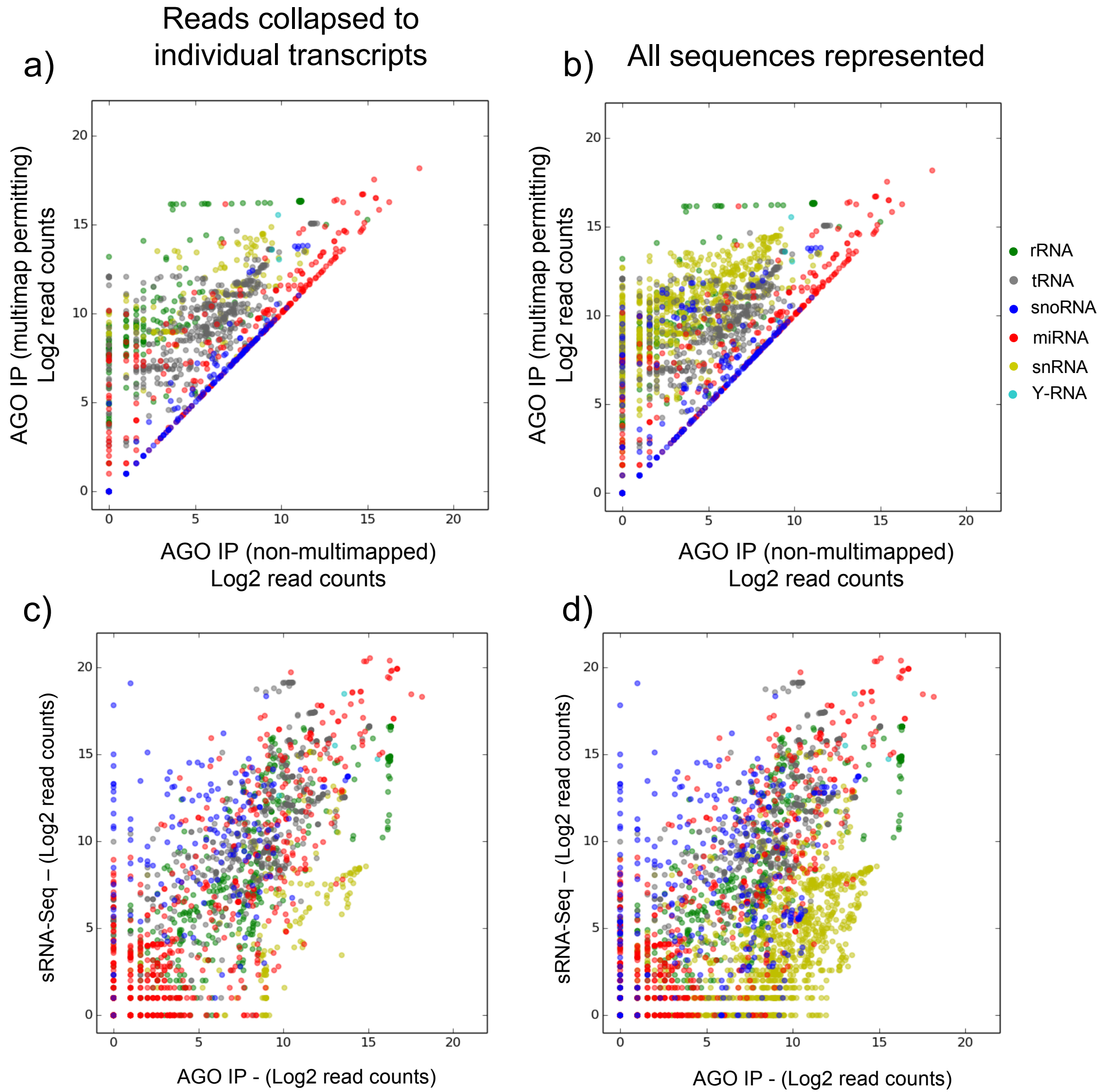
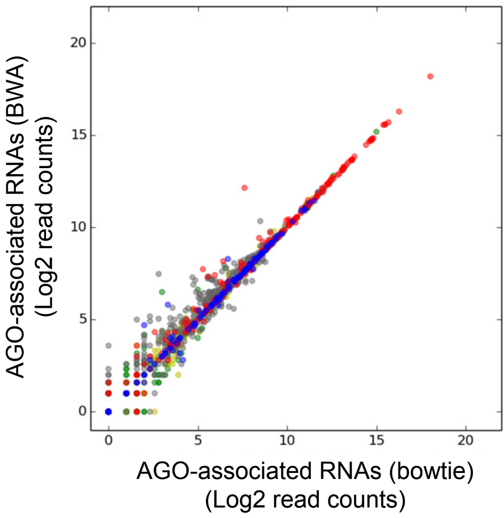
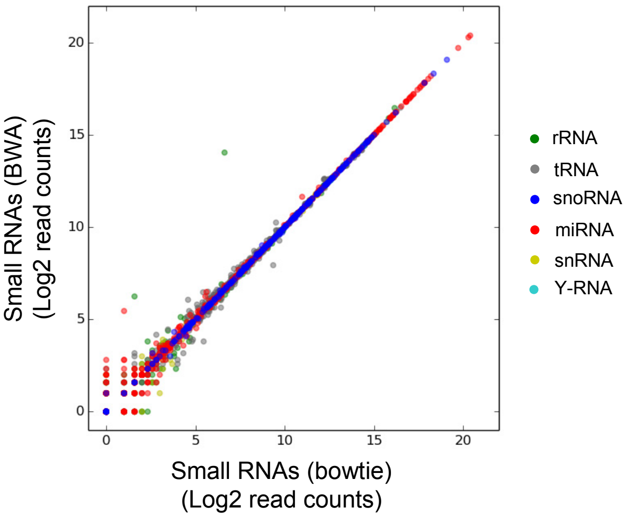
Supplementary Figure 1

