Evaluation of Genetic Variants in MicroRNA-Related Genes and Risk of Bladder Cancer

Hushan Yang,[1] Colin P. Dinney,[2] Yuanqing Ye,[1] Yong Zhu,[3] H. Barton Grossman,[2] and Xifeng Wu[1]

Departments of [1]Epidemiology and [2]Urology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas and [3]School of Public Health, Yale University School of Medicine, New Haven, Connecticut

Abstract

MicroRNAs (miRNA) are small noncoding RNA molecules involved in a diversity of cellular functions. Although it has been reported that global suppression of the miRNA biogenesis pathway leads to enhanced tumorigenesis, the effect of common genetic variants of miRNA-related genes on cancer predisposition is unclear. To better understand this effect, we genotyped 41 single-nucleotide polymorphisms (SNP) from 24 miRNA-related genes in a case-control study conducted in 746 Caucasian patients with bladder cancer and 746 matched controls. The homozygous variant genotype of a nonsynonymous SNP in the *GEMIN3* gene (rs197414) was associated with a significantly increased bladder cancer risk [odds ratios (OR), 2.40; 95% confidence interval (95% CI), 1.04–5.56]. Several additional miRNA-related SNPs were also identified that showed a borderline significant association with bladder cancer risk. Haplotype analysis indicated that a common haplotype of the *GEMIN4* gene was associated with a significantly increased bladder cancer risk with an OR of 1.25 (95% CI, 1.01–1.54). To assess the aggregate effects of the promising SNPs, we performed a combined unfavorable genotype analysis that included all SNPs showing at least a borderline statistical significance. We found that, compared with the low-risk reference group with less than two unfavorable genotypes, the medium-risk group with two unfavorable genotypes exhibited a 1.29-fold (0.92–1.81) increased risk whereas the high-risk group with more than two unfavorable genotypes exhibited a 1.92-fold (1.36–2.71) increased risk (P_{trend} < 0.0001). Overall, this is the first epidemiologic study showing that miRNA-related genetic variants may affect bladder cancer risk individually and jointly. [Cancer Res 2008;68(7):2530–7]

Introduction

MicroRNAs (miRNA) are a group of endogenous, small, noncoding RNA molecules of ~22 nucleotides (1). To date, more than 500 human miRNAs have been recorded in the miRBase registry, with the total number predicted to be at least 800 (2, 3). It has been conjectured that miRNAs regulate the expression of approximately one third of human genes (4, 5). The interaction of miRNAs and target genes is intricately regulated, in that one miRNA may modulate multiple target genes whereas one target gene may be regulated by various miRNAs.

MiRNAs negatively affect the expression level of their target genes through two distinctive mechanisms, depending on the degree of their complementarity to target sequences (6). In the first mechanism, a perfect or near-perfect match between miRNAs and their binding sequences within the 3′ untranslated regions (UTR) of their target mRNAs induces the RNA-mediated interference pathway. The RNA-induced silencing complex then recognizes the miRNA-mRNA interaction and cleaves the mRNA through an endonuclease activity. In the second mechanism, miRNAs control gene expression at the translational level through imperfect target matching (6, 7).

MiRNAs have been implicated in a wide diversity of basic cellular functions, such as organ development (8), insulin secretion (9), muscle differentiation (10), immune response (11), and cardiac regulation (12). Moreover, miRNAs have been extensively associated with the etiology and clinical outcome of human cancers (6, 13, 14). MiRNAs influence tumorigenesis through their regulation of specific proto-oncogenes and tumor suppressor genes. For instance, the let-7 miRNA inhibits lung tumorigenesis by repressing the expression of the *RAS* oncogene (15). The transcriptional factor E2F1 is also negatively regulated by two miRNAs, miR-17-5p and miR-20a, in the polycistronic *mir-17-92* cluster, the expression of which is controlled by the c-MYC onco-protein (16). In addition, by inhibiting the expression of the *LATS2* tumor suppressor gene, mir-372 and mir-373 promote the transformation of human testicular germ cell tumors (17). Moreover, the large-scale profiling of miRNA expressions using microarray or real-time PCR techniques has revealed significant associations between miRNA expression signatures and the etiology, early diagnosis, molecular classification, and prognosis of various cancers (4, 18–20).

MiRNAs are generated in a precisely coordinated two-step pathway. Most miRNAs reside in intergenic or intronic regions and are transcribed as a part of a long transcript through RNA polymerases II (6). These primary miRNA transcripts (pri-miRNA) are processed in the nucleus by the microprocessor machinery, which contains the Drosha RNase and the double-strand RNA binding protein DGCR8 (21). A hairpin precursor miRNA molecule of 70 to 100 nucleotides (pre-miRNA) is then produced, which translocates to the cytoplasm through the assistance of RAN GTPase and Exportin 5 (XPO5), where it is further processed by a protein complex that includes DICER, TRBP, AGO1, and AGO2, leading to the production of mature miRNAs (22, 23). The global or specific deregulation of key genes in the miRNA biogenesis pathway has been associated with malignant transformation (6, 22, 24).

