

MicroRNA genes are frequently located near mouse cancer susceptibility loci

Cinzia Seivignani*, George A. Calin†, Stephanie C. Nnadi*, Masayoshi Shimizu†, Ramana V. Davuluri†, Terry Hyslop†, Peter Demant§, Carlo M. Croce*¶, and Linda D. Siracusa*¶

*Department of Microbiology and Immunology, Kimmel Cancer Center, and †Department of Pharmacology and Experimental Therapeutics, Jefferson Medical College, Philadelphia, PA 19107; ‡Department of Molecular Virology, Immunology, and Medical Genetics and Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210; and §Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY 14263

Contributed by Carlo M. Croce, March 19, 2007 (sent for review January 25, 2007)

MicroRNAs (miRNAs) are short 19- to 24-nt 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. Because 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 a catalog of miRNA genes in inbred strains that could represent genes involved in the development and penetrance of solid tumors.

solid tumor | inbred strains | quantitative trait loci

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 and susceptibility to infectious diseases (refs. 1 and 2 and 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 use to search for allelic variants that confer susceptibility or resistance to different types of tumors (3). Quantitative trait loci 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 microRNA (miRNAs), could be responsible for susceptibility to solid tumors.

Several papers have shown that miRNAs play an important role in certain human cancers, including B cell chronic lymphocytic leukemias (11), and 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 down-regulate 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 and cancer-associated genomic regions (18).

Compilation of Known Tumor Susceptibility and Modifier Loci. We constructed a database that contained both susceptibility and modifier loci for eight 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 inbred 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 [supporting information (SI) Fig. 4] suggests that genes predisposing to solid tumors are not distributed randomly within the genome, but tend to be concentrated on certain Chrs. Third, Fig. 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*

Author contributions: C.S. and G.A.C. contributed equally to this work; C.S., G.A.C., and L.D.S. designed research; C.S., G.A.C., S.C.N., M.S., and R.V.D. performed research; R.V.D., T.H., P.D., and C.M.C. contributed new reagents/analytic tools; C.S., S.C.N., R.V.D., T.H., and L.D.S. analyzed data; and C.S., G.A.C., T.H., P.D., C.M.C., and L.D.S. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: miRNA, microRNA; Chr, chromosome; in-del, insertion–deletion; IRR, incidence rate ratio.

¶To whom correspondence may be addressed. E-mail: carlo.croce@osumc.edu or siracusa@mail.jci.tju.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0702177104/DC1.

