Canis lupus*) domestication, which started around 10,000 years ago [28, 29]. Since then, hundreds of dog breeds have been artificially selected, leading to very high levels of morphological and behavioral variation. Having shared the environment with humans ever since its appearance, the dog has been exposed to similar pathogens, and thus represents an important model system for the study of human diseases [30].

The first high quality draft genome of the dog appeared in 2005 [31]; the authors also generated a map of single nucleotide polymorphisms (SNPs), revealed long-range haplotype structure and explored possible approaches for genome-wide association studies of trait and disease loci. Successful GWAS studies followed, providing new insights into the biology of selected traits [32–34], while other studies lead to the identification of important candidate genes for heritable diseases [35–38] improving our understanding of the corresponding human conditions.

The link between miRNAs and pathological phenotypes in dog has been suggested by many studies. Altered expression of some miRNAs was identified in canine melanomas, lymphomas and osteosarcomas, providing insights into human cancer treatment [39–42]. Other studies include the identification of miRNAs expression profiles in a canine heart failure model [43, 44], retinal degeneration [45] and in dogs infected by canine influenza virus [46].

MiRBase currently provides a catalogue of 502 miRNAs annotated in the dog genome (ftp://mirbase.org/pub/mirbase/CURRENT/genomes/cfa.gff3). However, the recent publication of an improved canine genome build, *canFam3.1* [30] provides a unique chance to further improve this annotation. In this study, we used the latest genome assembly and small RNA sequencing data to 1) validate all previously annotated dog miRNAs, 2) identify novel miRNA loci *in silico* and 3) strengthen our predictions by the comparison of multiple, independent tissue samples.

Materials and Methods

Library generation

Small RNA libraries for 9 different tissues (blood, brain, heart, kidney, lung, ovary, skin, smooth muscle and testis) were prepared using NEB's *Multiplex Small RNA Library Prep Set for Ilumina* (product E7300S, following the included protocol), and sequenced on an Illumina *HiSeq*2000 machine. Generated small RNA data was first processed by removing adaptor sequences, then combined into a superset of 126,263,963 reads.

Annotation of known and novel miRNAs

We first aligned the complete set of all miRBase hairpin sequences to the canFam3.1 genomic sequence, using the command-line version of *BLASTN* (e-value $\leq 10^{-6}$). This gave us an initial set of putative miRNA homologues in the dog genome. We then mapped mature miRBase sequences to these putative pre-miRNA hairpins, and selected those for which at least one alignment with zero or 1 mismatch was observed.

In order to identify miRBase homologues, which were previously not annotated in dog, we downloaded the existing miRBase dog annotation (ftp://mirbase.org/pub/mirbase/CURRENT/genomes/cfa.gff3) and compared it to our complete set of predicted conserved miRNAs. We then excluded, using *Bedtools merge* [47], all of our miRNAs coordinates overlapping at least one entry in the existing miRBase annotation.

For the identification of novel loci in the combined set of *miRCat-miRDeep2* predictions, we similarly used *Bedtools merge* [47] and selected all predicted miRNAs whose coordinates do not overlap with any locus of our annotation of conserved miRNAs.

Predictions of homologous miRNAs

In order to identify putative homologous loci, we used *BLASTN* [48] with an e-value cutoff of 10^{-6} to align our set of novel hairpin sequences to the latest genome assembly (downloaded from Ensembl) of the following species: Human, Pig, Mouse, Ferret, Elephant, Microbat, Lizard, Chicken, Zebrafish and Nile Tilapia. Genomic hits were then filtered, selecting hits with alignment length of at least 60 bps. The ferret genome had the highest number of hits (10), followed by microbat (2) and human and chicken with a single hit. In order to gain more confidence on the reliability of our inferences, we mapped the corresponding dog mature sequences of these conserved miRNAs to the putative homologous loci, using *PatMaN* [49] with parameters–*g 0 –e 1* (no gaps, 1 mismatch allowed). The corresponding dog mature sequences successfully aligned to 8 predicted ferret loci and one microbat locus (S7 Table),

but not to the putative miRNA common to human and chicken from our predictions. We then decided to remove this locus from our final results.

Annotating miRNAs by genomic location

We downloaded the GTF file of canFam3.1 from the Ensembl website (http://www.ensembl. org/info/data/ftp/index.html), converted it into a BED file of genomic coordinates and removed any row corresponding to a known miRBase miRNA. Using *Bedtools* [47], we found all overlaps between our hairpin coordinates and the entries from the GTF file. We selected overlaps implicating at least 15% of the total hairpin sequences. We were then able to classify our miRNAs into the categories UTR and EXON, according to the label of the overlapping entry in the GTF file. Using *Genometools* [50], we were also able to generate a gff file containing intronic coordinates for the canFam3.1 assembly. We then used an identical approach to identify miRNA loci located inside an intron (or overlapping one by 15% of their sequence). Finally, all unclassified loci were assigned to the INTERGENIC category.

We also looked for miRtrons, using the following approach: first we selected introns of length 50–120 bps [51]; we then used *Bedtools intersect* to find overlapping intronic and hairpin sequences, such that at least 95% of the intronic sequence was covered. Next, we aligned the full set of small RNA reads to these selected introns, using *PatMaN* [49] with parameters*–g 0 –e 0* (no gaps or mismatches allowed). When two clear peaks of read alignments at the 5' and 3' ends of the intronic hairpin sequence were observed (taking into account the possible presence of a short 3' or 5' tail [52], trimmed as part of the post-transcriptional modifications), the locus was considered a putative miRtron.

Predicting piRNA clusters

Small RNA reads and CanFam3.1 annotation was used as input for the software *proTRAC-2.0.5* [53]. We used the following pipeline: initially we mapped our small RNA reads to the dog genome sequence; then we performed a hit counts normalization (based on uniquely mapping sequence reads) using the script *reallocate.pl* (part of *proTRAC* package). The resulting file of mapping reads with normalized counts was used as an input for the main *proTRAC* script, returning the piRNA predictions. For this step, we used the following command line parameters: *-map canine_miRNA.map.weighted -repeatmasker canFam3repeatmasker -geneset CanFam3.1.81.gtf -genome CanFam3.1.fa -clsize 1000 -1Tor10A 0.5 -distr 1–90 –nr–nh*. Finally, we merged overlapping coordinates from the combined set of predicted piRNA clusters across all tissues, using *Bedtools merge* [47].

Gene target prediction

Target predictions of all novel miRNAs, as well as all newly discovered conserved miRNAs, were carried out using *TargetScan7* software [16]. We used the UTR sequences and the code available for download (http://www.targetscan.org/cgi-bin/targetscan/data_download.cgi?db=vert_70) together with the CanFam3.1 coding sequences obtained from Biomart (http://www.biomart.org), reformatted to meet the requirements of TargetScan7.

Accession number

Small RNA data used for this project has been deposited at the National Center for Biotechnology Information (NCBI) under the BioProject accession PRJNA305157.

Results

miRCat and miRDeep2 identify known and novel miRNA loci

Size distribution (S1 and S2 Figs) and counts of genome matching reads for each tissue (S3 Fig) were generated as a preliminary quality control. We observed a clear enrichment for 22 nt long reads (57,642,938 or 45% of the total number), the typical size of a mature miRNA. Only the blood sample had a peak at size class 23. When we looked at the size distribution of genome matching reads (S3 Fig), however, the expected enrichment for 22 nt long reads was observed, although the count for 23 nt long reads was still high. Additional analyses showed how the peak at 23nt in the blood sample was mainly due to two highly abundant sequences that did not map to the reference genome. These sequences were searched against the miRBase database and returned perfect alignments to the first 22 nt of miR-486-5p (previously described as an important regulator of erythropoiesis) [54, 55]. The 3' nucleotide of the two miRNA sequences (A and U respectively) were non-templated bases suggesting that this miRNA undergoes uridylation (6,960,721 reads in the blood sample) and adenylation (1,161,985 reads) [56] in the blood sample studied (Figs 1 and 2). These modifications are of great interest as they can affect the stability of the mature miRNA [57] which can in turn affect regulation of target genes. The abundance of the canonical 22-nt mature miRNA and 23-nt adenylated sequences appeared to be similar across all 9 tissues, while the 23-nt uridylated isoform was found to be consistently the most abundant (with the only exception of the skin sample). However, the counts for all isoforms were significantly higher in blood than in any other sample (Figs 1 and 2) leading to a peak at 23nt in the size distribution of total trimmed reads in this sample.

