Supplemental methods

Analysis of sequencing data

Individual sequence reads with the base quality scores were produces by Illimina/Solexa. The identical sequence reads were counted to produce a 'read count' score. Then all duplicated sequences were eliminated from the initial dataset to produce a non-redundant set of the unique sequences, hereafter referred to as sequence tags. After trimming the 3' adaptor sequence (5'-TCGTATGCCGTCTTCTGCTTG-3') sequence tags were mapped onto chicken genome assembly galGal3 (May 2006), using BLAT software (Kent 2002) with the following parameters: -noHead -minMatch=1 -oneOff=1 -minIdentity=100 -tileSize=12. All partial alignments and alignments containing gaps were eliminated from further analyses.

To identify sequence tags originating from coding exons, repeats, rRNA, tRNA, snRNA and snoRNA we used UCSC "RefGene", "RepeatMasker" and NCBI "RefSeq" data (Karolchik et al. 2003; Pruitt and Maglott 2001), as well as our own set of ncRNA annotations compiled from the NCBI GenBank data (http://www.ncbi.nlm.nih.gov/). Sequence tags matching parts of chicken rRNA, tRNA, snRNA, and snoRNA sequences were recorded and excluded from further analyses.

For RNA structure prediction the genomic coordinates of the sequence tags were extended by 60 nt at their 5' and 3' ends, and the corresponding genomic sequences were retrieved and used as an input for the RNAfold software (Hofacker 2003). To identify potential miRNA precursors in the set of the predicted RNA structures we applied a filtering mechanism based on sequence, structural features, and minimum free energy of the known vertebrate miRNA precursors present in miRBase (Griffiths-Jones et al. 2006). In brief, we eliminated all predicted hairpin-like structures that were shorter than 53 nt in length, or had either of the following: a terminal loop larger than 22 nt, bulges larger than 18 nt, runs of more than 7 'Gs', runs of more than 8 'Cs', runs of more than 13 'As', runs of more than 11 'U/Ts', or had a minimum free energy above -15 kcal/mol. In addition, all remaining pre-miRNA candidates were analyzed for distribution of the sequence tags to ensure that the RNA hairpin structure and positions of sequence tags meet basic requirements of the Drosha/Dicer pre-miRNA processing complex.

Analysis of the evolutionary conservation of the newly identified miRNAs

The sequences of the predicted RNA hairpin structures identified as pre-miRNA candidates were used to search human (hg18), dog (canFam2), opossum (monDom4), zebra fish (danRer4), xenopus (xenTro2), and lizard (anoCar1) genomes using BLAT (Kent 2002) with the following parameters: -noHead -minMatch=1 -oneOff=1 -minIdentity=85 -tileSize=8. Sequence alignments covering at least 90% of the length of chicken pre-miRNA were considered as potential orthologs and used in further RNA secondary structure analyses. To identify potential avian-specific miRNA we searched sequence traces of the Zebra finch (*Taeniopygia guttata*) in the NCBI trace archive using BLAST. Zebra finch genomic traces that had sequence similarity of at least 90% to the corresponding chicken sequence and covered at least 90% of the sequence length were considered as candidate sequences.

Orthologous candidate sequences identified in human, dog, opossum, zebra fish, xenopus, lizard and zebra finch were used to predict RNA secondary structure as described above. New chicken miRNA sequences that did not have detectable orthologous sequences in any of the analyzed non-avian genomes using the described method, were considered as potential avian-specific miRNAs.

Analysis of the relative abundance of miRNAs in three RNA libraries

Similar to credibility interval approaches (as opposed to hypothesis testing) reported for the analysis of SAGE data (Vencio et al. 2003), we employed a semi-parametric approach to identify miRNAs showing statistically significant difference in relative abundance (as reflected by total count of individual sequence reads) between the three embryonic small RNA libraries. Firstly, for each sample, the abundance of each miRNA was expressed in reads per million (rpm) and this value was log-transformed to approximate a Gaussian distribution. Secondly, for each miRNA, the observed range in log-transformed rpm value across the three samples was recorded. Finally, miRNA with range values lying beyond three standard errors of the mean of ranges were deemed to be significantly different between the samples at $P < 0.01$.

Additional references

Griffiths-Jones, S., R.J. Grocock, S. van Dongen, A. Bateman, and A.J. Enright. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* **34:** D140-D144.

Hofacker, I.L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31:** 3429-3431.

Karolchik, D., R. Baertsch, M. Diekhans, T.S. Furey, A. Hinrichs, Y.T. Lu, K.M. Roskin, M. Schwartz, C.W. Sugnet, D.J. Thomas et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Research* **31:** 51-54.

Kent, W.J. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* **12:** 656-664.

Pruitt, K.D. and D.R. Maglott. 2001. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Research* **29:** 137-140.

Vencio, R.Z., H. Brentani, and C.A. Pereira. 2003. Using credibility intervals instead of hypothesis tests in SAGE analysis. *Bioinformatics* **19:** 2461-2464.