© 2007 by The National Academy of Sciences of the USA

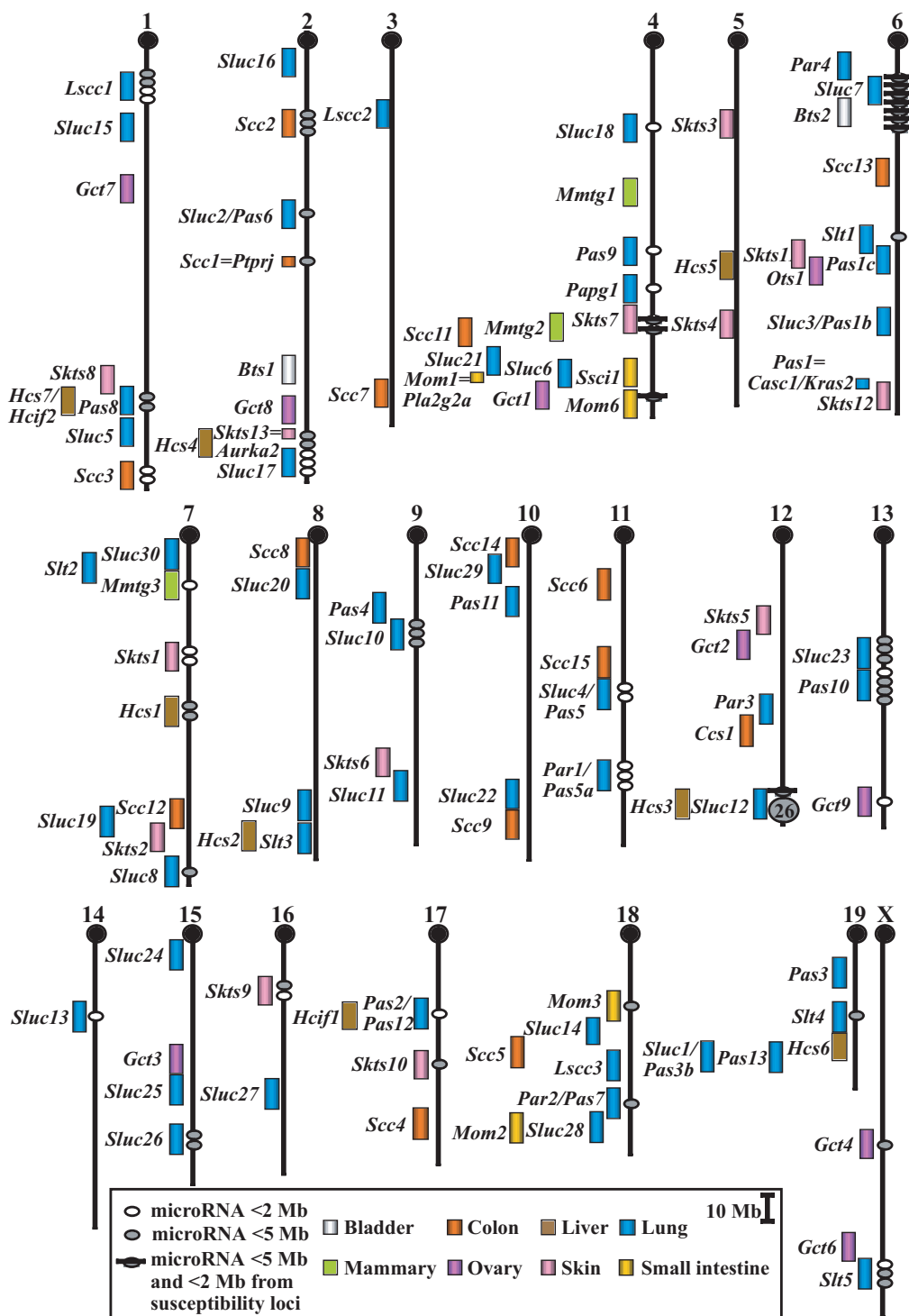


Fig. 1. Chromosomal positions of mouse tumor susceptibility loci and associated miRNAs. The MUSMIRSUS database was used to identify miRNAs at or near the peak of mouse tumor susceptibility loci (www.kimmellcancercenter.org/siracusa/musmirsus.htm). Mouse Chrs are shown with centromeres at the top. The map positions of mouse miRNAs (ovals) are compared with those of susceptibility loci (rectangles) for eight 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 (19); most of these miRNAs reside near the lung *Sluc12* locus.

on Chr 17, and *Pas3b* and *Sluc1* on Chr 19). Because these tumor susceptibility loci were discovered by using different test conditions and strains in different laboratories and yet, reside in close proximity, they may be caused by 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