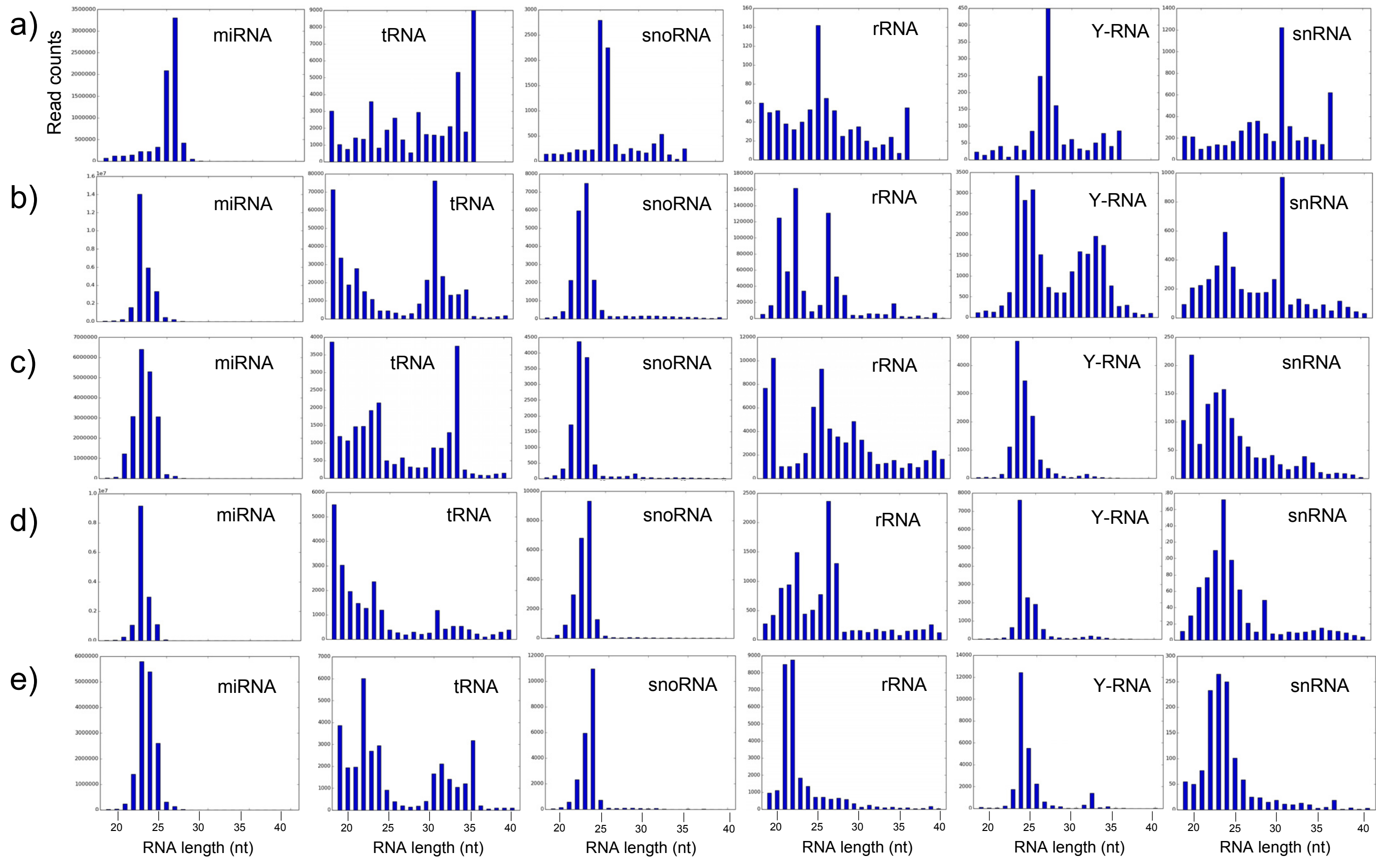


Supplementary Figure 2

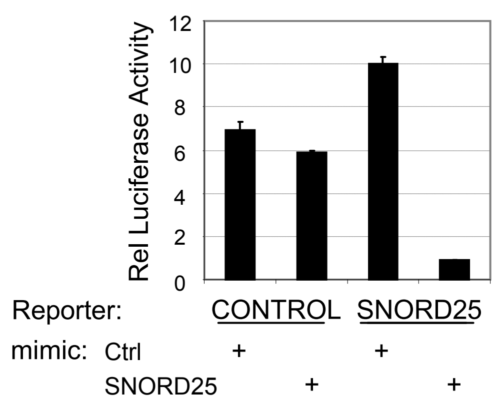




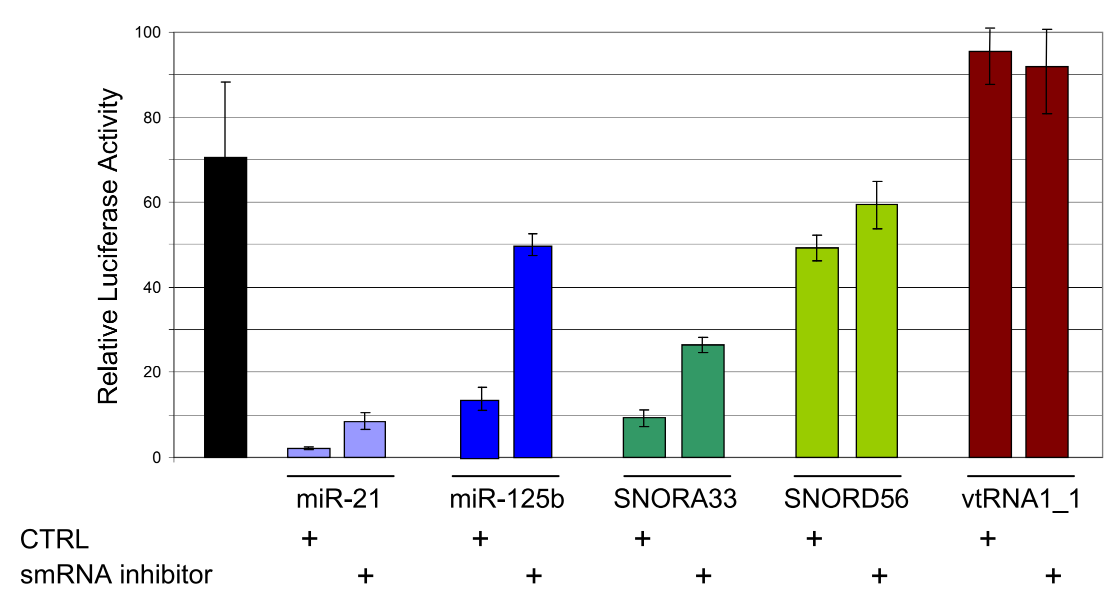
Supplementary Figure 3



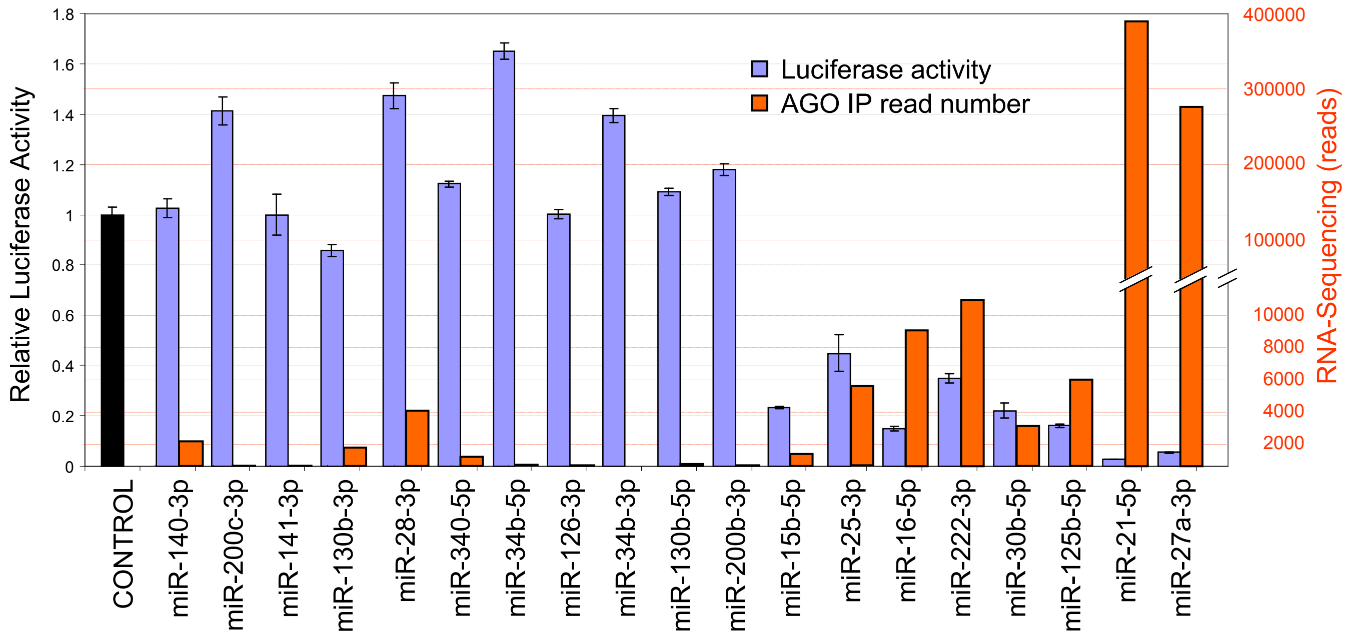
Supplementary Figure 4



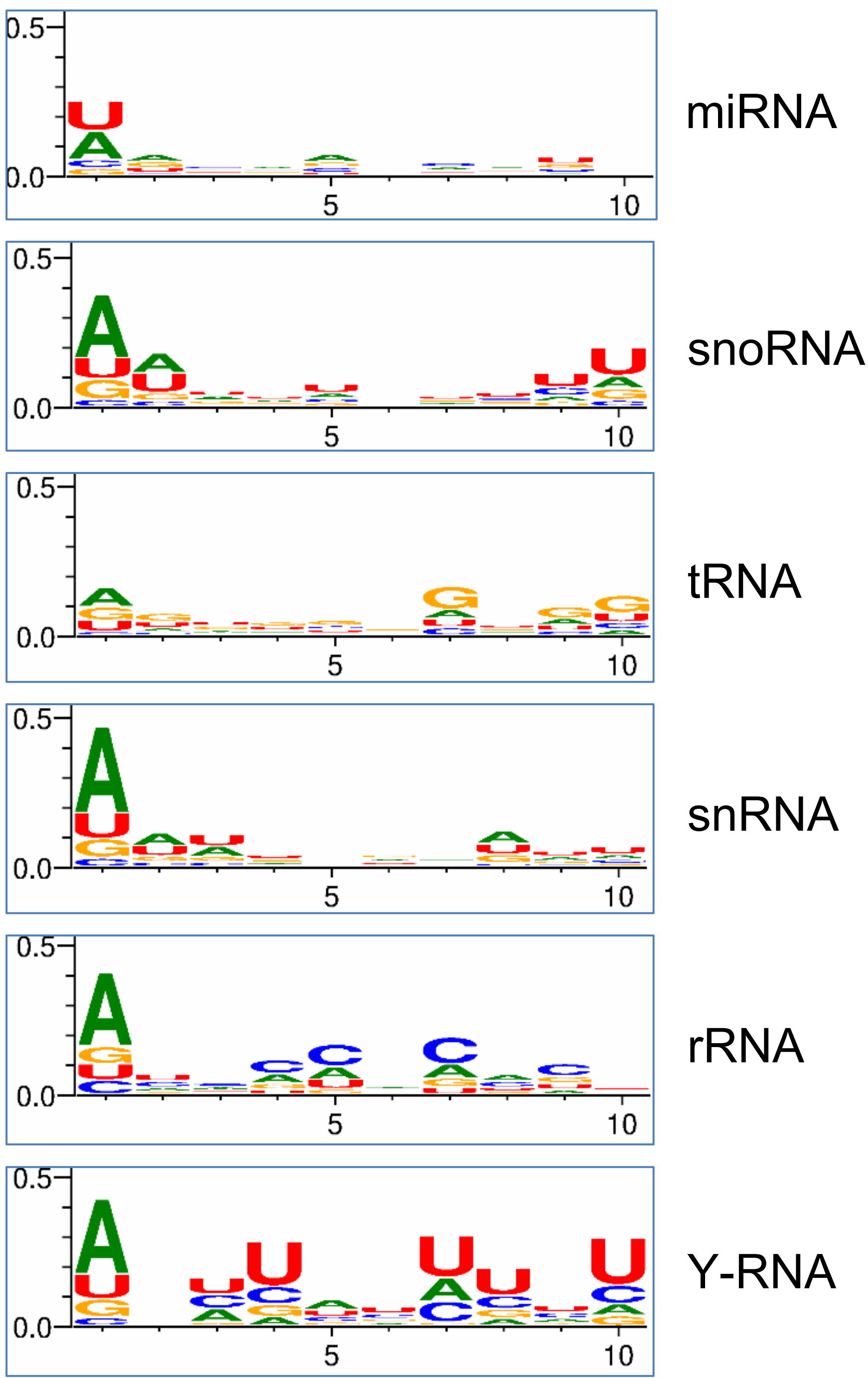
Supplementary Figure 5



Supplementary Figure 6

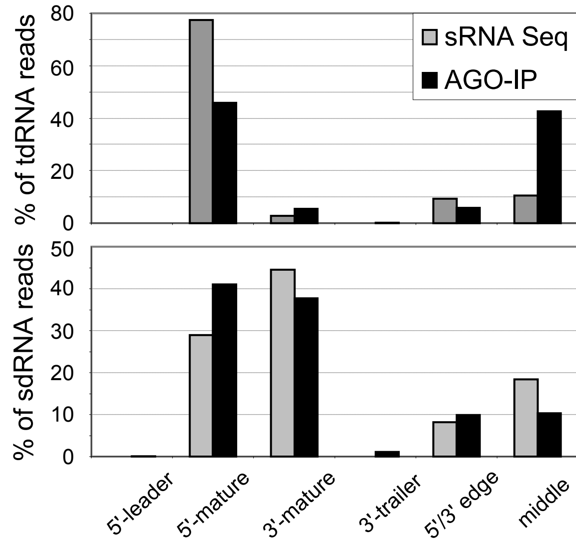


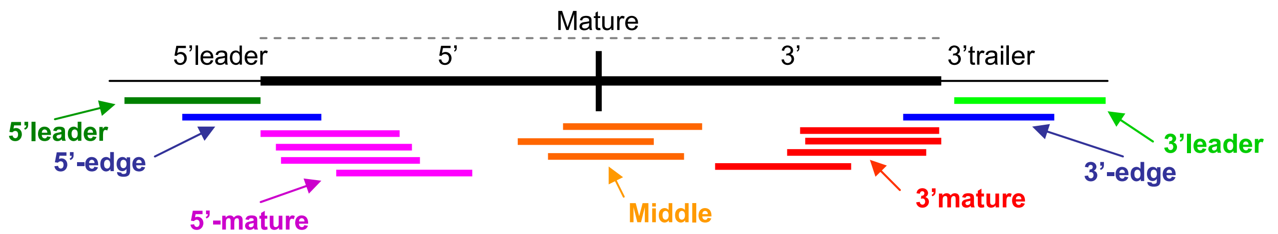
Supplementary Figure 7



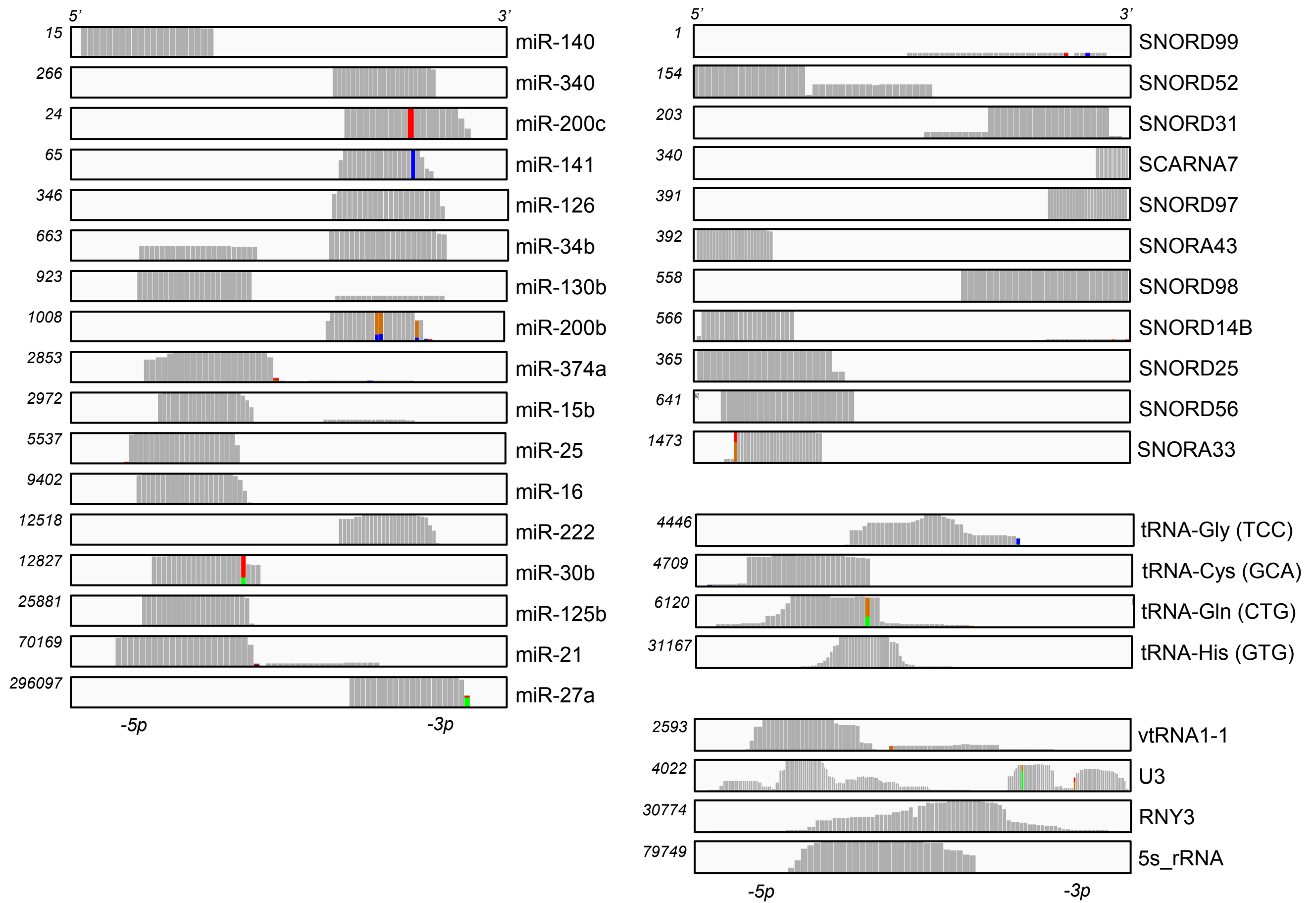
Supplementary Fig 8

		5'-leader	5'-mature	3'-mature	3'-trailer	5'/3' edge	middle