Next, we ran *miRCat* [20] and *miRDeep2* [22] on the total set of 85,810,181 trimmed, genome matching reads, thus obtaining two distinct sets of miRNA predictions (Fig 3). Genomic coordinates of the 589 hairpin sequences predicted by miRCat, and the 609 loci predicted by *miRDeep2* were merged using *Bedtools merge* [47], thus obtaining a non-overlapping combined set of 880 coordinates.

Re-annotation of the canFam3.1 assembly separates high and low confidence miRNAs

Hairpin sequences for both novel and conserved miRNA candidates were derived from their genomic coordinates. Small RNA reads were then aligned to these sequences, and we identified patterns of alignment which were consistent with Dicer and Drosha processing (typically, a precise "peak" at both the 5' end and 3' end of the sequence) [25, 58]. Where the presence of 3p-miRNA and 5p-miRNA sequences (S3 and S4 Tables) was supported by alignments (S5 and S6 Tables), hairpin secondary structures were generated (see Fig 4) using the Vienna-RNA package [59]. This software determines the most likely folding structure based on the minimum free-energy criterion. Based on the consistency of both the alignments and predicted secondary structure, 720 conserved miRNAs (S1 Table) and 91 high confidence novel loci (S2 Table) were identified (see also Fig 5). Each conserved miRNA was either classified as "high confidence" (consistent alignment pattern and secondary structure, 417 loci) or "low confidence" (where any of the above criteria was not met, 303 loci). The complete set of conserved and novel miRNA predicted structures can be found in S6 and S8 Figs, respectively. Small RNA reads provided evidence of expression of both the 3p- and 5p-miRNA of all newly identified loci but only 467 of the 720 conserved miRNAs (whose structures can be found in S7 Fig) received the same level of support. As we only sequenced nine different tissues it is entirely possible that the miRNAs we have annotated as low confidence could be supported by expression
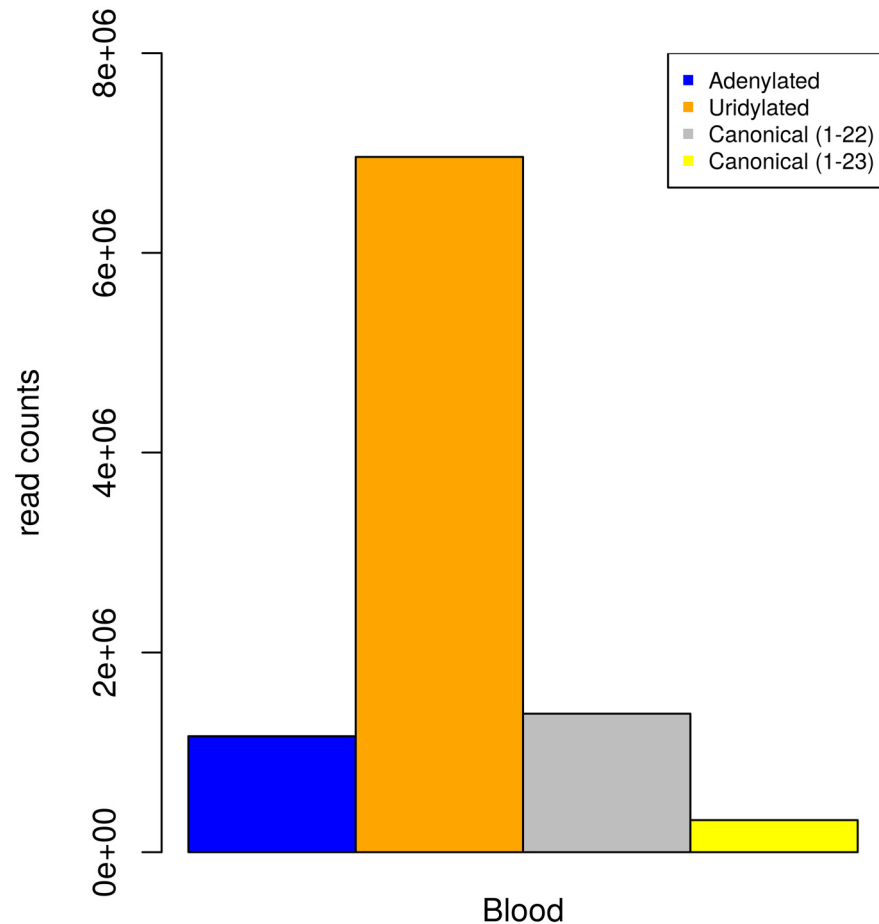
Fig 1. Counts of different isoforms of miRNA cfa-mir-486-5p in the blood sample.

doi:10.1371/journal.pone.0153453.g001

data in other tissues and therefore we have annotated them solely on similarity to miRNAs currently annotated in miRBase and conservation of the annotated mature miRNA.

By excluding all of our miRNAs coordinates overlapping at least one entry in the existing miRBase annotation (see Materials and Methods), we were also able to identify, from our set of 720 conserved loci, 207 newly annotated miRNAs, 65 of which were classified as "high confidence". Tissue specific expression plots for all novel (S4 Fig) and conserved miRNAs (S5 Fig) were also generated allowing visualization of expression data across each of the tissues sampled and demonstrating that predictions are expressed in multiple independent datasets (Fig 6B).

The complete annotation of both conserved and novel miRNAs is available online, as part of the UCSC genome browser "Broad Institute Canine Annotation" Track Hub (https://genome. ucsc.edu/cgi-bin/hgGateway?hgsid=432322929_FBTCtBIVIPVoA0nQXPzwAguagWPr).

Read abundance plots indicate arm switching across tissues

Generating expression plots as shown in Fig 6 allowed for the visualization of a very important phenomenon: the switching in abundance across tissues of the 5' and 3' mature sequences. These changes can be associated with important functional consequences, as different mature strands might target different mRNAs [60]. In order to gain a better understanding of the type of variation in our data, we generated abundance plots, using *Rstudio* (https://www.rstudio.com/), where
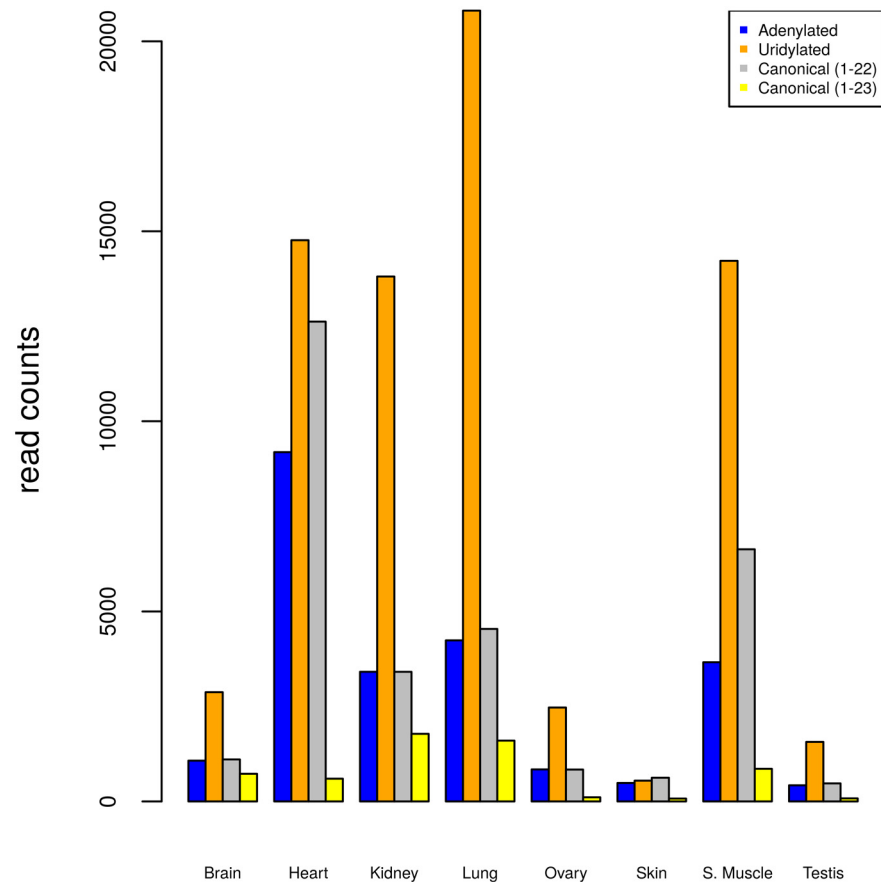
Fig 2. Counts of different isoforms of miRNA cfa-mir-486-5p in all samples except blood.

