May 2007

MicroRNA genes are frequently located near mouse cancer susceptibility loci.

Cinzia Sevignani
Thomas Jefferson University, cinzia.sevignani@jefferson.edu

George A. Calin
Ohio State University

Stephanie C. Nnadi
Thomas Jefferson University

Masayoshi Shimizu
Ohio State University

Ramana V. Davuluri
Ohio State University

See next page for additional authors

Authors
Cinzia Sevignani, George A. Calin, Stephanie C. Nnadi, Masayoshi Shimizu, Ramana V. Davuluri, Terry Hyslop, Peter Demant, Carlo M. Croce, and Linda D. Siracusa

MicroRNA Genes are Frequently Located near Mouse Cancer Susceptibility Loci

Cinzia Sevignani [1], George A. Calin [2]*, Stephanie C. Nnadi [1], Masayoshi Shimizu [2], Ramana V. Davuluri[2], Terry Hyslop [3], Peter Demant [4], Carlo M. Croce [2] & Linda D. Siracusa [1#]*

[1] Department of Microbiology and Immunology, Kimmel Cancer Center, Jefferson Medical College, Philadelphia, PA 19107.

[2] Department of Molecular Virology, Immunology and Medical Genetics and Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210.

[3] Department of Pharmacology and Experimental Therapeutics, Jefferson Medical College, Philadelphia, PA 19107.

[4] Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY 14263

#Correspondence to: Linda D. Siracusa at Kimmel Cancer Center, Thomas Jefferson University, 233 South 10th Street, BLSB 719, Philadelphia, PA 19107-5541 215-503-4536, Siracusa@mail.jci.tju.edu

* Note: C. Sevignani and G. A. Calin contributed equally to this work.

ABSTRACT

MicroRNAs (miRNAs) are short 19-24 nucleotide RNA molecules that have been shown to regulate the expression of other genes in a variety of eukaryotic systems. Abnormal expression of miRNAs has been observed in several human cancers, and furthermore, germline and somatic mutations in human miRNAs were recently identified in patients with chronic lymphocytic leukemia. Thus, human miRNAs can act as tumor suppressor genes or oncogenes, where mutations, deletions or amplifications can underlie the development of certain types of leukemia. In addition, previous studies have shown that miRNA expression profiles can distinguish among human solid tumors from different organs. Since a single miRNA can simultaneously influence the expression of two or more protein-coding genes, we hypothesized that miRNAs could be candidate genes for cancer risk. Research in complex trait genetics has demonstrated that genetic background determines cancer susceptibility or resistance in various tissues, such as colon and lung, of different inbred mouse strains. We compared the genome positions of mouse tumor susceptibility loci with those of mouse miRNAs. Here we report a statistically significant association between the chromosomal location of miRNAs and those of mouse cancer susceptibility loci that influence the development of solid tumors. Furthermore, we identified distinct patterns of flanking DNA sequences for several miRNAs located at or near susceptibility loci in inbred strains with different tumor susceptibilities. These data provide the first catalog of miRNA genes in inbred strains that could represent genes involved in the development and penetrance of solid tumors.

INTRODUCTION

The genetics of cancer penetrance is an evolving field that is leading to new strategies for cancer prevention and treatment. Susceptibility loci can influence not only cancer incidence and metastasis, but also the penetrance of other genetic disorders as well as susceptibility to infectious diseases (1-2, www.informatics.jax.org). The penetrance of tumor phenotypes results from the interaction between genetics and environment. The variability of environmental

factors and the genetic heterogeneity in human populations complicates the search for susceptibility loci. To control these factors, inbred mouse strains are useful models to search for allelic variants that confer susceptibility or resistance to different types of tumors (3). Quantitative trait loci (QTLs) analyses in mice have already identified chromosomal regions involved in cancer (3-5), and several protein-coding genes have a proven role as cancer susceptibility loci (6-10). We suggest that a newly discovered family of noncoding genes, the miRNAs, could be responsible for susceptibility to solid tumors.