tRNA	sRNA Seq	725	894275	31797	701	105483	122043
	AGO IP	67	47385	5716	126	5933	44011
snoRNA	sRNA Seq	4	124657	191499	22	35152	78970
	AGO IP	34	15114	13881	413	3640	3788





Supplementary Fig 9



Supplementary Figure 1. Representation of all ncRNA sequences irrespective of their genomic loci

Figures 1a and 5a are shown again here in (a) and (c). a,b) Transcript levels of AGO-associated small RNA genes are compared when determined using methods allowing read multimapping (vertical axis) or default alignment parameters (horizontal axis). In (a), one representative transcript is chosen per gene, for (b), all transcripts and loci are shown for each gene. c) Transcript levels of small RNA genes in the cell compared to those in the AGO immunoprecipitate. For (a), one representative transcript is shown per gene, for (b), all transcripts and loci are shown for each gene. For genes with highly similar sequences where multimapping is an consideration (e.g. snRNAs are encoded at multiple loci per gene, tRNA genes have high sequence similarity), a common problem caused by using default mapping parameters is underestimation of the possible contribution of these genes to the transcript pool as reads are divided between all valid loci or genes. Allowing multimapping addresses this by ensuring the transcript levels at each locus or gene are calculated as the maximum expression level possible for that gene. Whilst this inevitably results in over-estimation of the transcript levels at many genes and loci, it makes it possible to identify individual genes in multimapping groups which may be responsible for significant contributions to the transcript pool, a feat which is not possible using normal parameters.

Supplementary Figure 2. Default parameters for different sequence aligners yield similar results

For Bowtie (horizontal axis) and BWA (vertical axis) sequence alignment software, read counts obtained using default parameters were compared for both (a) the AGO immunoprecipitation and (b) whole cell RNA sequencing. For virtually all transcripts, the default parameters of each program produced highly similar results.

Supplementary Figure 3.