Table 1. Examples of miRNAs located near solid tumor susceptibility loci

Locus/marker	Chr	Mb	Tissue	Resistant strains	Susceptible strains	Distance in bp*	miRNA homolog in human cancers†
<i>mmu-mir-30a</i>	1	23.5				791,353	Low lung; high pancreas
<i>mmu-mir-30c-2</i>	1	23.5				771,919	High prostate, colon, pancreas; low breast (<i>miR-30c</i>)
<i>Lsc1/D1Mit169</i>	1	24.3	Lung	AKR/J, C57BL/6J, 129X1/SvJ	A/J, SWR/J, NIH Swiss		
<i>mmu-mir-194-1</i>	1	185.0				169,212	Human FRA1H
<i>mmu-mir-215</i>	1	185.0				168,950	Human FRA1H; high lung, stomach
<i>Scc3/D1Mit208</i>	1	185.1	Colon	BALB/cHeA	STS/A		
<i>Sluc17/D2Mit200</i>	2	179.5	Lung	B10.O20	O20		
<i>mmu-mir-1-1</i>	2	180.0				597,348	
<i>mmu-mir-133a-2</i>	2	180.0				606,679	
<i>mmu-mir-31</i>	4	87.2				402,170	High breast
<i>Pas9/D4Mit77</i>	4	87.6	Lung	BALB/c	A/J		
<i>Mmtg2/D4Mit200-308</i>	4	118.6	Mammary	FVB/Nmul-TgN	FVB/NJ		
<i>mmu-mir-30c-1</i>	4	119.1				469,491	High prostate, colon, pancreas; low breast (<i>miR-30c</i>)
<i>mmu-mir-30e</i>	4	119.1				472,563	
<i>Mom6/D4Mit64 - Tel</i>	4	146.7	Small intestine	AKR/J	C57BL/6J		
<i>mmu-mir-34a</i>	4	148.0				855,896	Human CAGR, high lung, prostate, solid cancer vs. normal
<i>Sluc4/Pas5/D11Mit15</i>	11	69.6	Lung	B10.O20	O20		
<i>mmu-mir-324</i>	11	69.6				63,439	
<i>mmu-mir-195</i>	11	69.8				286,438	High lung, prostate, solid cancer vs. normal
<i>mmu-mir-196a-1</i>	11	95.9				20,475	High lung, prostate (<i>miR-196-1</i>)
<i>Par1/Pas5a/D11Mit54</i>	11	95.9	Lung	C57BL/6J, BALB/c, SM/J, <i>M. spretus</i>	A/J		
<i>mmu-mir-10a</i>	11	96.0				31,526	High colon, lung, solid cancer vs. normal
<i>mmu-mir-152</i>	11	96.5				564,754	Low pancreas, solid cancer vs. normal
<i>mmu-mir-219-1</i>	17	32.5				719,176	Low lung
<i>Pas2/Pas12/D17Mit34</i>	17	33.2	Lung	C57BL/6J	A/J		

miRNAs located < 1 Mb from the peak marker of susceptibility loci on Chrs 1, 2, 4, 11, 17 are shown. For a complete list, see the MUSMIRSUS Database (www.kimmelcancercenter.org/siracusa/musmirsus.htm).

*The base pairs (bp) represent the distance between a miRNA and the closest susceptibility locus.

†Data from refs. 18, 23, and 24 (solid cancers vs. normal tissues are from tables 10, 11, and 12 in ref. 24).

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 web site (<http://microrna.sanger.ac.uk>). We combined and compared the Mb 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 suscepti-

bility loci, we used the random effect poisson regression model (18). For this analysis, the random effect used was Chr, in that data within a Chr is assumed correlated. Under this model, the number of miRNAs defined "events" and nonoverlapping 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 quantitative trait locus 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), two-sided 95% confidence interval of the IRR, and two-sided *P* values for testing the hypothesis that the IRR is 1.0. An IRR significantly >1 indicates an increase in the number of miRNAs within a region. All statistical computations were completed by using STATA version 7.0.