the read abundance at the 5' end for each tissue is plotted against the corresponding abundance at the 3' (S10 Fig). We observed cases of arm switching, possibly indicating functional diversification across tissues. For example, novel miRNA *18/50102465-50102527(+)* shows prevalence of the 3p-miRNA in blood and heart, but prevalence of 5p-miRNA in brain, lung and testis (and no evidence of expression in the remaining tissues). Similarly, *cfa-mir-18a* shows a clear prevalence of 3p-miRNA in blood, while in all other tissues the 5p-miRNA is more abundant. Interestingly, the 3p-miRNA counts in blood are much higher than those of the 5p-miRNA in the other tissues.

We also generated histograms for the read counts across the hairpin sequences (S9 Fig). Unsurprisingly, we observed highly 3' or 5' biased distributions of reads, suggesting that in most cases only one strand of the mature miRNA is retained and plays a role in gene regulation.

Comparative analyses of novel miRNAs reveal 9 putative homologues

Next we addressed the question whether some of our newly identified dog miRNAs are present in other vertebrates. We selected the following species: Human, Pig, Mouse, Ferret, Elephant, Microbat, Lizard, Chicken, Zebrafish and Nile Tilapia. In order to identify putative homologous loci in this diverse set of organisms, we used *BLASTN* to align our set of novel hairpin sequences to the latest genome assembly (downloaded from Ensembl) of each species. Selected hits (see Materials and Methods for more details) were further analysed by mapping the corresponding dog mature sequences to these putative homologous hairpins, allowing for 1
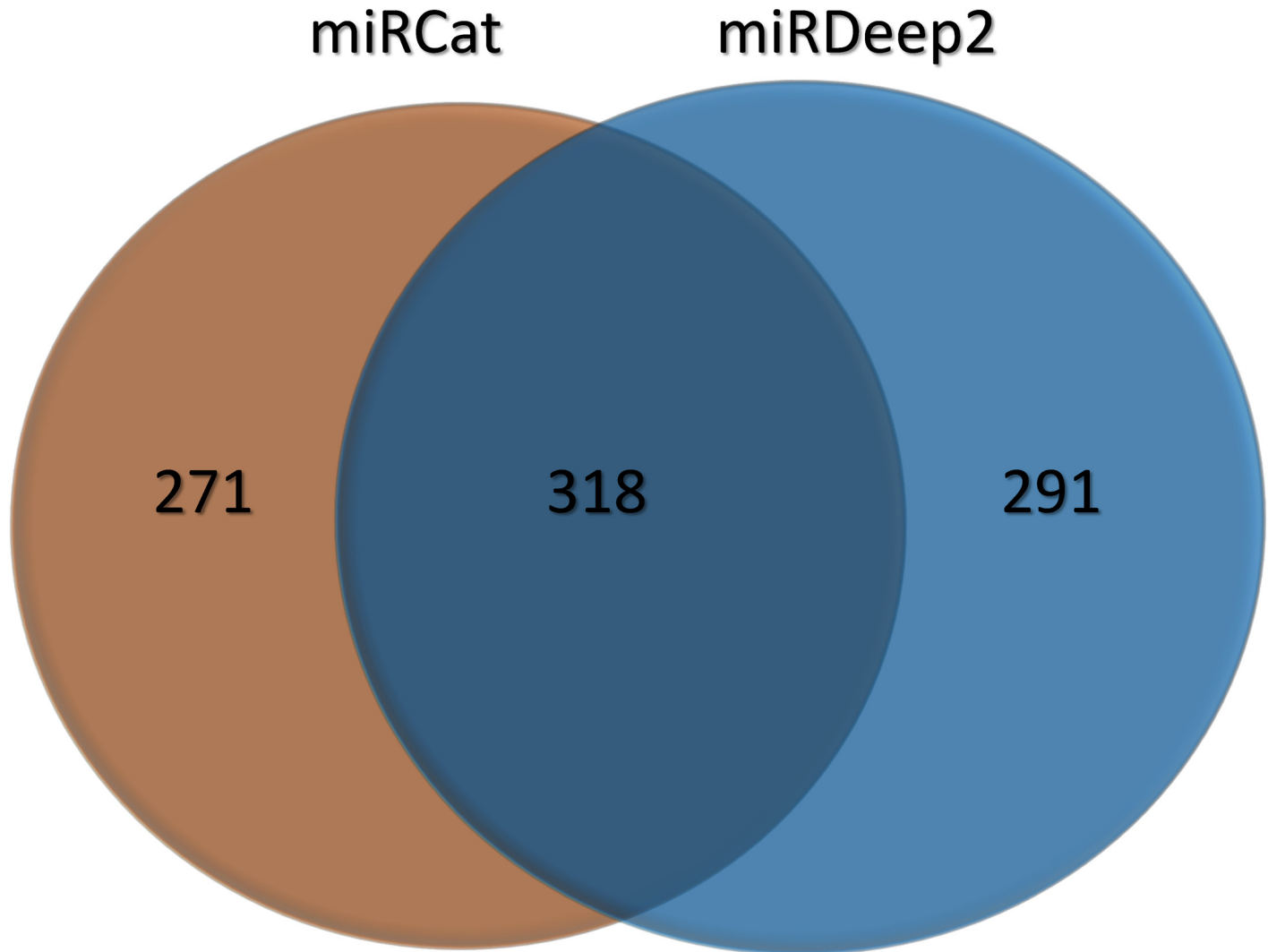
Predictions of miRNA loci

Fig 3. Results of the computational prediction of miRNA loci, using miRCat and miRDeep2.