Although genetic polymorphisms have been widely implicated in cancer development and treatment response (25, 26), such evidence is lacking for the miRNA-related genes. SNPs in miRNA-containing genomic regions have been reported to be rare and unlikely to be functionally important (27, 28). However, Duan et al. (29) identified a SNP within the seed region of *miR-125* that significantly affected the production of *miR-125a* precursors. A SNP

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Xifeng Wu, Department of Epidemiology, Unit 1340, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: 713-745-2485; Fax: 713-792-4657; E-mail: xwu@mdanderson.org.

doi:10.1158/0008-5472.CAN-07-5991

in the precursor stem-loop of *miR-K5*, a miRNA encoded by the human Kaposi's sarcoma-associated herpesvirus, has also been found to influence Drosha processing (30). These lines of evidence suggest that sequence variations affect the expression or function of their host miRNAs. However, as yet, no study has shown an association between the polymorphisms in the miRNA biogenesis pathway genes and cancer incidence.

In this case-control study, we tested the hypothesis that common sequence variants in genes of miRNA and of the miRNA biogenesis pathway affect bladder cancer susceptibility. We used a polygenic approach to evaluate the haplotypes and combined effects of 41 potentially functional miRNA-related SNPs on bladder cancer risk. To our knowledge, this is the first study of the involvement of polymorphisms in miRNA processing pathway genes in cancer predisposition.

Materials and Methods

Study population and epidemiologic data. The study population has been described elsewhere (26). Briefly, the population consisted of patients with newly diagnosed and histopathologically confirmed bladder cancer accrued at The University of Texas M. D. Anderson Cancer Center and Baylor College of Medicine. The control subjects were selected from a large control pool recruited through a collaboration with the Kelsey-Seybold Clinic, the largest private multi-specialty physician group in the Houston metropolitan area, consisting of 23 clinics and more than 300 physicians. The potential control subjects were identified through a short questionnaire during registration when they visit the Kelsey Seybold Clinic to determine their willingness to participate in the study and to provide preliminary demographic data for matching purposes. They were then contacted by telephone at a later date to confirm their willingness to participate and to schedule an interview at a Kelsey-Seybold clinic. On the day of the interview, the potential control subject came to the clinic specifically for the purpose of study participation. Controls had no cancer history (except nonmelanoma skin cancer) and were frequency matched to cases on age, gender, and ethnicity. This control selection strategy has been well described and proved to be feasible and effective for molecular epidemiologic studies in which population-based control selection poses a practical challenge (31). All subjects were interviewed using a structured questionnaire. Each participant had a 40-mL blood sample drawn into a coded, heparinized tube, which was sent to the laboratory for immediate molecular analysis. Laboratory personnel were blinded to case-control status. All participants signed written informed consent forms, and human subject approval was obtained from both M. D. Anderson and Kelsey-Seybold institutional review boards.

Selection of genes and polymorphisms. Through an extensive mining of the databases of the International HapMap Project (32), dbSNP (33), and miRBase registry (3), we identified 41 potential functional polymorphisms: 24 SNPs in eleven genes in the miRNA biogenesis pathway, 7 SNPs in seven pre-miRNAs, and 10 SNPs in eight pri-miRNAs (Table 1). All SNPs have a reported minor allele frequency (MAF) of >0.01 in Caucasians. In the miRNA biogenesis pathway, except for two *AGO1* SNPs (rs636832 and rs595961) located in introns, all other polymorphisms reside in functional regions, including exons, UTRs, and promoters (within 2 kb of the genes). In the case of multiple potentially functional SNPs within the same haplotype block (defined by the linkage coefficient $r^2 > 0.8$), only one SNP was included. Except for *GEMIN4* rs7813, none of these SNPs have been reported in previous studies. All SNPs identified from the pre-miRNA regions were included if the MAF was >0.01 in Caucasians. For SNPs in pri-miRNAs but not in pre-miRNAs, because we identified more than 200 such SNPs with an MAF of >0.01 in Caucasians, we included 10 SNPs from eight pri-miRNAs whose mature counterparts have been extensively implicated in cancer etiology or clinical outcome.

Genotyping. All polymorphisms were genotyped using the SNPlex assay according to the manufacturer's instructions (Applied Biosystems). SNPlex is a high-throughput genotyping approach that makes use of a set of preoptimized, universal assay reagents together with customized probe sets to generate genotyping data through oligonucleotide ligation, exonuclease purification, PCR amplification, probe hybridization, and capillary electrophoresis (34). Briefly, we customized and submitted a list of miRNA-related SNPs to Applied Biosystems. Based on this list, a pool of allele-specific oligonucleotide (ASO) probes and locus-specific oligonucleotide (LSO) probes was designed. Genomic DNA was fragmented at 99°C for 10 min and hybridized to a mixture of the probe pool and universal linkers that contain universal PCR primer-binding sequences, as well as sequences complementary to ASO and LSO probes. On perfect matching to the sequence at the target SNP site in the genomic DNA, ASO and LSO probes were ligated and the universal linkers were linked to the distal termini of the ASO and LSO probes. The mixture of unligated probes, linkers, and genomic DNAs was purified after exonuclease enzymatic digestion and amplified by PCR using a pair of universal PCR primers, one of which was biotinylated. The biotinylated amplicons were denatured and bound on streptavidin-coated microtiter plates. After the removal of nonbiotinylated strands, the single-stranded PCR amplicons were hybridized with a set of fluorescently labeled, mobility-modified ZipChute probes (Applied Biosystems), eluted into capillary electrophoresis buffer, and analyzed on an Applied Biosystems 3730 DNA Analyzer. Genotypes were called by GeneMapper software (Applied Biosystems) using a template file provided with each custom SNPlex assay. Internal quality controls and negative controls were used to ensure genotyping accuracy, and 5% of all samples were randomly selected and genotyped in duplicate with 100% concordance.