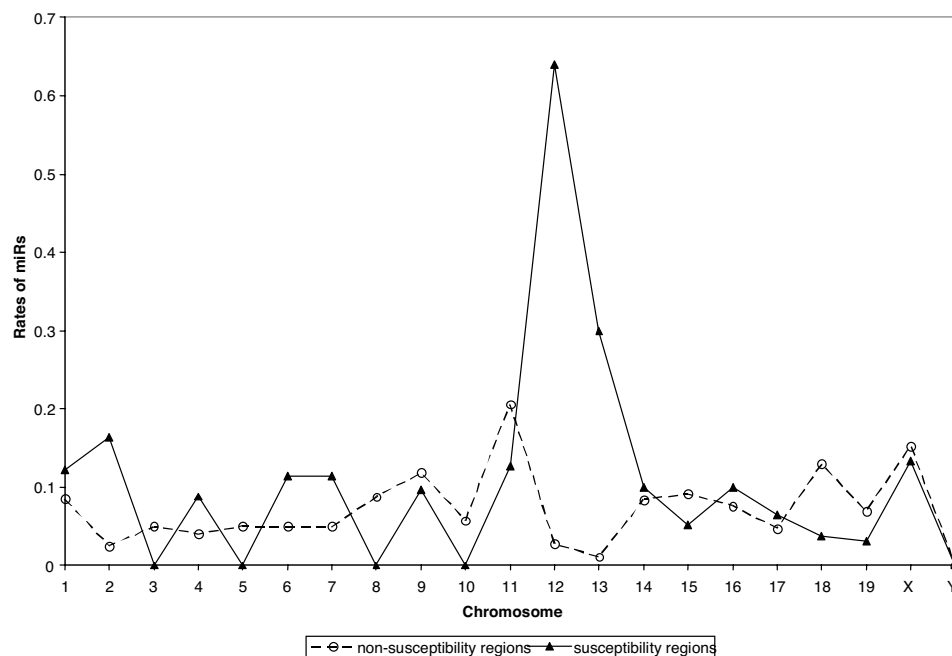


Fig. 2. Rates of miRNAs in tumor-susceptibility regions of mouse Chrs. The plot illustrates the estimated rate of miRNAs in susceptibility regions versus nonsusceptibility regions by Chr. For plotting purposes, rates of miRNAs within a region are computed as number of miRNAs per length of region (within each Chr). Rates of miRNAs are plotted by Chr and susceptibility region status (susceptibility region versus not). The estimated length in Mb of the genome was based on Ensembl version 33 (September 2005). There is a significant association ($P < 0.001$) of the incidence of miRNAs near the peak of tumor-susceptibility regions.

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 nonsusceptibility regions [IRR = 1.63; 95% confidence interval (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 Chrs, but quite large in Chrs 2, 6, 7, 13, and especially 12. Thus, a larger number of miRNAs than expected by chance occurs in these five Chrs. Regardless of location, the model indicates that miRNAs are significantly more likely to reside within tumor susceptibility regions than outside of them. In addition, a recent report suggests an association between the locations of mouse miRNAs and known sites of retroviral integration in mouse cancers (20).

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 *Sccl1* 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 *Gcl1* 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 six 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 and fragile sites as 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 (21). Because 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 (22). 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 *Sccl3* 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 *Sccl2* locus on Chr 2; homologous human miRNAs are located inside a cancer-associated genomic region 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 noncell 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 (23, 24).

Direct Sequence Screen of Inbred Strains. Discrete sequence alterations are frequently found in human cancers, and variation in the sequence of precursor miRNAs were identified in the normal human population (25) and the germ line of cancer patients (26). 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 (25). 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 *Lsccl1* locus on Chr 1.

We sequenced several miRNAs to determine whether inbred strains with different tumor susceptibilities possessed unique al-