mismatch. This lead to the identification of 8 miRNA loci in ferret and one in microbat ([S7 Table](#)). We also compared the genomic annotation of these homologous miRNAs: we found perfect correspondence intron to intron or exon to exon for all putative homologues, giving further support to our predictions ([Table 1](#)). Our results thus indicate that these novel miRNAs are not conserved across long evolutionary timescales; however, when we look at the two evolutionary most closely related species (ferret, with an estimated divergence from dog of about 46 MYA, and microbat, about ~82 MYA) (http://www.timetree.org) we find some putative conserved miRNAs.
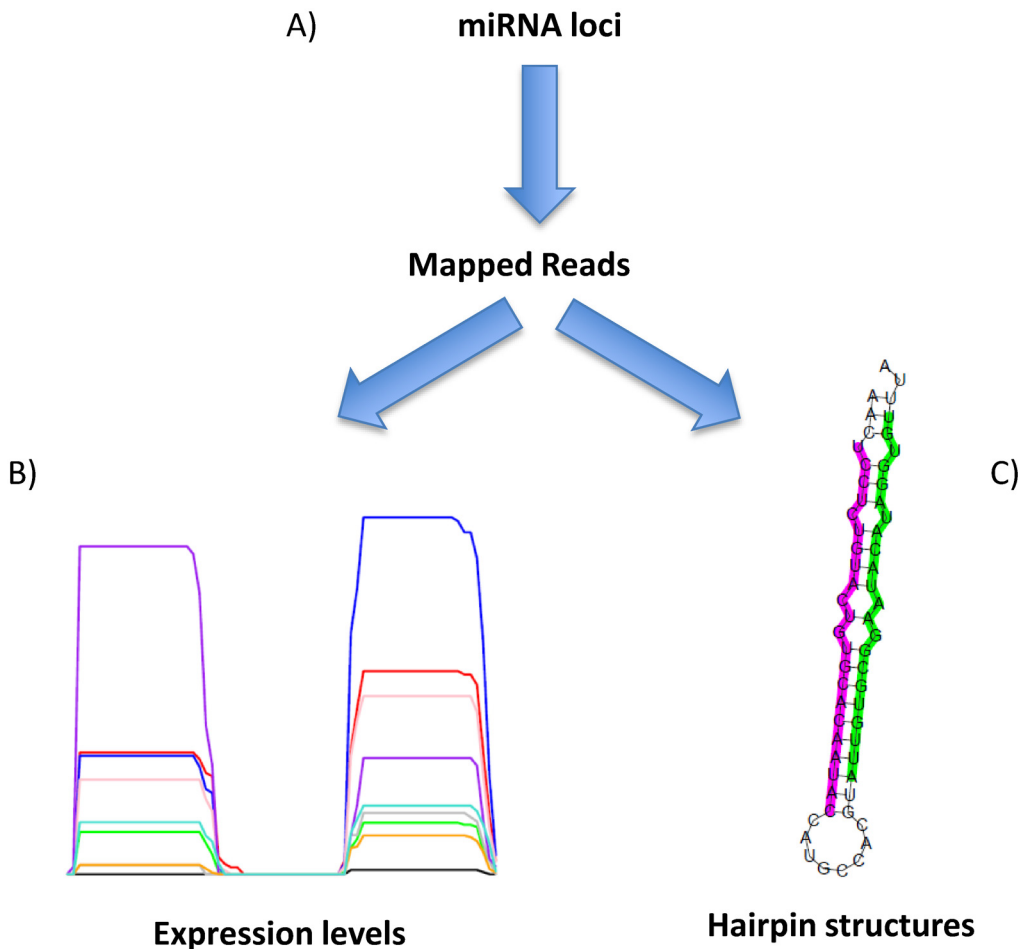
Fig 4. Workflow for the generation of miRNA expression profiles and secondary structures. Small RNA reads are first mapped (A) to the hairpin sequences, looking for two peak alignment patterns as shown in B (see S4 and S5 Figs). Expression levels across each hairpin (B) and secondary structures (C) are then derived (S6 and S7 Figs).

Canfam3.1 gene annotation reveals 4 possible MiRtrons in our dataset

Next, we used the latest dog genome annotation (available from the Ensembl website, http://www.ensembl.org/info/data/ftp/index.html) to classify our complete set of miRNAs, based on their genomic coordinates. As expected, the vast majority of our loci fell into the intronic or intergenic (all unclassified miRNAs) categories, with fewer located in UTR or CDS regions (see Fig 7). By comparing the genomic locations of conserved and novel miRNAs (S11 Fig) we observed much higher proportions of intronic loci in the novel set (66%) compared to the conserved miRNAs (25%). This is in agreement with what Meunier *et al.* [11] observed for recently evolved mammalian miRNAs. The overrepresentation of novel loci in intronic sequences is unsurprising if we consider that the transcription of the host coding gene will favor the evolution of a novel miRNA [6]. Moreover, the interaction between Drosha and the spliceosome can lead to the formation of an intronic miRNA precursor, excised from the primary mRNA transcript [61].

A particularly interesting class of intronic miRNAs is represented by miRtrons, whose pre-miRNA is generated by mRNA splicing, rather than by Drosha cleavage. Given the observed high number of intronic loci in our dataset, we addressed the question whether some of them
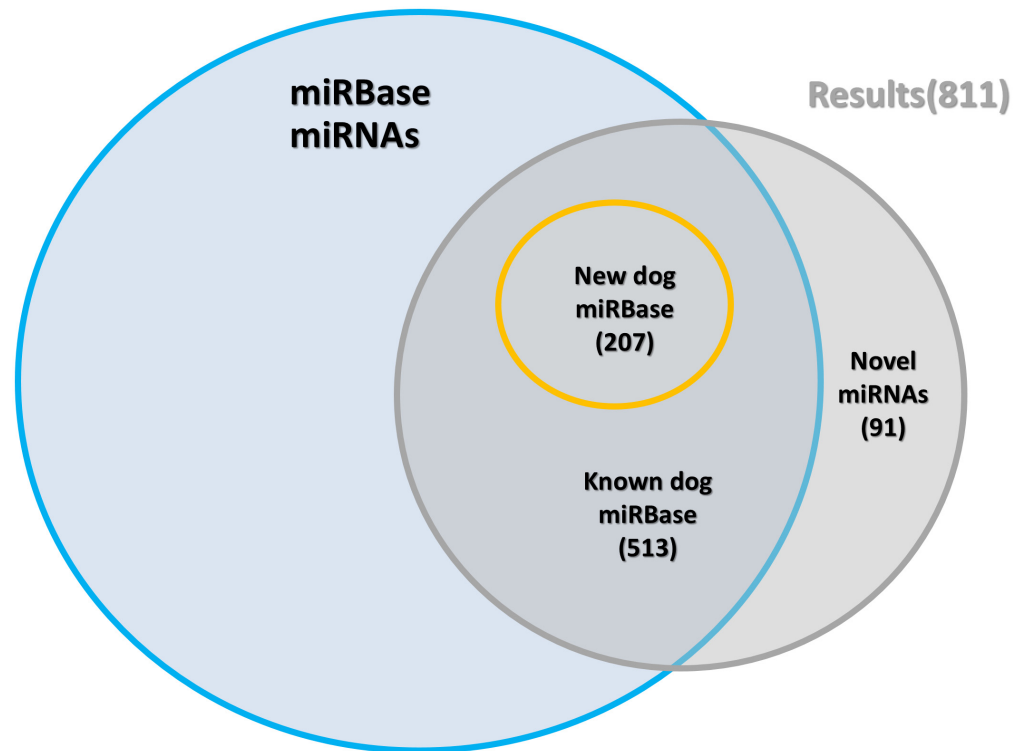
Fig 5. Identification of known and novel miRNA loci.

might represent miRtrons. We would expect these pre-miRNAs to cover the majority of the short intronic sequences they are processed from, with the 3p-miRNA and 5p-miRNA mapping to the corresponding ends of the intron. By aligning the full set of small RNA reads to these selected introns (as previously done with all predicted miRNA loci), we could identify 4 putative miRtrons: *12/406899-406968(+)_mir-877* (whose human homologue has been proposed as a miRtron) and novel miRNAs *4/35856564-35856639(+)*, *9/22412453-22412517* and *13/37816081-37816136*. These results represent the first description of putative miRtrons in the dog genome, and give further relevance to the novel predictions of this study.

Annotation of piRNAs

The size distribution of our combined small RNA dataset suggested not only the presence of miRNAs, but also longer, 24–30 bp sequences likely to represent piRNAs. These are an additional class of non-coding RNA molecules involved in the silencing of transposons and other repetitive elements [62–64]. We were thus able to identify a total of 263 predicted clusters (S8 Table). Unsurprisingly [65], we observed a clear enrichment in testis (211 clusters) compared to all other tissues (see Fig 8). We then selected the 231 mono-strand predicted clusters and used *Bedtools merge* [47] to combine overlapping coordinates. This returned a set of 205 merged coordinates, which are listed in S9 Table.