Statistical analysis. Due to the small number of minority participants, we limited all our analyses to Caucasians. Statistical analyses were done using Intercooled STATA software (STATA Corp.). χ^2 analysis was used to assess the differences between cases and controls with regard to categorical variables such as gender and smoking status. Student's t test was used to test for continuous variables, including age and pack-years. The Hardy-Weinberg equilibrium was tested using a goodness-of-fit χ^2 analysis. The bladder cancer risks were estimated as odds ratios (OR) and 95% confidence intervals (95% CI) using unconditional multivariate logistic regression adjusted for age, gender, smoking status, and pack-years, where appropriate. The definitions of smoking status were the same as those that have previously been described (35). Haplotypes were inferred using the expectation-maximization algorithm implemented in the HelixTree software (Golden Helix, Inc.). Haplotypes with a probability of <95% were excluded from the final analysis. The adjusted OR and 95% CI for each haplotype were assessed using multivariate logistic regression under a 1 degree of freedom model that, for each haplotype, combines all other haplotypes as the reference group. The unfavorable genotype analysis included those SNPs showing at least a borderline statistical significance in the main analysis. The unfavorable genotypes were collapsed together and categorized according to the tertiles (low, medium, and high risk) of the number of unfavorable genotypes in controls. Using the low-risk group as the reference group, we calculated the ORs and 95% CIs for the medium-risk and high-risk groups using unconditional multivariate logistic regression adjusted for age, gender, and smoking status. All P values were two sided. $P \leq 0.05$ was considered the threshold of statistical significance.

Results

Characteristics of the study population. The final study population consisted of 1,492 Caucasians, composed of 746 bladder cancer patients and 746 cancer-free controls (Table 2). No significant differences were identified between cases and controls with regard to age [cases versus controls (mean ± SD), 63.5 ± 10.9 versus 63.3 ± 10.6 years; $P = 0.70$] and gender ($P = 1.00$). As expected, there was a significantly higher percentage of ever smokers among cases (73.5%) than among controls (55.1%; $P < 0.001$). Among ever smokers, cases also reported a significantly greater cigarette consumption than did controls, as assessed by the mean number of pack-years (cases versus controls, 42.7 ± 29.9 versus 30.3 ± 28.1; $P < 0.001$).

Main effects and stratified analyses by individual polymorphisms. The associations of the 41 SNPs with bladder cancer risk are listed in Supplementary Table S1. The genotyping completion rate ranged from 90% to 99% for all SNPs except for *DROSHA* rs10719 (65%) and was similar between cases and controls (average rate, 96.5% for cases and 97.2% for controls). Three SNPs (*DROSHA* rs10719, *RAN* rs14035, and *let7f-2* rs17276588) showed a significant deviation from the Hardy-Weinberg equilibrium and were excluded from further analyses. In all, seven SNPs exhibited at least a borderline significant association with bladder cancer risk under either a dominant model (variant-containing genotypes versus homozygous wild-type genotype) or a recessive model (homozygous variant genotype versus wild-type-containing genotypes; Table 3 and Supplementary Table S1). Among them, the AA genotype of *GEMIN3* rs197414 was associated with a 2.5-fold (95% CI, 1.08–5.78; $P = 0.03$) increased risk when compared with the combined CC/CA genotypes. In stratified analyses, this risk remained significant in young subjects (OR, 3.19; 95% CI, 1.00–10.19; $P = 0.05$) and light smokers (OR, 2.97; 95% CI, 1.01–24.31; $P = 0.05$; Table 4). Moreover, an altered risk association was also identified for *TRBP* rs784567

Table 1. MiRNA-related genes and polymorphisms evaluated in this study

Gene name	Polymorphisms	Position	Nucleotide change	Amino acid change	MAF (%)*
MiRNA biogenesis pathway					
DROSHA	rs10719	3′ UTR	C>T		23
	rs6877842	Promoter	G>C		18
DGCR8	rs417309	3′ UTR	G>A		11
	rs3757	3′ UTR	G>A		27
	rs1640299	3′ UTR	T>G		47
XPO5	rs11077	3′ UTR	A>C		40
RAN	rs14035	3′ UTR	C>T		27
DICER	rs3742330	3′ UTR	A>G		12
	rs13078	3′ UTR	T>A		14
TRBP	rs784567	Promoter	C>T		48
AGO1	rs636832	Intron 8	G>A		9
	rs595961	Intron 10	A>G		15
AGO2	rs4961280	Promoter	C>A		13
GEMIN4	rs910924	Promoter	C>T		35
	rs2740348	nsSNP	G>C	Glu>Gln	19
	rs7813	nsSNP	C>T	Arg>Cys	46
	rs910925	nsSNP	C>G	Ala>Gly	48
	rs3744741	nsSNP	C>T	Arg>Gln	13
	rs1062923	nsSNP	T>C	Ile>Thr	18
	rs4968104	nsSNP	T>A	Val>Glu	35
GEMIN3	rs197414	nsSNP	C>A	Arg>Ser	19
	rs197388	Promoter	T>A		29
	rs197412	nsSNP	T>C	Ile>Thr	47
HIWI	rs1106042	nsSNP	G>A	Arg>Lys	8
Pre-miRNAs					
mir146a	rs2910164	Pre-miRNA	G>C		24
mir196a-2	rs11614913	Pre-miRNA	C>T		44
mir423	rs6505162	Pre-miRNA	A>C		43
mir492	rs2289030	Pre-miRNA	C>G		8
mir604	rs2368392	Pre-miRNA	C>T		25
mir608	rs4919510	Pre-miRNA	C>G		18
mir631	rs5745925	Pre-miRNA	CT>-		7
Pri-miRNAs					
let7f-2	rs17276588	5′ region	G>A		2
mir26a-1	rs7372209	5′ region	C>T		29
mir30a	rs1358379	5′ region	A>G		4
mir30c-1	rs16827546	5′ region	C>T		4
mir100	rs1834306	5′ region	C>T		44
mir124-1	rs531564	5′ region	C>G		12
mir219-1	rs107822	5′ region	G>A		23
	rs213210	3′ region	T>C		6
mir373	rs12983273	5′ region	C>T		13
	rs10425222	3′ region	C>A		3

Abbreviation: nsSNP, nonsynonymous SNP.
*MAF in Caucasians, as reported in dbSNP database.