AGO binding profiles of small RNAs are shown, with the vertical axis representing read counts and the horizontal axis, the length of the small RNA transcript. All data were re-analysed from published datasets: a) derived from THP-1 cells (21), b-e) derived from HEK-293T, A549, SY5Y, C8166 cells (46). The small RNA : AGO immunoprecipitation data we present in this paper was prepared using the HITS-CLIP methodology in which AGO : miRNA : mRNA complexes are UV-crosslinked and longer associated mRNAs partially digested to more accurately identify the miRNA binding site (to within ~ 40-60 nucleotides). AGO-associated small RNAs shown here were purified

independently of cross-linking and enzymatic digestion. Regardless of the methodology used, an abundance of 18-30nt RNAs from diverse sources are endogenously co-purified with AGO.

Supplementary Figure 4. Over-expression of a sdRNA mimic indicates miRNA-like functionality

MDA-MB-231 cells were transfected with either a control luciferase reporter or a luciferase reporter containing a single fully complementary site to a small RNA derived from SNORD25. A mimic of this SNORD25-derived small RNA (20nM) was co-transfected with the reporter and relative luciferase activity determined 48 hours post-transfection. Luciferase assays were performed as biological triplicates with error bars depicting standard error of mean.

Supplementary Figure 5. Sequestering small RNA inhibitors increase miR-21 and SNORA33 reporter activity

MDA-MB-231 cells were initially transfected with sequestering small RNA inhibitors (30uM), then 24 hours later transfected again with additional inhibitors (30uM) and luciferase reporter vectors (100nM) as indicated. 24 hours later cells were lysed and relative luciferase activity determined. Luciferase assays were performed in triplicate with error bars depicting standard deviation.

Supplementary Figure 6. The abundance of miRNAs within cells partially indicates functionality

Relative luciferase activity for a panel of established miRNAs were determined as with Figure 2 and graphed alongside the number of sequence reads for these miRNAs from whole cells. In figure 2, when luciferase activity was represented along with abundance in the AGO immunoprecitation, a strong and consistent correlation was seen between AGO association and the degree of reporter suppression. This relationship is less consistent with total miRNA expression, as exemplified through strong reporter repression by modestly expressed miRNAs (ie: miR-15b-5p) and the absence of repression for miRNAs detected at reasonably high levels (ie: miR-140-3p).

Supplementary Figure 7. A/U 5’ nucleotide bias in AGO-bound small RNAs

Nucleotide level abundance (for the first 10 nucleotides from the 5’ end) was determined for the different classes of small RNAs immunoprecipitated with AGO. Only unique molecules were analysed to avoid possible skewing of data by very highly abundant sequences. Each class of small RNA (with the exception of tRNAs) display an A and / or U 5’ bias.

Supplementary Figure 8. Location of snoRNA and tRNA-derived small RNAs

The location of small RNAs (relative to the parent ncRNA) are shown for both snoRNA and tRNA-derived reads in both the whole cell and the AGO immunoprecipitate. Read numbers are shown in the table (above) and percentages of snoRNA or tRNA-derived reads are represented in the histogram. For sequences to be classified 5’ or 3’-mature, the entire length of the small RNA must fall within the 5’ or 3’ half of the mature ncRNA. The vast majority of these fragments start or end at, or within several nucleotides of, the mature 5’ or 3’ termini. Reads attributed to the “5’/3’ edge” extend beyond the annotated mature termini into the precursor leader or trailer sequences. “5’-leader” or “3’-trailer” sequences map entirely within this region, though the numbers of these fragments are most likely underestimated given the typically short length of these sequences (often ~ 8-12 nucleotides) which were discarded due to the promiscuity of the mapping of such short sequences. “Middle” reads map across the 5’ and 3’ halves of the mature ncRNA.

Supplementary Figure 9. Profiles of various AGO-associated small RNAs

Read abundance (vertical axis) after AGO immunoprecipitation is shown for all small RNAs for which reporter assays have been conducted. The position along the transcript is represented on the horizontal axis plotted 5’ – 3’. Small AGO-associated RNAs of miRNA or snoRNA origin have typically well defined termini, whilst small RNAs of other origin have typically poorly defined ends.