Target prediction analysis

We performed target predictions of all new miRNAs (S10 Table), as well as all newly annotated conserved loci (S11 Table), using *TargetScan7* software [16]. Results provide an overview of possible gene targets, as well as information about the predicted strength of the miRNA-
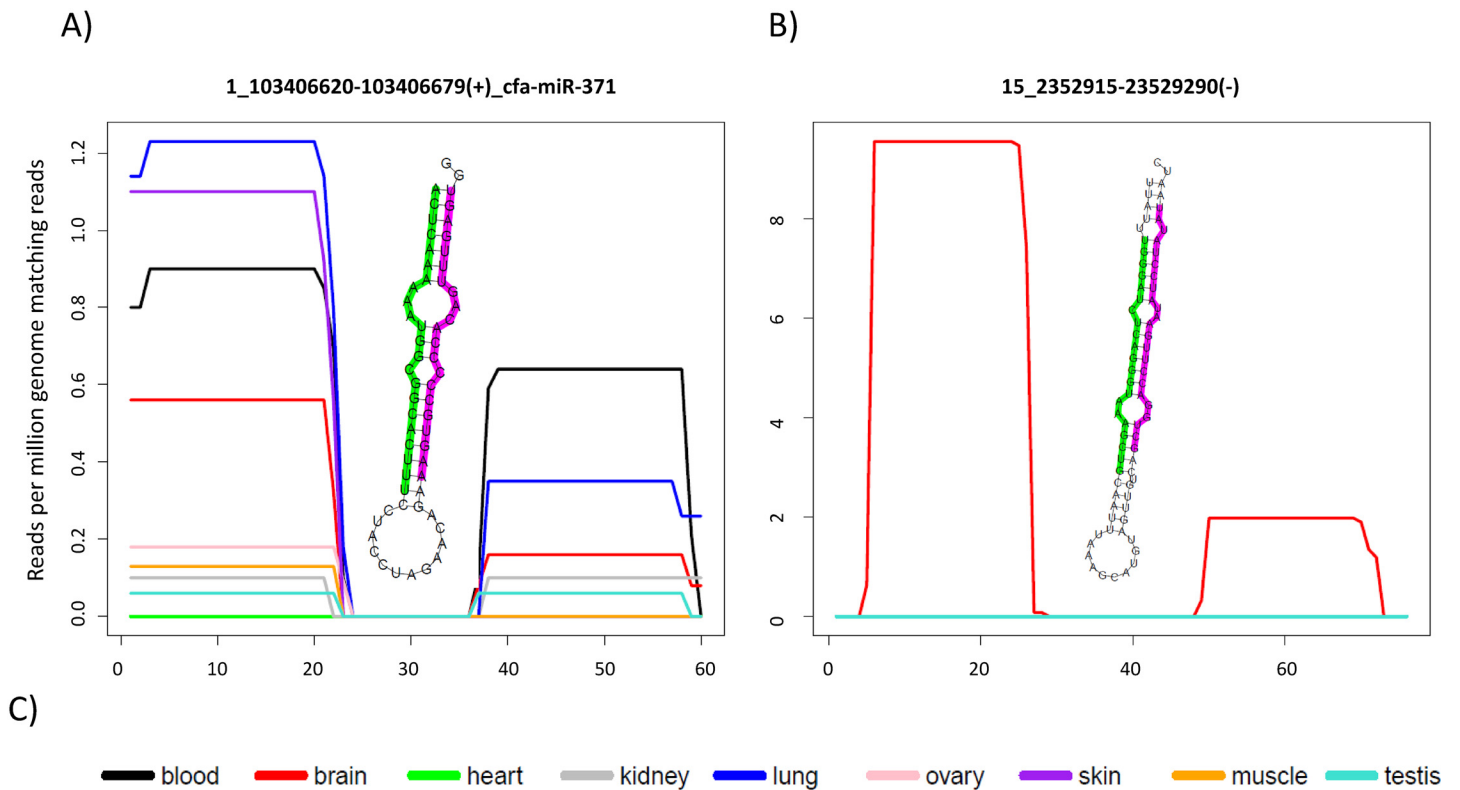
Fig 6. MiRNA expression plots and predicted secondary structures. Different colours in the plot correspond to different tissues as described in C). Green and purple-labelled bases in the hairpin figure indicate the most abundant read mapping to the 5' and 3' ends; conserved miRNA cfa-miR-371 (A) appears to be absent in heart and shows variable expression levels in the other tissues. B) Plots for a novel miRNA located in chromosome 15, supported by small RNA reads from the brain sample.

mRNA interaction (expressed by the context++ score). For a detailed description of the analyses, see section Materials and Methods.

Discussion

Our study provides an important improvement of the dog miRNA annotation, summarized by the classification of 720 conserved miRNAs into the high and low confidence categories, along with the identification of 91, previously undescribed loci. Analysis of 9 tissues gives a broad

Table 1. Putative miRNA loci, homologous to some novel dog miRNAs, identified in ferret and microbat.

| Species	Chromosome	Start	End	CanFam3.1 locus	Strand	Dog annotation	Species annotation
Ferret	GL896934.1	8105928	8105988	12/37307857-37307917(+)	+	INTRON	INTRON
Ferret	GL897283.1	41870	41948	15/23529215-23529290(-)	+	INTRON	INTRON
Ferret	GL896969.1	5729349	5729420	18/49182272-49182343(+)	-	INTRON	INTRON
Ferret	GL896988.1	2200068	2200144	3/84489968-84490044(-)	-	INTRON	INTRON
Ferret	GL896988.1	2218967	2219043	3/84489968-84490044(-)	+	INTRON	INTRON
Ferret	GL896900.1	21231318	21231382	8/42665000-42665064(-)	-	CDS	CDS
Ferret	GL896965.1	8980331	8980407	X/15639699-15639775(-)	-	INTRON	INTRON
Ferret	GL896945.1	1693147	1693222	X/32945413-32945490(-)	-	INTRON	INTRON
Microbat	GL429826	6144016	6144076	12/37307857-37307917(+)	+	INTRON	INTRON

MiRNA classification
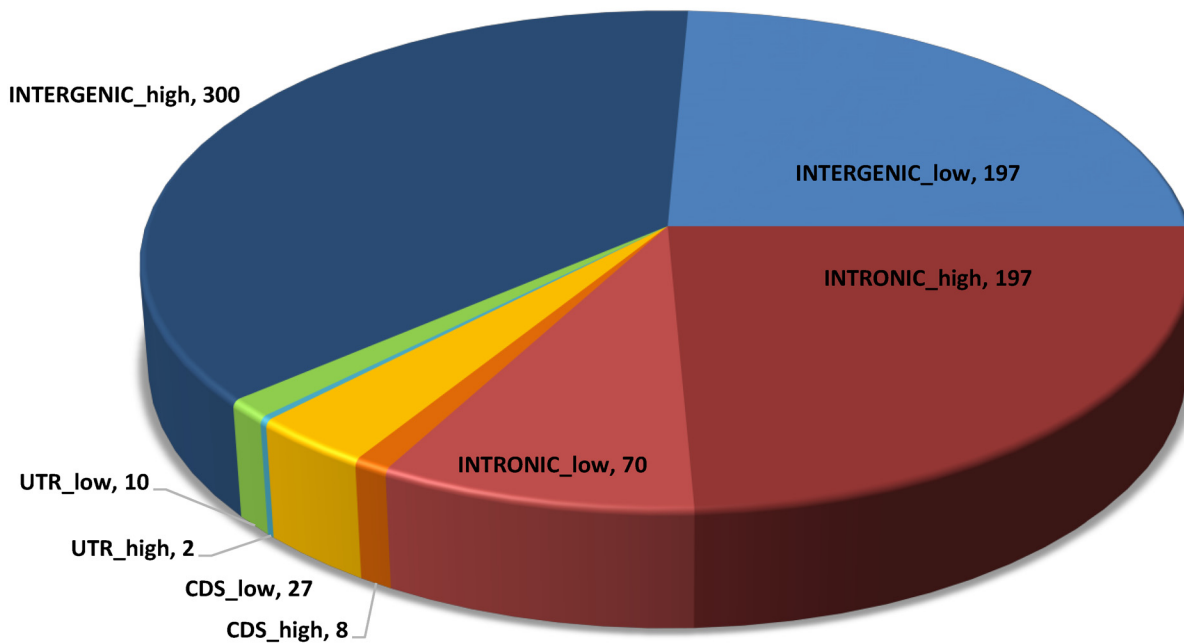
Fig 7. Annotation of the complete set of miRNAs, based on genomic location.

doi:10.1371/journal.pone.0153453.g007

overview of miRNA expression levels (S4 and S5 Figs), providing support across independent samples for a subset of 417 conserved miRNAs and all 91 novel miRNA predictions. Our improved microRNA annotation (available online as part of the UCSC genome browser "Broad Institute Canine Annotation" Track Hub) represents a valuable resource for the community, as it provides a possible starting point to both biomedical and evolutionary studies on the dog model system. Many polymorphisms in miRNAs target sites have already been identified which are associated with pathological conditions [66]. Investigating the functional role of these novel miRNAs has the potential of further elucidating the biology of several human diseases, and the evolution of artificially selected phenotypic traits in dog.