Table 2. Distribution of selected host characteristics by case-control status in Caucasians

Variables	Cases, *n* (%)	Controls, *n* (%)	*P**
Age (mean ± SD)	63.5 ± 10.9	63.3 ± 10.6	0.70
Gender			
Male	586 (79)	586 (79)	
Female	160 (21)	160 (21)	1.00
Smoking status			
Never	195 (26)	333 (45)	
Ever	467 (73.5)	350 (55.1)	<0.001
Pack-years (mean ± SD)†	42.7 ± 29.9	30.3 ± 28.1	<0.001

*P values were derived from the χ^2 test for categorical variables (gender and smoking status) and t test for continuous variables (age and pack-years).
† Among ever smokers.

in young subjects (OR, 0.69; 95% CI, 0.48–0.98; $P = 0.04$) and ever smokers (OR, 0.74; 95% CI, 0.54–1.00; $P = 0.05$), for *mir423* rs6505162 in males (OR, 1.34; 95% CI, 1.00–1.79; $P = 0.05$), for *mir492* rs2289030 in females (OR, 2.67; 95% CI, 1.26–5.62; $P = 0.01$), for *mir26a-1* rs7372209 in females (OR, 0.36; 95% CI, 0.13–0.94; $P = 0.04$), and for *mir124-1* rs531564 in old subjects (OR, 4.85; 95% CI, 1.02–23.01; $P = 0.05$).

Haplotype analyses. Table 5 summarizes the relative risks associated with the common haplotypes of genes in this study. The only haplotype associated with bladder cancer risk was the H3 (WMMMWWW; W, wild-type allele; M, variant allele) haplotype of the *GEMIN4* gene consisting of a promoter SNP and six nonsynonymous SNPs in the following order: rs910924, rs2740348, rs7813, rs910925, rs3744741, rs1062923, and rs4968104. Compared with the reference group combining all other *GEMIN4* haplotypes, this haplotype was associated with a 1.25-fold (95% CI, 1.01–1.54; $P = 0.04$) increased bladder cancer risk.

Combined effects of the unfavorable genotypes. We further evaluated the combined effects of the high-risk genotypes on bladder cancer risk by collapsing the unfavorable genotypes of the seven risk-conferring SNPs shown in Table 3. We found a progressively increased gene-dosage effect when the subjects were grouped on the basis of an increasing number of unfavorable genotypes (Table 6). That is, compared with the low-risk group consisting of subjects with less than two unfavorable genotypes, the medium-risk group with two unfavorable genotypes was at a 1.29-fold (95% CI, 0.92–1.81; $P = 0.14$) increased risk whereas the high-risk group with more than two unfavorable genotypes was at a 1.92-fold (95% CI, 1.36–2.71; $P < 0.0001$) increased risk ($P_{trend} < 0.0001$; Table 6).

Discussion

In this study, in which we assessed the effects of 41 SNPs in genes of the miRNA biogenesis pathway, pre-miRNAs, and pri-miRNAs on bladder cancer predisposition, we found that a nonsynonymous SNP in *GEMIN3* and a common haplotype of *GEMIN4* were associated with a significantly increased bladder cancer risk. We also identified several additional miRNA-related SNPs showing a borderline significant association with bladder cancer risk. In addition, we showed that the combined unfavorable genotypes of selected SNPs might be used jointly to predict bladder cancer risk. To our knowledge, this is the first study to evaluate the associations of miRNA-related polymorphisms and cancer susceptibility.

A few studies have been done that examined sequence variations in miRNA regions. For example, Iwai and Naraba (36) sequenced 173 human pre-miRNA genomic regions in 96 subjects and identified 10 polymorphisms. They suspected that an A/C SNP in the mature *miR-30c-2* may alter target gene selection and thus have biological consequences (36). Through an extensive database interrogation, Saunders et al. (27) identified 65 SNPs in 474 pre-miRNAs. However, many of these SNPs may not actually be important to population genetics due to the lack of frequency data (27), which is consistent with the notion that genetic variants in pre-miRNA regions are scarce and unlikely to be functionally

Table 3. Association of selected MiRNA-related polymorphisms with bladder cancer risk

Gene	SNP	Genotype	Case/control	OR (95% CI)*	*P*
TRBP	rs784567 (C>T)	CC	200/168	1 (reference)	
		CT/TT	537/561	0.80 (0.62–1.02)	0.07
GEMIN4	rs7813 (T>C)	TT/TC	225/222	1 (reference)	
		CC	511/514	0.78 (0.60–1.02)	0.07
GEMIN3	rs197414 (C>A)	CC/CA	717/734	1 (reference)	
		AA	23/8	2.50 (1.08–5.78)	0.03
mir423	rs6505162 (A>C)	AA/AC	553/588	1 (reference)	
		CC	170/145	1.25 (0.96–1.61)	0.09
mir492	rs2289030 (C>G)	CC	638/655	1 (reference)	
		CG/GG	101/79	1.35 (0.98–1.87)	0.07
mir26a-1	rs7372209 (C>T)	CC/CT	686/666	1 (reference)	
		TT	42/62	0.68 (0.45–1.04)	0.07
mir124-1	rs531564 (C>G)	CC/CG	724/733	1 (reference)	
		GG	15/7	2.44 (0.96–6.20)	0.06

*Adjusted for age, gender, and smoking status.