Several papers have shown that miRNAs play an important role in different cancers, including B cell chronic lymphocytic leukemias (11), Burkitts lymphoma, colon cancer, lung cancer, and breast cancer (12-14). Furthermore, miRNA expression profiles can be used to classify human solid tumors and leukemias (10, 15). Some miRNAs can downregulate large numbers of target mRNAs (16), and it is estimated that miRNAs could regulate 33% of the human genome (17). These targets may be oncogenes and tumor suppressor genes, supporting the idea that miRNAs could influence susceptibility and predispose individuals to different cancers. Because miRNAs are responsible for fine regulation of gene expression, “tuning” cellular phenotypes during delicate processes like development and differentiation, they represent a family of candidate genes for cancer susceptibility. Here, we evaluate the correlation of miRNAs with cancer susceptibility loci and identify miRNAs with a distinct pattern of flanking DNA regions in inbred mouse strains with diverse tumor phenotypes. Our work supports the hypothesis that miRNAs could represent a new family of susceptibility genes affecting the risk of cancer.

RESULTS AND DISCUSSION

A genome-wide approach was used to investigate associations between miRNAs and mouse cancer susceptibility loci. This type of investigation was previously successful in identifying a significant association between the genomic position of human miRNAs and that of fragile sites (FRAs) and cancer-associated genomic regions (CAGRs) (18).

Compilation of known tumor susceptibility and modifier loci. We constructed a database that contained both susceptibility and modifier loci for 8 types of solid tumors in the mouse: bladder, colon, liver, lung, mammary gland, ovary, small intestine, and skin. The list was built by locus symbol, peak marker, chromosome (Chr), tissue, and strains. To obtain loci positions, we used the Mouse Genome Informatics Database (www.informatics.jax.org) and PubMed (www.pubmed.gov). Published data were evaluated to identify the peak location of each region (**Fig. 1**).

Several observations were derived from this analysis. First, Chr 4 had the largest number of susceptibility loci (13 total), followed by Chr 6 (12 loci). Second, Chr Y had no susceptibility loci, followed by Chr 14 (1 locus). The frequency distribution (**Fig. 4,** published as supporting information on the PNAS website) differs from a normal distribution, suggesting that genes predisposing to solid tumors are not distributed randomly within the genome, but tend to be concentrated on certain chromosomes. Third, **Figure 1** shows that several susceptibility loci identified for the same organ overlap (such as for lung, *Pas6* and *Sluc2* on Chr 2, *Pas1b* and *Sluc3* on Chr 6, *Par1* and *Pas5a* on Chr 11, *Pas2* and *Pas12* on Chr 17, and *Pas3b* and *Sluc1* on Chr 19). Since these tumor susceptibility loci were discovered using different test conditions and strains in different laboratories and yet, reside in close proximity, they may be due to the same gene(s). Fourth, several susceptibility loci identified for different organs overlap (such as *Hcs7/Hcif2* and *Pas8* on Chr 1, *Hcs3* and *Sluc12* on Chr 12, and the 9 loci on distal Chr 4). These findings suggest that genes controlling the integrity and/or homeostasis of multiple solid organs may be concentrated within the genome; there may be only one or multiple, closely linked genes in each of these regions. Alternatively, future studies that define more susceptibility loci in these and other organs may prove otherwise.

Construction of the MUSMIRSUS Database. A total of 229 mouse miRNA gene locations were downloaded from the July 2005 miRNA registry website (microrna.sanger.ac.uk). We combined and compared the positions of mouse miRNAs with those of known susceptibility and modifier loci (www.ensembl.org); this ordered list was used to identify miRNAs that mapped at or near the peak of tumor loci (see **Table 1** as an example). The database was named MUSMIRSUS, for Mus

miRNA susceptibility and is contained in its entirety at www.kimmelcancercenter.org/siracusa/musmirsus.htm.