Our data has also highlighted an example of highly abundant adenylated and uridylated mature miRNA in the blood sample (Fig 1). This is in agreement with a study by Koppers-Lalic *et al.* [67], who observed high proportions of non-canonical, monoadenylated mir-486-5p sequences. The total count for the two modified cfa-mir-486-5p sequences in blood (8,122,706) represents as much as 43% of all size selected, trimmed reads in the sample. These modifications appear to be associated with increased stability of the mature miRNA [57] and hence can have an effect on target genes expression. Given the implication of mir-486 in erythropoiesis [54, 55], the observed modifications are of great interest, as they are likely to be linked to important regulatory mechanisms in blood.

PIRNA PREDICTIONS

Testis, 211

Smooth Muscle, 8

Skin, 1

Ovary, 9

Lung, 3

Kidney, 16

Heart, 6   Brain, 6   Blood, 3
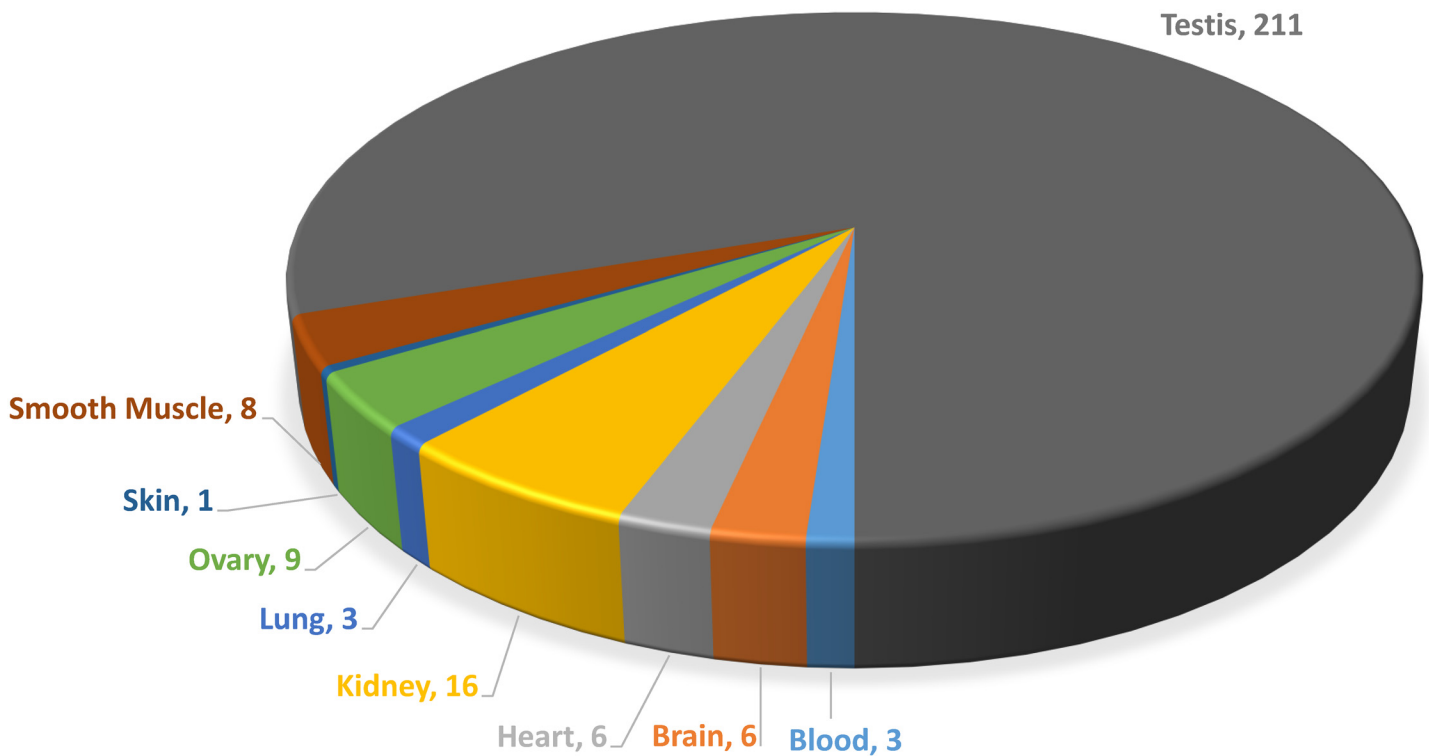
Fig 8. Counts of proTRAC piRNA predictions across tissues.

doi:10.1371/journal.pone.0153453.g008

An additional source of variation in our data is represented by a switching in the ratio of abundance between 5p- and 3p- miRNA sequences across samples. These results support the idea that distinct mature miRNA strands can be retained, and thus be functionally active, in different tissues. Such variation can be associated with differences in gene regulation, as a result of differential targeting or changes in targeting efficacy. Investigating the contribution of this variation in tissue specific regulation represents an important and interesting challenge.

miRNAs appear to be a key component in the evolution of life [6, 68]. Differential expression, target switching across tissues and organisms, together with the birth and death of miRNA loci in an evolutionary lineage, provide a wide spectrum of possibilities for phenotypic diversification. As opposed to the emergence of novel gene functions, functional miRNA genes represent a much easier way to explore new evolutionary pathways. Subtle modifications in how every gene is regulated in each tissue may have well represented, in some cases, the necessary base of a wider phenotypic diversity, or additional organismal complexity. Further research is needed in order to better elucidate the link between functional evolution and the observed diversity of miRNA expression, appearance and loss across taxa.

Supporting Information

S1 Fig. Read length distribution of trimmed, size filtered ($\geq$16 bps) reads of each sample. (PDF)

S2 Fig. Read length distribution of trimmed, size filtered (≥16 bps) genome matching reads for each sample.
(PDF)

S3 Fig. Counts of all reads, trimmed ≥16 bps and trimmed genome matching reads across all samples.
(PDF)

S4 Fig. Line plots of expression levels across the hairpin for all novel miRNAs. The base position along the hairpin (x axis) is plotted against the normalized read count (reads per million genome matching reads). Different colors in the plot correspond to different tissues, as described by the legend on the bottom of each plot.
(PDF)

S5 Fig. Line plots of expression levels across the hairpin for all conserved miRNA candidates. The base position along the hairpin (x axis) is plotted against the normalized read count (reads per million genome matching reads). Different colors in the plot correspond to different tissues, as described by the legend on the bottom of each plot.
(PDF)

S6 Fig. Predicted secondary structures for all conserved miRNA candidates.
(PDF)

S7 Fig. Predicted secondary structures for a subset of 467 conserved miRNA candidates, supported by evidence of expression of both the 5p- and 3p miRNAs. Purple and green labelled regions respectively correspond to the 5p- and 3p miRNA sequences, as predicted by small RNA reads alignments to the hairpin sequence.
(PDF)

S8 Fig. Predicted secondary structures for all novel miRNAs. Purple and green labelled regions respectively correspond to the 5p- and 3p miRNA sequences, as predicted by small RNA reads alignments to the hairpin sequence.
(PDF)

S9 Fig. Sample specific histograms of the proportion of 5' reads (considering all 811 miRNAs), divided into 10 intervals. High counts for the first and tenth bin indicate the expected extreme values for the 3'/5' proportions.
(PDF)

S10 Fig. Abundance plots of all miRNAs across different tissues. Counts of reads at the 5' end are plotted against the counts at the 3' end. Different symbols indicate different tissues, as described by the legend.
(PDF)

S11 Fig. Counts of novel and conserved miRNAs, classified by genomic location.
(PDF)

S1 Table. Genomic coordinates and hairpin sequences of all conserved miRNAs.
(XLSX)

S2 Table. Genomic coordinates and hairpin sequences of all novel miRNAs.
(XLSX)

S3 Table. Genomic coordinates and sequences of all conserved mature miRNAs.
(XLSX)

S4 Table. Genomic coordinates and sequences of all novel mature miRNAs.
(TXT)

S5 Table. Alignments of small RNA reads to the hairpin sequences of conserved miRNAs.
(ZIP)

S6 Table. Alignments of small RNA reads to the hairpin sequences of novel miRNAs.
(ZIP)

S7 Table. Genomic coordinates for 9 predicted miRNAs (homologous to some novel miRNA) in ferret and microbat.
(TXT)

S8 Table. Genomic coordinates of all piRNA clusters predicted by the software proTRAC.
(TXT)

S9 Table. Merged genomic coordinates of all piRNA clusters predicted by the software pro-TRAC.
(ZIP)

S10 Table. List of all novel miRNAs target predictions, as returned by the software TargetScan7.
(BZ2)

S11 Table. List of target predictions for all newly annotated conserved miRNAs, as returned by the software TargetScan7.
(BZ2)

Author Contributions

Conceived and designed the experiments: KLT RS FDP JA DS SM. Performed the experiments: RS. Analyzed the data: LPD SM JJ. Contributed reagents/materials/analysis tools: SM. Wrote the paper: LPD SM JA KLT FDP.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116(2):281–97. PMID: 14744438.

2. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136 (4):642–55. doi: 10.1016/j.cell.2009.01.035 PMID: 19239886; PubMed Central PMCID: PMC2675692.

3. Guo J, Ren F, Wang Y, Li S, Gao Z, Wang X, et al. miR-764-5p promotes osteoblast differentiation through inhibition of CHIP/STUB1 expression. J Bone Miner Res. 2012; 27(7):1607–18. doi: 10.1002/jbmr.1597 PMID: 22407479.

4. Stroynowska-Czerwinska A, Fiszer A, Krzyzosiak WJ. The panorama of miRNA-mediated mechanisms in mammalian cells. Cell Mol Life Sci. 2014; 71(12):2253–70. doi: 10.1007/s00018-013-1551-6 PMID: 24468964; PubMed Central PMCID: PMC4031385.

5. Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, et al. microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. Oncogene. 2011; 30 (47):4750–6. doi: 10.1038/onc.2011.193 PMID: 21625215; PubMed Central PMCID: PMC4292930.

6. Berezikov E. Evolution of microRNA diversity and regulation in animals. Nat Rev Genet. 2011; 12 (12):846–60. doi: 10.1038/nrg3079 PMID: 22094948.

7. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014; 15(8):509–24. doi: 10.1038/nrm3838 PMID: 25027649.

8. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993; 75(5):843–54. PMID: 8252621.

9. Mraz M, Pospisilova S. MicroRNAs in chronic lymphocytic leukemia: from causality to associations and back. Expert Rev Hematol. 2012; 5(6):579–81. doi: 10.1586/ehm.12.54 PMID: 23216588.

10. 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. doi: 10.1038/nature03702 PMID: 15944708.

11. Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, et al. Birth and expression evolution of mammalian microRNA genes. Genome Res. 2013; 23(1):34–45. doi: 10.1101/gr.140269.112 PMID: 23034410; PubMed Central PMCID: PMC3530682.

12. Michael MZ, SM OC, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res. 2003; 1(12):882–91. PMID: 14573789.

13. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002; 99(24):15524–9. doi: 10.1073/pnas.242606799 PMID: 12434020; PubMed Central PMCID: PMC137750.

14. Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, et al. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A. 2005; 102(10):3627–32. doi: 10.1073/pnas. 0500613102 PMID: 15738415; PubMed Central PMCID: PMC552785.

15. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. Cell. 2005; 120(5):635–47. doi: 10.1016/j.cell.2005.01.014 PMID: 15766527.

16. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015; 4. doi: 10.7554/eLife.05005 PMID: 26267216; PubMed Central PMCID: PMC4532895.

17. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, et al. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet. 2009; 41(5):609–13. doi: 10.1038/ng.355 PMID: 19363479.

18. de Pontual L, Yao E, Callier P, Faivre L, Drouin V, Cariou S, et al. Germline deletion of the miR-17 approximately 92 cluster causes skeletal and growth defects in humans. Nat Genet. 2011; 43 (10):1026–30. doi: 10.1038/ng.915 PMID: 21892160; PubMed Central PMCID: PMC3184212.

19. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63. doi: 10.1038/nrg2484 PMID: 19015660; PubMed Central PMCID: PMC2949280.

20. Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L, et al. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics. 2012; 28(15):2059–61. doi: 10.1093/bioinformatics/bts311 PMID: 22628521; PubMed Central PMCID: PMC3400958.

21. Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol. 2008; 26(4):407–15. doi: 10.1038/ nbt1394 PMID: 18392026.

22. Friedlander 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. doi: 10.1093/nar/gkr688 PMID: 21911355; PubMed Central PMCID: PMC3245920.

23. Bentwich I. Prediction and validation of microRNAs and their targets. FEBS Lett. 2005; 579(26):5904–10. doi: 10.1016/j.febslet.2005.09.040 PMID: 16214134.

24. Jain M, Chevala VV, Garg R. Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. J Exp Bot. 2014; 65(20):5945–58. doi: 10.1093/jxb/ eru333 PMID: 25151616; PubMed Central PMCID: PMC4203128.

25. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014; 42(Database issue):D68–73. doi: 10.1093/nar/gkt1181 PMID: 24275495; PubMed Central PMCID: PMCPMC3965103.

26. Buza T, Arick M 2nd, Wang H, Peterson DG. Computational prediction of disease microRNAs in domestic animals. BMC Res Notes. 2014; 7:403. doi: 10.1186/1756-0500-7-403 PMID: 24970281; PubMed Central PMCID: PMC4091757.

27. Sunkar R, Jagadeeswaran G. In silico identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol. 2008; 8:37. doi: 10.1186/1471-2229-8-37 PMID: 18416839; PubMed Central PMCID: PMC2358906.

28. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 2013; 495(7441):360–4. doi: 10.1038/nature11837 PMID: 23354050.

29. Pang JF, Kluetsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, et al. mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. Mol Biol Evol. 2009; 26(12):2849–64. doi: 10.1093/molbev/msp195 PMID: 19723671; PubMed Central PMCID: PMC2775109.

30. Hoeppner MP, Lundquist A, Pirun M, Meadows JR, Zamani N, Johnson J, et al. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. PLoS One. 2014; 9(3):e91172. doi: 10.1371/journal.pone.0091172 PMID: 24625832; PubMed Central PMCID: PMC3953330.

31. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005; 438(7069):803–19. doi: 10.1038/nature04338 PMID: 16341006.

32. Dreger DL, Parker HG, Ostrander EA, Schmutz SM. Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in Basset Hounds and Pembroke Welsh Corgis. J Hered. 2013; 104(3):399–406. doi: 10.1093/jhered/est012 PMID: 23519866.

33. Bannasch D, Young A, Myers J, Truve K, Dickinson P, Gregg J, et al. Localization of canine brachy-cephaly using an across breed mapping approach. PLoS One. 2010; 5(3):e9632. doi: 10.1371/journal.pone.0009632 PMID: 20224736; PubMed Central PMCID: PMC2835769.

34. Hoopes BC, Rimbault M, Liebers D, Ostrander EA, Sutter NB. The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. Mamm Genome. 2012; 23(11–12):780–90. doi: 10.1007/s00335-012-9417-z PMID: 22903739; PubMed Central PMCID: PMC3511640.

35. Olsson M, Meadows JR, Truve K, Rosengren Pielberg G, Puppo F, Mauceli E, et al. A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. PLoS Genet. 2011; 7(3):e1001332. doi: 10.1371/journal.pgen.1001332 PMID: 21437276; PubMed Central PMCID: PMC3060080.

36. Seppala EH, Jokinen TS, Fukata M, Fukata Y, Webster MT, Karlsson EK, et al. LGI2 truncation causes a remitting focal epilepsy in dogs. PLoS Genet. 2011; 7(7):e1002194. doi: 10.1371/journal.pgen.1002194 PMID: 21829378; PubMed Central PMCID: PMC3145619.