Table 4. Association of selected MiRNA-related polymorphisms with bladder cancer risk stratified by host characteristics

Gene/SNP	Age		Gender		Smoking status		Smoking intensity	
	Young*	Old	Male	Female	Never	Ever	Light	Heavy
	Case/control	Case/control	Case/control	Case/control	Case/control	Case/control	Case/control	Case/control
	OR (95% CI) †		OR (95% CI)		OR (95% CI)		OR (95% CI)	
TRBP/rs784567								
CC	107/79	93/89	439/425	48/36	48/76	152/92	47/49	100/43
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
CT/TT	265/292	272/269	141/145	109/123	145/249	391/302	108/150	264/144
	0.69 (0.48–0.98)	0.93 (0.66–.131)	0.86 (0.65–1.13)	0.62 (0.37–1.04)	0.92 (0.61–1.40)	0.74 (0.54–1.00)	0.75 (0.43–1.32)	0.99 (0.67–1.47)
	P = 0.04	*P* = 0.68	*P* = 0.28	*P* = 0.07	*P* = 0.71	*P* = 0.05	*P* = 0.32	*P* = 0.97
GEMIN4/rs7813								
TT/TC	307/295	299/279	472/452	134/122	160/257	445/317	132/159	303/151
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
CC	65/84	65/78	104/124	26/38	32/71	98/91	24/43	71/47
	0.77 (0.53–1.13)	0.81 (0.56–1.19)	0.84 (0.63–1.14)	0.60 (0.33–1.08)	0.71 (0.45–1.14)	0.82 (0.59–1.14)	0.75 (0.43–1.32)	0.80 (0.52–1.22)
	P = 0.18	*P* = 0.29	*P* = 0.27	*P* = 0.09	*P* = 0.16	*P* = 0.24	*P* = 0.32	*P* = 0.30
GEMIN3/rs197414								
CC/CA	359/378	358/356	563/575	154/159	189/327	527/407	149/201	366/198
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
AA	15/4	8/4	18/7	5/1	4/3	19/5	8/2	10/3
	3.19 (1.00–10.19)	1.94 (0.56–6.68)	2.19 (0.88–5.47)	4.6 (0.51–41.52)	2.32 (0.51–10.49)	2.58 (0.94–7.09)	2.97 (1.01–24.31)	1.57 (0.42–5.93)
	P = 0.05	*P* = 0.29	*P* = 0.09	*P* = 0.17	*P* = 0.28	*P* = 0.07	*P* = 0.05	*P* = 0.50
mir423/rs6505162								
AA/AC	279/301	284/287	433/457	130/131	149/260	413/328	119/162	284/159
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
CC	89/73	81/72	143/117	27/28	42/69	128/76	35/35	90/40
	1.35 (0.93–1.95)	1.15 (0.80–1.67)	1.34 (1.00–1.79)	0.89 (0.49–1.64)	1.07 (0.69–1.66)	1.35 (0.98–1.87)	1.40 (0.82–2.39)	1.28 (0.84–1.96)
	P = 0.11	*P* = 0.45	*P* = 0.05	*P* = 0.716	*P* = 0.76	*P* = 0.07	*P* = 0.22	*P* = 0.26
mir492/rs2289030								
GG	327/336	311/319	504/509	134/146	165/296	472/359	138/180	321/172
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
GC/CC	47/39	54/40	76/67	25/12	29/31	72/48	16/21	56/26
	1.338 (0.86–2.21)	1.35 (0.86–2.11)	1.14 (0.79–1.63)	2.67 (1.26–5.62)	1.68 (0.98–2.88)	1.21 (0.81–1.80)	1.02 (0.51–2.05)	1.25 (0.75–2.09)
	P = 0.18	*P* = 0.19	*P* = 0.49	*P* = 0.01	*P* = 0.06	*P* = 0.36	0.96	*P* = 0.39
mir26a-1/rs7372209								
CC/CT	345/342	341/324	537/526	149/140	179/293	507/473	144/182	352/183
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
TT	21/28	21/34	36/44	6/18	10/33	31/29	10/14	20/15
	0.68 (0.36–1.26)	0.65 (0.37–1.16)	0.89 (0.54–1.45)	0.36 (0.13–0.94)	0.49 (0.23–1.02)	0.82 (0.48–1.40)	0.98 (0.42–2.28)	0.71 (0.35–1.44)
	P = 0.22	*P* = 0.15	*P* = 0.63	0.04	*P* = 0.06	0.47	0.96	*P* = 0.34
mir124-1/rs531564								
CC/CG	366/376	358/357	568/574	156/159	187/325	536/408	153/202	370/198
	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
GG	6/5	9/2	12/7	3/0	6/4	9/3	2/1	7/2
	1.43 (0.40–5.10)	4.85 (1.02–23.01)	2.09 (0.79–5.52)	NA	2.69 (0.74–9.73)	2.16 (0.57–8.25)	2.30 (0.19–28.09)	1.96 (0.39–9.78)
	P = 0.58	*P* = 0.05	*P* = 0.14		*P* = 0.13	0.36	*P* = 0.51	*P* = 0.41

*The median age in controls was used as the cutoff point.