A significant number of miRNAs are located close to tumor susceptibility loci. To test hypotheses about the relationship of the incidence of miRNAs and their association with tumor susceptibility loci, we used the Random Effect Poisson Regression Model (18). For this analyses, the random effect used was chromosome, in that data within a chromosome is assumed correlated. Under this model, the number of miRNAs defined "events" and non-overlapping lengths of susceptibility regions defined "time". The "length" of a susceptibility region was ±0.5 Mb if the gene is known, or estimated as ±5 Mb from the peak marker. Although a QTL need not be at the peak marker, the critical distance selected here covers the most likely interval. A miRNA was considered within a susceptibility region if it was located ±5 Mb from the peak location of a given locus. The fixed effect in the model consisted of an indicator variable for the presence/absence of each susceptibility region. We report the incidence rate ratio (IRR), 2-sided 95% confidence interval of the incidence rate ratio, and 2-sided p-values for testing the hypothesis that the incident rate ratio is 1.0. An IRR significantly >1 indicates an increase in the number of miRNAs within a region. All statistical computations were completed using STATA v7.0.

Overall, 96 miRNAs (41.9%) were located <5 Mb from the peak of the closest locus (or loci), representing a highly significant association between the incidence of miRNAs and tumor susceptibility regions. Specifically, the relative incidence of miRNAs occurred at a rate 1.5 times higher in susceptibility regions than non-susceptibility regions (IRR=1.63, 95% CI {1.25, 2.14}, $p<0.001$). Therefore, miRNAs were found more often in (or near) susceptibility regions than in other areas of the genome (**Fig. 2**). The estimated rate of miRNAs is moderate in most chromosomes, but quite large in Chrs 2, 6, 7, 13, and especially Chr 12. Thus, a larger number of miRNAs than expected by chance occurs in these 5 chromosomes. Regardless of location, the model indicates that miRNAs are significantly more likely to reside within tumor susceptibility regions than outside them. In addition, a recent report suggests an association between the locations of mouse miRNAs and known sites of retroviral integration in mouse cancers (21).

Specific miRNAs as candidates for tumor susceptibility loci. Thirty-five of the 229 miRNAs (15.3%) are located in close proximity (<2 Mb) to 24 susceptibility loci, some of which overlap with the same miRNA. For example, *mir-30c-1* and *mir-30e* are 0.47 Mb distal to the mammary *Mmtg2* locus, <2.5 Mb proximal to the colon *Scc11* locus, and <3.6 Mb distal to the skin *Skts7* locus on Chr 4. Similarly, *mir-34a* is 0.86 Mb distal to the small intestine *Mom6* locus and <4.6 Mb distal to the ovary *Gct1* locus on Chr 4. In addition, the bladder *Bts2* locus and the lung *Sluc7* locus overlap on Chr 6, and reside 0.4–4.8 Mb from a cluster of 6 miRNAs. These miRNAs are all <1 Mb from at least one susceptibility locus and also reside <5 Mb from a second susceptibility locus for a different organ.

Eighteen mouse miRNAs (7.9%) localized near (<5 Mb) susceptibility loci correspond to human miRNAs located in human cancer associated regions (CAGRs) and fragile sites (FRAs) as previously reported (18). Some of these 18 miRNAs are members of the *let* family (*let-7a-1, let-7a-2, let-7d, let-7f-1*), and reduced expression of *let-7* miRNAs is associated with shortened postoperative survival in human lung cancer (22). Since the expression of *let-7* and RAS protein are opposite in lung tumors versus normal tissue, RAS was shown to be regulated by the *let-7* family (23). In fact, members of *let-7* are located within 4 Mb of loci predisposing to lung cancer: *let-7a-2* at *Sluc10* on Chr 9, *let-7f-2* at *Slt5* on Chr X, and the cluster *let-7d*, *let-7f-1,* and *let7a-1* at *Sluc23* on Chr 13. The list of miRNAs at or near sites commonly altered in human cancers includes *mir-215* and *mir-194-1* which are <0.17 kb from *D1Mit208*, a marker that defines the peak of the colon *Scc3* locus on Chr 1. Homologous human miRNAs are located inside the FRA1H at Chr 1q42.1**;** this fragile site is an HPV16 integration site involved in cervical cancer (18). In addition, *mir-181b-2* and *mir-199b* are 3.3 Mb from the peak of the colon *Scc2* locus on Chr 2; homologous human miRNAs are located inside a CAGR at Chr 9q33-34.1, which is a region commonly deleted in bladder cancer (18).