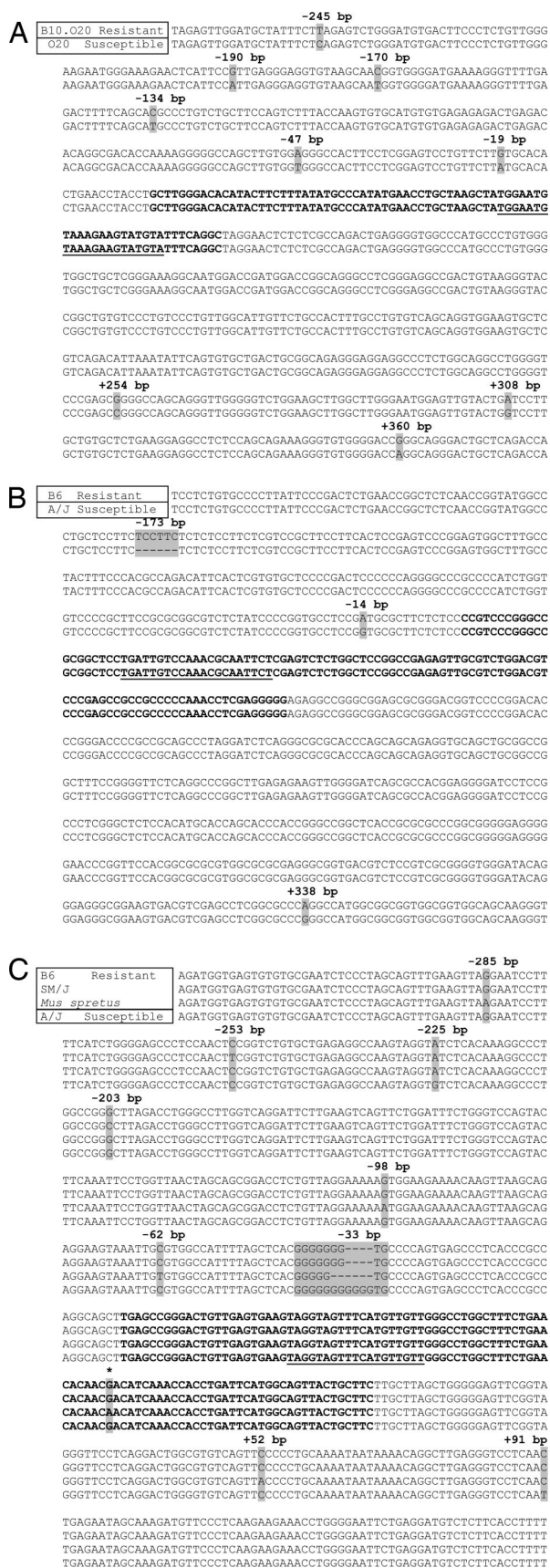


Fig. 3. miRNA gene sequences differ between resistant and susceptible inbred strains. The nucleotide sequences of *mir-1-1* on Chr 2 (near the lung *Sluc17* locus) (A), *mir-219-1* on Chr 17 (near the lung *Pas2/Pas12* locus) (B), and

leles. 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, eight of these loci have more than one miRNA that lies <1 Mb away from their peak marker. Selection of miRNAs was based on criteria that included their proximity to susceptibility loci, the same miRNA being present in human cancer-associated genomic regions, 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*, and *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 (27). A highly conserved motif ≈200 bp upstream of the miRNA stem loop was found in most independently (intergenic) transcribed nematode miRNAs (28). 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 (29).

Seven of the 13 miRNAs showed sequence differences between inbred strains (Fig. 3 and SI Figs. 5–8). Fig. 3A shows that the flanking region of *mir-1-1* (near *Sluc17*) has six substitutions upstream of the stem loop in the susceptible O20 compared with the resistant B10.O20 strain. Fig. 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. Fig. 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 show 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 *Lsccl*), the resistant AKR and B6 strains had different alleles, with six substitutions and one insertion–deletion (in-del). The susceptible A/J strain has a third allele (SI Fig. 5). For *mir-133a-2* (near *Sluc17*), a deletion of TATATGTA was found in the susceptible O20 compared with the resistant B10.O20 strain. Furthermore, there are four substitutions upstream and two substitutions downstream of the *mir-133a-2* stem loop (SI Fig. 6). For *mir-34a* (near *Mom6*), six 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 (SI Fig. 7). For *mir-10a* (near *Par1/Pas5a*), an in-del upstream of the stem loop was detected between the susceptible A/J versus the resistant B6 and SM/J strains (as well as BALB). *M. spretus* and CAST had additional unique alleles (SI Fig. 8).

mir-196a-1 on Chr 11 (near the lung *Par1/Pas5a* locus) (C) 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 Chr; *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).