† Adjusted for age, gender, smoking status, and pack-years for the overall analysis; adjusted for age, gender, and pack-years for the analysis stratified with smoking status; adjusted for age, gender, and smoking status for the analysis stratified with smoking level; adjusted for age, smoking status, and pack-years for the analysis stratified with gender; adjusted for gender, smoking status, and pack-years for the analysis stratified with age.

relevant, possibly due to the constraint imposed by natural selection on the evolutionarily conserved pre-miRNA sequences (27, 28). In contrast, Duan et al. (29) identified 323 SNPs in 227 human miRNAs, among which 12 were localized within precursor regions. One SNP in the seed region of mature *miR-125a* was found to be essential to the accurate recognition of target mRNA sequences. *In vivo* functional characterization further revealed that this SNP significantly blocked the maturation of miR-125a (29). Additional evidence of the physiologic relevance of miRNA polymorphisms came from the study of Gottwein et al. (30), who

found that a miRNA precursor SNP influences Drosha processing. Taken together, these findings suggested that genetic variations affect the production or function of the host miRNAs. However, to date, there have not been any studies investigating the relevance of the polymorphisms in the miRNA biogenesis pathway genes to cancer risk.

In our study, the *GEMIN3* and *GEMIN4* genes were found to be associated with bladder cancer risk. Both proteins are core components of a large macromolecular complex that interacts with the survivor or motor neuron protein and plays an essential role in pre-mRNA splicing and ribonucleoprotein assembly (37, 38). Mourelatos et al. (39) found that the GEMIN3 and GEMIN4 proteins are also present in a 15S ribonucleoprotein complex containing eIF2C, a member of the AGO protein family pivotal to miRNA processing. The additional identification of numerous miRNAs in this complex (39), concordant with several other independent observations (40–42), strongly suggests the involvement of GEMIN proteins in the processing of miRNA precursors through their interaction with key components of the RNA-induced silencing complex (42, 43). No functional effect on *GEMIN3* rs197414 has been reported. However, Wan et al. (44) found that the variant allele of *GEMIN4* rs7813 was associated with a significant *in vitro* growth-inhibitory effect in hepatocellular carcinoma cell lines as compared with the wild-type allele, suggesting that the amino acid change caused by this SNP might have physiologic significance. However, whether this SNP produces the altered risk through a similar mechanism in bladder cancer needs to be investigated in an *in vivo* setting.

Besides the significant SNP/haplotype identified for the *GEMIN* genes, borderline significant associations with bladder cancer risk

Table 5. Association of haplotypes of MiRNA-related genes with bladder cancer risk

Gene haplotype	Haplotype sequence	Haplotype frequency (%)	Case/control	OR (95% CI)*	*P*
DROSHA (rs10719-rs6877842)					
H1	WW†	65	683/562	0.96 (0.78–1.19)	0.73
H1	MW	18	160/137	0.93 (0.72–1.19)	0.55
H3	WM	15	149/98	1.24 (0.94–1.64)	0.13
H4	MM	2	6/11	NA‡	NA
DGCR8 (rs417309-rs3757-rs1640299)					
H1	WWW	45	666/650	1.04 (0.89–1.21)	0.62
H2	WWM	24	327/355	0.92 (0.77–1.10)	0.36
H3	WMM	23	235/340	0.96 (0.80–1.15)	0.66
H4	MWW	8	127/108	1.17 (0.89–1.53)	0.26
DICER (rs3742330-rs13078)					
H1	WW	72	1054/1048	1.06 (0.88–1.28)	0.52
H2	WM	19	228/254	0.88 (0.72–1.07)	0.21
H3	MW	9	116/99	1.15 (0.86–1.52)	0.35
AGO1 (rs636832-rs595961)					
H1	WW	85	1278/1243	1.20 (0.97–1.49)	0.09
H2	MM	9	119/140	0.88 (0.67–1.15)	0.34
H3	WM	6	79/93	0.80 (0.58–1.10)	0.17
GEMIN4 (rs910924-rs2740348-rs7813-rs910925-rs3744741-rs1062923-rs4968104)					
H1	MWMMWWM	26	358/374	0.94 (0.79–1.12)	0.50
H2	WWWWWWW	23	314/288	1.07 (0.89–1.29)	0.45
H3	WMMMWWW	17	259/211	1.25 (1.01–1.54)	0.04
H4	WWWWWMW	17	222/254	0.87 (0.71–1.08)	0.21
H5	WWWMWW	13	167/191	0.88 (0.70–1.11)	0.27
H6	MWWWWWW	1	32/26	NA	NA
GEMIN3 (rs197414-rs197388-rs197412)					
H1	WWW	59	864/862	1.00 (0.85–1.17)	0.98
H2	WWM	20	287/281	1.00 (0.83–1.21)	0.99
H3	MMM	13	190/172	1.01 (0.81–1.27)	0.92
H4	WMM	8	110/115	0.99 (0.75–1.31)	0.95
mir 219-1 (rs107822-rs213210)					
H1	WW	75	1068/1043	0.99 (0.82–1.20)	0.94
H2	MW	18	237/244	0.97 (0.79–1.20)	0.78
H3	MM	6	88/87	NA	NA
mir 373 (rs12983273-rs10425222)					
H1	WW	82	1183/1173	1.10 (0.89–1.36)	0.39
H2	MW	14	183/199	0.89 (0.71–1.11)	0.29
H3	WM	3	36/30	1.22 (0.74–2.01)	0.43

*Adjusted for age, gender, and smoking status.