Some miRNAs can act as oncogenes or tumor suppressor genes (12, 14). If our hypothesis that miRNAs are candidates for tumor susceptibility genes is true, it is expected that some miRNAs will act cell autonomously and be abnormally expressed in cancer cells, whereas others will act non-cell autonomously to

influence cancer development (4). The expression of human homologs of candidate miRNAs for mouse tumor susceptibility loci was extensively reported. Nineteen of the 35 (54.3%) miRNAs located <2 Mb from the peak of a tumor susceptibility locus are abnormally expressed in at least one type of human solid cancer, with expression being significantly different from normal controls (19-20). Studies to discriminate differences in expression of candidate miRNAs between tumor susceptible and resistant strains are underway in our laboratories.

Direct sequence screen of inbred strains. Discrete sequence alterations are frequently found in human cancers (21) and variation in the sequence of precursor miRNAs were identified in the normal human population (24) and in the germline of cancer patients (25). It was shown that a C to A polymorphism in the mature *mir-30c-2* sequence may alter target selection and thus exert profound biological effects (24). According to the miRanda algorithm, 10 of 12 predicted targets were not considered likely for the A-type variant *mir-30c-2.* Interestingly, the mouse homolog is located <1 Mb from the peak marker of the lung *Lscc1* locus on Chr 1.

We sequenced several miRNAs to determine if inbred strains with different tumor susceptibilities possessed unique alleles. Out of 229 miRNAs, 10% are located <1 Mb from the peak marker that defines 15 solid tumor susceptibility loci (see **Table 1** as an example). In fact, 8 of these loci have >1 miRNA that lie within <1 Mb of their peak. Selection of miRNAs was based on criteria that included their proximity to susceptibility loci, the same miRNA being present in human CAGRs, and the available inbred strains. We sequenced 13 miRNAs (*mir-1-1, mir-7-1, mir-10a, mir-30c-2, mir-31, mir-34a, mir-133a-2, mir-196a-1, mir-207, mir-219-1, mir-324, mir-346, mir-448*) comparing susceptible, resistant, and/or common inbred strains.

Genomic regions corresponding to each precursor miRNA were amplified including flanking 5' and 3' ends, because these regions are needed for miRNA expression (26). A highly conserved motif ~200 bp upstream of the miRNA stem-loop was found in most independently (intergenic) transcribed nematode miRNAs (27). Moreover, independent transcription units in which the precursor miRNA is

sufficient for processing have been found in plants, with their promoter elements being located upstream of the stem-loop (28).

Seven of the 13 miRNAs showed sequence differences between inbred strains **(Figs. 3** and **5-8,** published as supporting information on the PNAS website). **Figure 3A** shows that the flanking region of *mir-1-1* (near *Sluc17*) has 6 substitutions upstream of the stem-loop in the susceptible O20 compared to the resistant B10.O20 strain. **Figure 3B** shows a deletion of TCCTTC for *mir-219-1* (near *Pas2/Pas12*) in the susceptible A/J compared with the resistant B6 strain. Furthermore, A/J carries a base substitution upstream and downstream of the stem-loop. **Figure 3C** shows a dramatic difference upstream of the *mir-196a-1* stem-loop (near *Par1/Pas5a*). The susceptible A/J carries the longest stretch of guanines compared with the resistant B6, SM/J, and *Mus spretus* strains. Together with substitutions, the data shows that there are four *mir-196a-1* alleles. In addition, *M. spretus* has the only substitution found within a stem-loop structure.

