Supplemental Methods

Bioinformatics analysis of sequencing data. The two samples described in **Figure 1** (i.e., the pool of 473 synthetic miRNAs and small RNAs from normal prostate tissue) each had three technical replicates sequenced in different lanes using the Illumina Genome Analyzer platform. The three lanes of synthetic miRNA pool produced a total of 6,429,663 raw reads, and the three lanes of normal prostate tissue produced a total of 10,605,938 raw reads. Reads with either a sequencing error (designated by a "." in the 35 nt read), or a mononucleotide run of adenosines were filtered out, as well as sequences not matching the first six nucleotides (CTGTAG) of the 17 nt Illumina 3' linker. The first six nucleotides were used due to the degradation of sequence quality at the 3' end in Illumina sequences after about the 32nd nucleotide. This filtering step left us with 4,946,602 synthetic and 9,417,361 prostate high-quality reads carried through for further analysis.

To compare miRNAs common to both datasets, we started with the list of 473 spiked-in synthetic miRNAs. Because these were synthesized based on Release 8.0 of miRBase, and some canonical nucleotide sequences have changed from release to release, we limited our list to those miRNAs that remained the same through Release 12.0 of miRBase. We then required a miRNA to have been detected at an abundance of at least 10 reads in both datasets, giving us a list of 167 shared, abundant miRNAs.

Using this list of 167 miRNAs, we then calculated the total matching reads (TMR) and fraction of total reads with 3' additions (%NTA) for each miRNA. In calculating the TMR, the canonical miRNA sequence was compared to all the reads in the dataset and any read which exactly matched the canonical sequence, or matched with trailing nucleotides, was counted towards the TMR. Note that this includes all additions be they templated (matching the miRNA precursor sequence), or not. This method also counts sequence that might have a 5' addition, which were infrequent and not examined further in this study. The %NTA was calculated to be the fraction of the TMR that had non-templated 3' nucleotide additions beyond the canonical miRNA sequence (i.e. those not matching the precursor sequence). Both the prostate tissue and the synthetic miRNA samples had an additional 40 non-human synthetic miRNAs spiked in. This set contained 26 *C. elegans*, 10 *O. sativa*, and 4 *A. thalania* miRNAs listed in the table below.

For these miRNAs, all additions would be non-templated additions so all sequences with 3' additions would count towards the %NTA (in contrast to the human miRNAs).

Table of 40 non-human spiked-in synthetic oligonucleotide sequences.

miRNA species	miRNA name	Sequence
A. thaliana	ath-miR-402	TTCGAGGCCTATTAAACCTCTG
A. thaliana	ath-miR-406	TAGAATGCTATTGTAATCCAG
A. thaliana	ath-miR-413	ATAGTTTCTCTTGTTCTGCAC
A. thaliana	ath-miR-416	GGTTCGTACGTACACTGTTCA
C. elegans	cel-miR-237	TCCCTGAGAATTCTCGAACAGCTT
C. elegans	cel-miR-238	TTTGTACTCCGATGCCATTCAGA
C. elegans	cel-miR-248	TACACGTGCACGGATAACGCTCA
C. elegans	cel-miR-250	TCACAGTCAACTGTTGGCATGG
C. elegans	cel-miR-254	TGCAAATCTTTCGCGACTGTAGG
C. elegans	cel-miR-258	GGTTTTGAGAGGAATCCTTTT
C. elegans	cel-miR-259	AAATCTCATCCTAATCTGGTA
C. elegans	cel-miR-261	TAGCTTTTTAGTTTTCACG
C. elegans	cel-miR-262	GTTTCTCGATGTTTTCTGAT
C. elegans	cel-miR-267	CCCGTGAAGTGTCTGCTGCA
C. elegans	cel-miR-2	TATCACAGCCAGCTTTGATGTGC
C. elegans	cel-miR-357	TAAATGCCAGTCGTTGCAGGA
C. elegans	cel-miR-37	TCACCGGGTGAACACTTGCAGT
C. elegans	cel-miR-392	TATCATCGATCACGTGTGATGA
C. elegans	cel-miR-39	TCACCGGGTGTAAATCAGCTTG
C. elegans	cel-miR-40	TCACCGGGTGTACATCAGCTAA
C. elegans	cel-miR-41	TCACCGGGTGAAAAATCACCTA
C. elegans	cel-miR-50	TGATATGTCTGGTATTCTTGGGTT
C. elegans	cel-miR-51	TACCCGTAGCTCCTATCCATGTT
C. elegans	cel-miR-54	TACCCGTAATCTTCATAATCCGAG
C. elegans	cel-miR-55	TACCCGTATAAGTTTCTGCTGAG
C. elegans	cel-miR-59	TCGAATCGTTTATCAGGATGATG
C. elegans	cel-miR-65	TATGACACTGAAGCGTAACCGAA
C. elegans	cel-miR-66	CATGACACTGATTAGGGATGTGA
C. elegans	cel-miR-70	TAATACGTCGTTGGTGTTTCCAT
C. elegans	cel-miR-75	TTAAAGCTACCAACCGGCTTCA
O. sativa	osa-miR-413	CTAGTTTCACTTGTTCTGCAC
O. sativa	osa-miR-414	TCATCCTCATCATCATCGTCC
O. sativa	osa-miR-416	TGTTCGTCCGTACACTGTTCA
O. sativa	osa-miR-420	TAAATTAATCACGGAAATGAT
O. sativa	osa-miR-426	TTTTGGAAGTTTGTCCTTACG
O. sativa	osa-miR-435	TTATCCGGTATTGGAGTTGA
O. sativa	osa-miR-437	AAAGTTAGAGAAGTTTGACTT
O. sativa	osa-miR-438	TTCCCACGCGTTATAGTGAAA
O. sativa	osa-miR-440	AGTGTCTCCTGATGATCGGGACAA
O. sativa	osa-miR-442	TGACGTGTAAATTGCGAGACGAAT

For the rest of the analyses, a variety of next generation sequencing datasets spanning various disease states, tissue types, and species were analyzed. These datasets and their GEO accession numbers are listed in **Supplemental Table S1**. For the mouse sequencing datasets, we downloaded 6 lanes of Illumina sequencing data for 47 samples totaling 11,408,875 reads. The reads were deconvoluted by barcode from sequencing lanes into biological samples, and then biological replicates were combined into 11 sequencing datasets (**Supplemental Table S1,** m1-m11). For the 12 *C. elegans* sequencing datasets, the datasets were downloaded from GEO already filtered and processed into unique sequences and read counts. The TMR and %NTA for the mouse and *C. elegans* miRNAs were calculated in the same way as for the human datasets. The TMR and %NTA for the human, mouse and *C. elegans* sequencing datasets are given in **Supplemental Tables S3, S4, and S5**, respectively. When plotting the histograms in **Figure 2** and **Supplemental Fig. S2**, all miRNAs in all datasets for all species were required to have a minimum of 10 reads to be included in the %NTA histogram. When constructing the boxplots in **Figure 3**, a miRNA was required to have at least 40 reads in 5 out of 9 datasets (human), 5 out of 11 datasets (mouse) or 5 out of 12 datasets (*C. elegans*). The resulting list of miRNAs for each species was then sorted by mean %NTA and the bottom 10 and top 10 miRNAs were plotted in **Figure 3**. For the pie charts in **Figure 4** and **Supplemental Fig S3**, the fraction of total additions for each nucleotide was calculated on a per-miRNA basis, and then averaged over all the miRNAs in the dataset. MicroRNAs that had 10 or more reads (but no minimum %NTA) were included in the calculation of nucleotide distribution.