†M, variant allele; W, wild-type allele.

‡NA, not applicable. Represents those haplotypes that could not be resolved under the one degree of freedom additive model.

Table 6. Joint effects of unfavorable genotypes

No. unfavorable genotypes*	Cases/controls	OR (95% CI) †	*P*
Reference group ‡ ≤1	78/112	1.00	Reference
2	306/350	1.29 (0.92–1.81)	0.14
≥3	326/245	1.92 (1.36–2.71)	<0.0001
P_{trend}			<0.0001

*Definition of unfavorable genotypes: *TRBP* rs784567, WW; *GEMIN4* rs7813 WW+WM, *GEMIN3* rs197414 MM, *mir423* rs6505162 MM, *mir492* rs2289030 WM+MM, *mir26a-1* rs7372209 WW+WM, and *mir124-1* rs531564 MM. W, wild-type allele; M, variant allele.

† Adjusted for age, gender, and smoking status.

‡ Because the subject number in group "0" was sparse (2 cases and 5 controls), the subjects with <2 unfavorable genotypes were combined as the reference group. Because the subject number in group "4" and "5" was also sparse (68 cases and 40 controls for group "4," and 6 cases and 4 controls for group "5"), they were combined with group "3" as the high-risk group.

were also identified for SNPs in several other genes, including *TRBP, mir423, mir492, mir26a-1*, and *mir124-1* (Table 3). In particular, the variant allele of rs784567, which is located in the 5′ UTR of the *TRBP* gene, was associated with a 20% risk reduction (P = 0.07). TRBP is an integral component of a DICER-containing complex important to the cytoplasmic processing of pre-miRNAs into mature miRNAs (23). TRBP recruits the DICER complex to the Argonaute 2 protein, which is the catalytic unit of the RNA-induced silencing complex. Depletion of TRBP was found to result in significantly impaired miRNA biogenesis (23). *mir423* and *mir492* have been found to possess a high-frequency SNP in the pre-miRNA region. We identified fewer than 10 miRNA genes with polymorphisms in the pre-miRNA region, consistent with previous findings indicating the elimination of genetic variants by natural selection pressure in this area (27–29). *mir26a-1* is localized to chromosome 3p21, a region frequently deleted in several epithelial cancers (28, 45). Loss of expression of mir124-1 as a result of CpG island hypermethylation has also been observed in multiple cancer cell lines and is associated with the dysregulation of the oncogene *CDK6* and the tumor suppressor gene *RB* (46). In addition, the expression levels of both *mir26a-1* and *mir124-1* are down-regulated in lung cancer (19). To further explore the implications of our identified miRNAs in bladder tumorigenesis, we used the TargetScan program (47) to identify a list of candidate transcripts targeted by each of these miRNAs. The number of transcripts ranges from 11 (*mir423*) to 773 (*mir124-1*; Supplementary Table S2). By using a normal bladder tissue microarray database (accession no. GSM44682) in the National Center for Biotechnology Information Gene Expression Omnibus (48), we found that, for each miRNA, ~50% of the targeted transcripts are expressed in normal bladder tissues (range, 41–55%). Moreover, 43% to 55% of these expressed transcripts exhibited differential expression patterns between normal and tumor bladder tissues (ref. 49; Supplementary Table S2). Some of these target genes have been implicated in bladder tumorigenesis. For instance, the *PTEN* gene is identified by TargetScan as a putative target of *mir26a-1*. Mutations of *PTEN* have been reported in a wide variety of tumors including bladder cancer (50). If the expressions of these miRNAs in bladder tissues are recognized and the associations between these SNPs and bladder cancer risk are validated, the next key question would be which gene(s) are the targets of these miRNAs in the development of bladder cancer.

There is a possibility that the associations observed in this study were attained by chance, given the loss of robustness after multiple comparison adjustments (data not shown). Therefore, it is likely that the risk-conferring effect of any single polymorphisms may only be minimal. To more powerfully elucidate the influences of miRNA polymorphisms on bladder cancer risk, we used a pathway-based polygenic strategy in which we collapsed all the unfavorable genotypes with at least a borderline significant association to assess their combined effects on tumorigenesis. Through such an approach, we identified a trend toward an increasing bladder cancer risk with an increasing number of unfavorable genotypes that occurred in a dose-dependent manner. This finding reinforces the notion that bladder cancer is a polygenic process and thus a combined analysis of multiple factors may have a greater ability to characterize high-risk populations.

The strength of our study includes our use of a large and homogeneous population with strict matching criteria to eliminate the potential confounding effects of age and gender. We have also constructed a comprehensive catalogue of potentially functional SNPs in most currently known miRNA biogenesis genes. This list can be readily used by independent researchers for replication studies of different cancer sites. Nevertheless, it is likely that some associations we presented here are chance findings. Further epidemiologic and functional studies are needed to validate these results.

Overall, we present the first epidemiologic evidence supporting the involvement of genetic polymorphisms in miRNA genes and miRNA biogenesis pathway genes in cancer development. Our results suggest that individual as well as combined genotypes of miRNA-related variants may be used to predict the risk of bladder cancer.

Acknowledgments

Received 10/25/2007; revised 1/4/2008; accepted 1/10/2008.