For *mir-30c-2* (near *Lscc1*), the resistant AKR and B6 strains had different alleles with 6 substitutions and one insertion-deletion (in-del). The susceptible A/J strain has a third allele (**Fig. 5**). For *mir-133a-2* (near *Sluc17*), a deletion of TATATGTA was found in the susceptible O20 compared to the resistant B10.O20 strain. Furthermore, there are four substitutions upstream and two substitutions downstream of the *mir-133a-2* stem-loop **(Fig. 6)**. For *mir-34a* (near *Mom6*), 6 substitutions and one in-del was found between the susceptible B6 and the resistant AKR strains. AKR and C3H had the same allele, B6, BALB and A/J had a second allele, and CAST had a third allele (**Fig. 7).** For *mir-10a* (near *Par1/Pas5a*), an in-del upstream of the stem-loop was detected in the susceptible A/J compared to the resistant B6 and SM/J strains (as well as BALB). *M. spretus* and CAST had additional unique alleles (**Fig. 8**).

Computational prediction of miRNA promoters: Since only a few miRNA promoters have been identified experimentally (30-32), we used an automated computational pipeline to determine miRNA promoters, depending on their relative genomic location to the annotated protein-coding genes. The annotation system combines the genomic alignments of mRNAs and miRNAs with *ab initio* pol II

promoter predictions of FirstEF (33). The genomic sequences and alignments of RefSeq mRNAs and known mRNAs were downloaded from the UCSC website (genome.ucsc.edu). Genomic alignments of miRNAs were determined by the BLAT program (34).

For intragenic miRNAs that overlap with protein-coding genes and in the sense strand, the promoter of the host protein coding gene is considered as the promoter of the miRNA, since the 'host' transcript and miRNAs usually portray similar expression profiles indicating that these miRNAs are transcribed as part of long host transcription units (35-37). For the rest, we ran the FirstEF program on the 5' upstream region of all miRNAs, with cut-off values for posterior probabilities of donor, promoter, and first-exon of 0.4, 0.3, and 0.4, respectively, to predict the pol II promoters. Values of the posterior probabilities range from 0 to 1, where higher probabilities indicate a greater confidence that the corresponding predictions are correct (**Table 3,** published as supporting information on the PNAS website).

Of 7 miRNAs that showed sequence differences between tumor susceptible and resistant strains, 5 miRNAs had changes within their predicted promoter regions (**Fig. 3** and **Table 3**). The majority of alterations resided in the 5' flanking and promoter regions (**Figs. 3** and **5-8**). Most variants were substitutions and in-dels; the in-dels were primarily in simple repeats. Since control of miRNA expression levels resides primarily in the 5' flanking regions, the implications of these findings are that subtle changes in the nucleotide sequence of miRNA promoters could significantly affect expression of miRNA transcripts.

MiRNAs in development, differentiation, and cancer. MiRNAs have significant roles in development and differentiation (38-40). Since a single miRNA can simultaneously target multiple genes, alterations in the amount and/or sequence of a mature miRNA can have significant effects on the expression of target genes. In addition, mutations in the promoter regions of miRNAs could affect the spatial, temporal, and/or absolute levels of miRNAs, without changing the sequence of the mature miRNA. The downstream consequences of such alterations would be manifest as changes in cellular physiology and phenotype. For example, if the fine-tuning of gene expression controlled by miRNAs were subtly altered, leading to

enhanced rates of proliferation and/or decreased rates of apoptosis, such changes could significantly influence an individual's risk of developing cancer over time. Our studies have shown that there is a statistically significant correlation between the map locations of mouse tumor susceptibility loci and the map locations of miRNAs. Furthermore, substitutions and in-dels were predominantly in promoter regions of several miRNAs between susceptible and resistant inbred strains. Taken together, this report supports the idea that miRNAs should be considered candidates for tumor susceptibility loci. MiRNAs could represent a new family of genes involved in the penetrance of human cancer and other diseases. Identification of such genes would be important to clarify the mechanisms involved in tumor initiation, growth, and progression as well as to reveal new markers for molecular diagnosis and novel targets for drug therapy.