37. Downs LM, Wallin-Hakansson B, Boursnell M, Marklund S, Hedhammar A, Truve K, et al. A frameshift mutation in golden retriever dogs with progressive retinal atrophy endorses SLC4A3 as a candidate gene for human retinal degenerations. PLoS One. 2011; 6(6):e21452. doi: 10.1371/journal.pone.0021452 PMID: 21738669; PubMed Central PMCID: PMC3124514.

38. Wilbe M, Jokinen P, Truve K, Seppala EH, Karlsson EK, Biagi T, et al. Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet. 2010; 42(3):250–4. doi: 10.1038/ng.525 PMID: 20101241.

39. Noguchi S, Mori T, Hoshino Y, Yamada N, Maruo K, Akao Y. MicroRNAs as tumour suppressors in canine and human melanoma cells and as a prognostic factor in canine melanomas. Vet Comp Oncol. 2013; 11(2):113–23. doi: 10.1002/vco.306 PMID: 23638671.

40. Noguchi S, Kumazaki M, Mori T, Baba K, Okuda M, Mizuno T, et al. Analysis of microRNA-203 function in CREB/MITF/RAB27a pathway: comparison between canine and human melanoma cells. Vet Comp Oncol. 2014. doi: 10.1111/vco.12118 PMID: 25280339.

41. Pazzaglia L, Leonardi L, Conti A, Novello C, Quattrini I, Montanini L, et al. miR-196a expression in human and canine osteosarcomas: a comparative study. Res Vet Sci. 2015; 99:112–9. doi: 10.1016/j.rvsc.2014.12.017 PMID: 25599934.

42. Uhl E, Krimer P, Schliekelman P, Tompkins SM, Suter S. Identification of altered MicroRNA expression in canine lymphoid cell lines and cases of B- and T-Cell lymphomas. Genes Chromosomes Cancer. 2011; 50(11):950–67. doi: 10.1002/gcc.20917 PMID: 21910161.

43. Melman YF, Shah R, Danielson K, Xiao J, Simonson B, Barth A, et al. Circulating MicroRNA-30d Is Associated With Response to Cardiac Resynchronization Therapy in Heart Failure and Regulates Cardiomyocyte Apoptosis: A Translational Pilot Study. Circulation. 2015; 131(25):2202–16. doi: 10.1161/CIRCULATIONAHA.114.013220 PMID: 25995320; PubMed Central PMCID: PMC4479965.

44. Chen Y, Wakili R, Xiao J, Wu CT, Luo X, Clauss S, et al. Detailed characterization of microRNA changes in a canine heart failure model: Relationship to arrhythmogenic structural remodeling. J Mol Cell Cardiol. 2014; 77:113–24. doi: 10.1016/j.yjmcc.2014.10.001 PMID: 25315712.

45. Genini S, Guziewicz KE, Beltran WA, Aguirre GD. Altered miRNA expression in canine retinas during normal development and in models of retinal degeneration. BMC Genomics. 2014; 15:172. doi: 10.1186/1471-2164-15-172 PMID: 24581223; PubMed Central PMCID: PMC4029133.

46. Zhao FR, Su S, Zhou DH, Zhou P, Xu TC, Zhang LQ, et al. Comparative analysis of microRNAs from the lungs and trachea of dogs (Canis familiaris) infected with canine influenza virus. Infect Genet Evol. 2014; 21:367–74. doi: 10.1016/j.meegid.2013.11.019 PMID: 24333372.

47. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014; 47:11 2 1–2 34. doi: 10.1002/0471250953.bi1112s47 PMID: 25199790; PubMed Central PMCID: PMC4213956.

48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215(3):403–10. doi: 10.1016/S0022-2836(05)80360-2 PMID: 2231712.

49. Prufer K, Stenzel U, Dannemann M, Green RE, Lachmann M, Kelso J. PatMaN: rapid alignment of short sequences to large databases. Bioinformatics. 2008; 24(13):1530–1. doi: 10.1093/bioinformatics/btn223 PMID: 18467344; PubMed Central PMCID: PMC2718670.

50. Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans Comput Biol Bioinform. 2013; 10(3):645–56. doi: 10.1109/TCBB.2013.68 PMID: 24091398.

51. Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxevanis AD. MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore Mnemiopsis leidyi. BMC Genomics. 2012; 13:714. doi: 10.1186/1471-2164-13-714 PMID: 23256903; PubMed Central PMCID: PMCPMC3563456.

52. Ladewig E, Okamura K, Flynt AS, Westholm JO, Lai EC. Discovery of hundreds of mirtrons in mouse and human small RNA data. Genome Res. 2012; 22(9):1634–45. doi: 10.1101/gr.133553.111 PMID: 22955976; PubMed Central PMCID: PMCPMC3431481.

53. Rosenkranz D, Zischler H. proTRAC—a software for probabilistic piRNA cluster detection, visualization and analysis. BMC Bioinformatics. 2012; 13:5. doi: 10.1186/1471-2105-13-5 PMID: 22233380; PubMed Central PMCID: PMC3293768.

54. Lulli V, Romania P, Morsilli O, Cianciulli P, Gabbianelli M, Testa U, et al. MicroRNA-486-3p regulates gamma-globin expression in human erythroid cells by directly modulating BCL11A. PLoS One. 2013; 8 (4):e60436. doi: 10.1371/journal.pone.0060436 PMID: 23593217; PubMed Central PMCID: PMCPMC3617093.

55. Wang LS, Li L, Li L, Chu S, Shiang KD, Li M, et al. MicroRNA-486 regulates normal erythropoiesis and enhances growth and modulates drug response in CML progenitors. Blood. 2015; 125(8):1302–13. doi: 10.1182/blood-2014-06-581926 PMID: 25515961; PubMed Central PMCID: PMCPMC4335083.

56. Chen Y, Sinha K, Perumal K, Reddy R. Effect of 3' terminal adenylic acid residue on the uridylation of human small RNAs in vitro and in frog oocytes. RNA. 2000; 6(9):1277–88. PMID: 10999605; PubMed Central PMCID: PMCPMC1370001.

57. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S, Baba T, et al. Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes Dev. 2009; 23(4):433–8. doi: 10.1101/gad.1761509 PMID: 19240131; PubMed Central PMCID: PMCPMC2648654.

58. Brown M, Suryawanshi H, Hafner M, Farazi TA, Tuschl T. Mammalian miRNA curation through next-generation sequencing. Front Genet. 2013; 4:145. doi: 10.3389/fgene.2013.00145 PMID: 23935604; PubMed Central PMCID: PMCPMC3731538.

59. Hofacker IL. RNA secondary structure analysis using the Vienna RNA package. Curr Protoc Bioinformatics. 2009; Chapter 12:Unit12 2. doi: 10.1002/0471250953.bi1202s26 PMID: 19496057.

60. Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. EMBO Rep. 2011; 12(2):172–7. doi: 10.1038/embor.2010.191 PMID: 21212805; PubMed Central PMCID: PMC3049427.

61. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009; 10 (2):126–39. doi: 10.1038/nrm2632 PMID: 19165215.

62. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011; 12(4):246–58. doi: 10.1038/nrm3089 PMID: 21427766.

63. Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. Development. 2008; 135 (1):3–9. doi: 10.1242/dev.006486 PMID: 18032451.

64. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009; 10 (2):94–108. doi: 10.1038/nrg2504 PMID: 19148191; PubMed Central PMCID: PMCPMC2724769.

65. Ha H, Song J, Wang S, Kapusta A, Feschotte C, Chen KC, et al. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. BMC Genomics. 2014; 15:545. doi: 10.1186/1471-2164-15-545 PMID: 24981367; PubMed Central PMCID: PMCPMC4094622.

66. Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. Trends Genet. 2008; 24(10):489–97. doi: 10.1016/j.tig.2008.07.004 PMID: 18778868.

67. Koppers-Lalic D, Hackenberg M, Bijnsdorp IV, van Eijndhoven MA, Sadek P, Sie D, et al. Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. Cell Rep. 2014; 8 (6):1649–58. doi: 10.1016/j.celrep.2014.08.027 PMID: 25242326.

68. Dai ZH, Chen LB. [The impact of microRNAs on the evolution of metazoan complexity]. Yi Chuan. 2010; 32(2):105–14. PMID: 20176553.