Grant support: National Cancer Institute grants CA 74880 and CA 91846.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

1. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75:843–54.
2. Bentwich I, Avniel A, Karov Y, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 2005;37:766–70.
3. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006;34:D140–4.
4. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834–8.
5. Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs down-regulate large numbers of target mRNAs. Nature 2005;433:769–73.
6. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer 2006;6:259–69.
7. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281–97.
8. Schratt GM, Tuebing F, Nigh EA, et al. A brain-specific

microRNA regulates dendritic spine development. Nature 2006;439:283–9.
9. Poy MN, Eliasson L, Krutzfeldt J, et al. A pancreatic islet-specific microRNA regulates insulin secretion. Nature 2004;432:226–30.
10. Chen JF, Mandel EM, Thomson JM, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 2006;38:228–33.
11. Li QJ, Chau J, Ebert PJ, et al. miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell 2007;129:147–61.
12. Care A, Catalucci D, Felicetti F, et al. MicroRNA-133 controls cardiac hypertrophy. Nat Med 2007;13:613–8.
13. Tricoli JV, Jacobson JW. MicroRNA: potential for cancer detection, diagnosis, and prognosis. Cancer Res 2007;67:4553–5.
14. Hammond SM. MicroRNAs as tumor suppressors. Nat Genet 2007;39:582–3.
15. Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. Cell 2005;120:635–47.
16. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature 2005;435:839–43.
17. Voorhoeve PM, le Sage C, Schrier M, et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell 2006;124:1169–81.
18. Meng F, Henson R, Lang M, et al. Involvement of human micro-RNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines. Gastroenterology 2006;130:2113–29.
19. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 2006;9:189–98.
20. Calin GA, Liu CG, Sevignani C, et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc Natl Acad Sci U S A 2004;101:11755–60.
21. Han J, Lee Y, Yeom KH, et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 2006;125:887–901.
22. Yue J, Tigyi G. MicroRNA trafficking and human cancer. Cancer Biol Ther 2006;5:573–8.
23. Chendrimada TP, Gregory RI, Kumaraswamy E, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 2005;436:740–4.
24. Chiosea S, Jelezcova E, Chandran U, et al. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. Cancer Res 2007;67:2345–50.
25. Gu J, Grossman HB, Dinney CP, Wu X. The pharmacogenetic impact of inflammatory genes on bladder cancer recurrence. Pharmacogenomics 2005;6:575–84.
26. Wu X, Gu J, Grossman HB, et al. Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. Am J Hum Genet 2006;78:464–79.
27. Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A 2007;104:3300–5.
28. Diederichs S, Haber DA. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. Cancer Res 2006;66:6097–104.
29. Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Hum Mol Genet 2007;16:1124–31.
30. Gottwein E, Cai X, Cullen BR. A novel assay for viral microRNA function identifies a single nucleotide polymorphism that affects Drosha processing. J Virol 2006;80:5321–6.
31. Hudmon KS, Honn SE, Jiang H, et al. Identifying and recruiting healthy control subjects from a managed care organization: a methodology for molecular epidemiological case-control studies of cancer. Cancer Epidemiol Biomarkers Prev 1997;6:565–71.
32. The International HapMap Project. Nature 2003;426:789–96.
33. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res 1999;9:677–9.
34. Tobler AR, Short S, Andersen MR, et al. The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. J Biomol Tech 2005;16:398–406.
35. Leibovici D, Grossman HB, Dinney CP, et al. Polymorphisms in inflammation genes and bladder cancer: from initiation to recurrence, progression, and survival. J Clin Oncol 2005;23:5746–56.
36. Iwai N, Naraba H. Polymorphisms in human pre-miRNAs. Biochem Biophys Res Commun 2005;331:1439–44.
37. Shpargel KB, Matera AG. Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. Proc Natl Acad Sci U S A 2005;102:17372–7.
38. Meister G, Buhler D, Laggerbauer B, Zobawa M, Lottspeich F, Fischer U. Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. Hum Mol Genet 2000;9:1977–86.
39. Mourelatos Z, Dostie J, Paushkin S, et al. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev 2002;16:720–8.
40. Dostie J, Mourelatos Z, Yang M, Sharma A, Dreyfuss G. Numerous microRNPs in neuronal cells containing novel microRNAs. RNA 2003;9:180–6.
41. Nelson PT, Hatzigeorgiou AG, Mourelatos Z. miRNP: mRNA association in polyribosomes in a human neuronal cell line. RNA 2004;10:387–94.
42. Murashov AK, Chintalgattu V, Islamov RR, et al. RNAi pathway is functional in peripheral nerve axons. FASEB J 2007;21:656–70.
43. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science 2002;297:2056–60.
44. Wan D, He M, Wang J, et al. Two variants of the human hepatocellular carcinoma-associated HCAP1 gene and their effect on the growth of the human liver cancer cell line Hep3B. Genes Chromosomes Cancer 2004;39:48–58.
45. Protopopov A, Kashuba V, Zabarovska VI, et al. An integrated physical and gene map of the 3.5-Mb chromosome 3p21.3 (AP20) region implicated in major human epithelial malignancies. Cancer Res 2003;63:404–12.
46. Lujambio A, Ropero S, Ballestar E, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 2007;67:1424–9.
47. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20.
48. Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol 2006;411:352–69.
49. Dyrskjot L, Kruhoffer M, Thykjaer T, et al. Gene expression in the urinary bladder: a common carcinoma *in situ* gene expression signature exists disregarding histopathological classification. Cancer Res 2004;64:4040–8.
50. Tsuruta H, Kishimoto H, Sasaki T, et al. Hyperplasia and carcinomas in Pten-deficient mice and reduced PTEN protein in human bladder cancer patients. Cancer Res 2006;66:8389–96.