MATERIALS AND METHODS

Mice: The A/J, AKR/J (AKR), BALB/cJ (BALB), C57BL/6J (B6), C3H/HeJ (C3H), CAST/Ei (CAST), *M. spretus*, and SM/J inbred strains were purchased from The Jackson Laboratory (Bar Harbor, ME) and maintained at the Thomas Jefferson University Animal Facility. O20/ADem and B10.O20/Dem mice are maintained at the Department of Laboratory Animal Resources of Roswell Park Cancer Institute. Both facilities are AAALAC accredited.

Sequencing: High-molecular weight kidney DNA was isolated using the Qiagen DNeasy Tissue Kit (Valencia, CA). The genomic region of each miRNA was amplified and sequenced using primers listed in **Table 2** (published as supporting information on the PNAS website). Sequence analyses were performed on genomic DNA from two mice from each strain.

ACKNOWLEDGEMENTS

GAC was supported by a Kimmel Foundation Award and a CLL Global Research Foundation Grant. SCN is supported by an NCI Minority Supplement Award. Research supported by NCI P01CA76259 and P01CA81534 to CMC and RO1CA72027 to LDS.

REFERENCES

1. Hunter K, Welch DR, Liu ET (2003) *Nat Genet* 34:23-24.
2. Siracusa LD, Silverman KA, Koratkar R, Markova M, Buchberg A (2004) in *Oncogenomics: Molecular Approaches to Cancer*, eds Brenner C, Duggan D (Wiley-Liss, New Jersey) pp 255-290.
3. Mao J-H, Balmain A (2003) *Curr Opin Genet Dev.* 13:14-19.
4. Demant P (2003) *Nat Rev Genet* 4:721-734.
5. Dragani T (2003) *Cancer Res* 63:3011-3018.
6. Cormier RT, Hong KH, Halberg RB, Hawkins TL, Richardson P, Mulherkar R, Dove WF, Lander ES (2000) *Nat Genet* 17:88-91.
7. Ruivenkamp CA, van Wezel T, Zanon C, Stassen AP, Vlcek C, Csikos T, Klous AM, Tripodis N, Perrakis A, Boerrigter L *et al.* (2002) *Nat Genet* 31:295-300.
8. Ewart-Toland A, Briassouli P, de Koning JP, Mao JH, Yuan J, Chan F, MacCarthy-Morrogh L, Ponder BAJ, Nagase H, Burn J *et al*. (2003) *Nat Genet* 34:403-412.
9. Youngren KK, Coveney D, Peng X, Bhattacharya C, Schmidt LS, Nickerson ML, Lamb BT, Deng JM, Behringer RR, Capel B *et al.* (2005) *Nature* 435:360-364.
10. Liu P, Wang Y, Vikis H, Maciag A, Wang D, Lu Y, Liu Y, You M (2006) *Nat Genet* 38:888-895.
11. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K *et al.* (2002) *Proc Natl Acad Sci USA* 99:15524-15529.
12. Esquela-Kerscher A, Slack FJ (2006) *Nat Rev Cancer* 6:259-269.
13. Sevignani C, Calin GA, Siracusa LD, Croce CM (2006) *Mamm Genome* 17:189-202.
14. Calin GA, Croce CM (2006) *Nat Rev Cancer* 6:857-866.
15. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA *et al.* (2005) *Nature* 435:834-838.
16. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) *Nature* 433:769-773.
17. Bartel DP (2004) *Cell* 23:281-297.
18. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M *et al.* (2004) *Proc Natl Acad Sci USA* 101:2999-3004.
19. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamaoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T *et al.* (2006) *Cancer Cell* 9:189-198.
20. Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M *et al.* (2006) *Proc Natl Acad Sci USA* 103:2257-2261.
21. Huppi R, Volfovsky N, Mackiewicz M, Runfola T, Jones TL, Martin SE, Stephens R, Caplen NA (2007) *Sem Cancer Biol* 17:65-73.

22. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y *et al.* (2004) *Cancer Res* 64:3753-3756.
23. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ (2005) *Cell* 120:635-647.
24. Iwai N, Naraba H (2005) *Biochem Biophys Res Comm* 331:1439-1444.
25. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik S, Iorio MV, Visone R, Sever NI, Fabbri M *et al.* (2005) *N Engl J Med* 353:1793-1801.
26. Chen CZ, Li L, Lodish HF, Bartel DP (2004) *Science* 303:83-86.
27. Ohler U, Yekta S, Lim LP, Bartel DP, Burge CB (2004) RNA 10:1309-1322.
28. Parizotto EA, Dunoyer P, Rahm N, Himber C, Voinnet O (2007) *Genes & Dev* 18:2237-2242.
29. Seitz H, Royo H, Lin SP, Youngson N, Ferguson-Smith AC, Cavaille J (2004) *Biol Chem* 385:905-911.
30. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) *EMBO J* 23:4051-4060.
31. Cai X, Hagedorn CH, Cullen BR (2004) *RNA* 10:1957-1966.
32. Kim VN (2005) *Nat Rev Mol Cell Biol* 6:376-385.
33. Davuluri RV, Grosse I, Zhang MQ (2001) *Nat Genet* 29:412-417.
34. Kent WJ (2002) *Genome Res* 12:656-664.
35. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) *Science* 294:858-862.
36. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) *Science* 294:853-858.
37. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) *Genome Res* 14:1902-1910.
38. Ambros V (2004) *Nature* 431:350-355.
39. Costa FF (2005) *Gene* 357:83-94.
40. Mendell JT (2005) *Cell Cycle* 4:1179-1184.

FIGURE LEGENDS

Fig. 1. Chromosomal positions of mouse tumor susceptibility loci and associated miRNAs. The MUSMIRSUS Database was used to identify miRNAs at or near the peak of mouse tumor susceptibility loci (www.kimmelcancercenter.org/siracusa/musmirsus.htm). Mouse chromosomes are shown with centromeres at the top. The map positions of mouse miRNAs (ovals) are compared with those of susceptibility loci (rectangles) for 8 types of solid tumors. MiRNAs <5 Mb from the peak of a susceptibility locus are shown. A large cluster of 28 miRNAs on distal Chr 12 includes tissue-specific miRNA genes (*mir-127*, *mir-*

136, *mir-134*, *mir-154*) that map to an imprinted region (29); most of these miRNAs reside near the lung *Sluc12* locus.

Fig. 2. Rates of miRNAs in susceptibility regions of mouse chromosomes. This plot illustrates the estimated rate of miRNAs in susceptibility regions versus non-susceptibility regions by chromosome. For plotting purposes, rates of miRNAs within a region are computed as number of miRNAs/length of region (within each chromosome). Rates of miRNAs are plotted by chromosome and by susceptibility region status (susceptibility region versus not). The estimated length in Mb of the genome was based on Ensembl v33 (September 2005). There is a significant association ($p<0.001$) of the incidence of miRNAs near the peak of tumor susceptibility regions.

Fig. 3. miRNA gene sequences are different between inbred strains. The nucleotide sequences of A) *mir-1-1*, B) *mir-219-1,* and C) *mir-196a-1* are shown in the strains indicated in the top left boxes. Gray shading highlights sequence variants (substitutions and/or in-dels). The miRNA stem-loop sequences are in bold and the mature miRNA sequences are underlined. The tumor susceptibility loci that map close to each miRNA are shown in parentheses. For *mir-1-1*, several substitutions were found between the resistant B10.O20 and the susceptible O20 strain. The sequence of B10.O20 is identical to that of the B6 strain. For *mir-219-1*, a 6 bp in-del and two substitutions were found between the resistant B6 and the susceptible A/J strains. For *mir-196a-1*, four alleles were detected; the sequence of B6 is identical to that of the evolutionarily divergent CAST strain. The asterisk indicates the position of a substitution found within the *mir-196a-1* stem-loop. *Mir-196a-2* was also sequenced, because it has the same mature sequence as *mir-196a-1*, but is located on a different chromosome; *mir-196a-2* had two substitutions upstream of the stem-loop in *M. spretus*, but no difference was found between susceptible and resistant strains (data not shown).