

An miR-502–Binding Site Single-Nucleotide Polymorphism in the 3′-Untranslated Region of the *SET8* Gene Is Associated with Early Age of Breast Cancer Onset

Fengju Song,¹ Hong Zheng,¹ Ben Liu,¹ Sheng Wei,² Hongji Dai,¹ Lina Zhang,¹ George A. Calin,³ Xishan Hao,¹ Qingyi Wei,² Wei Zhang,⁴ and Kexin Chen¹

Abstract **Purpose:** MicroRNAs regulate gene expression by binding to the 3′-untranslated region (UTR) of target genes. Single-nucleotide polymorphisms of critical genes may affect their regulation by microRNAs. We have identified a single-nucleotide polymorphism within the miR-502 seed binding region in the 3′-UTR of the *SET8* gene. *SET8* methylates TP53 and regulates genome stability. We investigated the role of this *SET8* single-nucleotide polymorphism and in concert with the *TP53* codon 72 single-nucleotide polymorphism in the propensity for onset of breast cancer. **Experimental Design:** We measured the *SET8* single-nucleotide polymorphisms in a case-control study on 1,110 breast cancer cases and 1,097 controls. **Results:** The *SET8* CC and *TP53* GG genotypes were independently associated with an earlier age of breast cancer onset in an allele-dose-dependent manner (for *SET8*, 52.2 years for TT, 51.4 for TC, and 49.5 for CC; and for *TP53*, 53.1 years for CC, 51.5 for GC, 50.7 for GG). Individuals with combined *SET8* CC and *TP53* GG genotypes developed cancer at a median age of 47.7 years as compared with 54.6 years for individuals with combined *SET8* TT and *TP53* CC genotypes. In the 51 breast cancer tissue samples tested, the *SET8* CC genotype was associated with reduced *SET8*, but not miR-502, transcript levels. **Conclusions:** These data suggest that the miR-502–binding site single-nucleotide polymorphism in the 3′-UTR of *SET8* modulates *SET8* expression and contributes to the early development of breast cancer, either independently or together with the *TP53* codon 72 single-nucleotide polymorphism. Larger studies with multiethnic groups are warranted to validate our findings. (Clin Cancer Res 2009;15(19):6292–300)

Breast cancer is the most common cancer in women worldwide, accounting for 23% of all cancers (1). Mutations in the *BRCA1* and *BRCA2* genes are the major factors that predispose women to familial breast cancer (2). Other genetic events such as mutations in the *TP53* tumor suppressor gene and amplification of the *ErbB2* oncogene are also important factors in breast cancer etiology (3, 4). The *TP53* mutations associated with Li-Fraumeni syndrome contribute to an early age of onset of several cancers, including breast cancer (5, 6). Although mutational events in the breast cancer genome will likely be revealed further by whole-genome sequencing efforts, breast cancers do not necessarily initiate from a somatic mutation event. Instead,

polymorphisms in critical genes may have a subtle but accumulating impact on cells. When coupled with environmental exposure, these polymorphisms may gradually lead to changes in the genome and development of cancer.

Epidemiologic studies have begun to systematically interrogate single-nucleotide polymorphisms in critical genes to unravel specific genotypes that contribute to an increased risk for cancer development and/or a specific pattern of cancer development such as early age of onset (7). The single-nucleotide polymorphism at codon 72 of the *TP53* gene resulting in two different amino acids (Pro or Arg) is one of the most extensively studied single-nucleotide polymorphisms that have been

Authors' Affiliations: ¹Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Hospital and Institute, Tianjin, People's Republic of China and Departments of ²Epidemiology, ³Experimental Therapeutics, and ⁴Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas

Received 4/2/09; revised 6/14/09; accepted 6/29/09; published OnlineFirst 9/29/09.

Grant support: National Natural Science Foundation of China grant 30872172 (K. Chen).

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.

Note: F. Song and H. Zheng contributed equally to this work.

The Tissue Bank is jointly supported by Tianjin Cancer Institute and Hospital and The National Foundation for Cancer Research.

Requests for reprints: Kexin Chen, Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Hospital and Institute, Tianjin, People's Republic of China 300060. Phone: 86-22-23340123-5226; Fax: 86-22-23372231; E-mail: chenkexin1963@yahoo.com and Wei Zhang, Department of Pathology, Unit 85, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. E-mail: wzhang@mdanderson.org.

© 2009 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-09-0826

Translational Relevance

In this study, we investigated for the first time the role of a common single-nucleotide polymorphism of the *SET8* gene in the etiology of breast cancer. We found that the *SET8* C allele interacts synergistically with the *TP53* G allele in a dose-dependent manner to lower the age of onset of breast cancer. Individuals with the *SET8* CC and *TP53* GG genotypes developed cancer at a median age of 47.7 years as compared with 54.6 years for individuals with the *SET8* TT and *TP53* CC genotypes. Our findings suggest that genetic variations in microRNA-regulated genes are important for identifying individuals at high risk for developing breast cancer, particularly at an early age.

shown to influence cancer incidence in a number of cancer types, including cervical cancer (8). Single-nucleotide polymorphisms in other TP53 pathway-related genes, such as those in the promoter regions of the *MDM2* and *TP53BP1* genes, have also been shown to influence cancer risk (9, 10). Therefore, the picture that emerges is that single-nucleotide polymorphisms in genes of the TP53 pathway, which is critical for genome stability and cellular metabolism (11, 12), may be more important than single-nucleotide polymorphisms in other genes as risk factors for cancer development.

MicroRNAs are a class of noncoding small RNAs that regulate gene expression by base pairing with sequences within the 3'-untranslated regions (UTR) of target mRNAs, leading to mRNA cleavage and/or translational repression (13). To date, >700 microRNAs have been identified in humans, and these microRNAs are proposed to regulate the expression of at least 30% of all protein-coding genes (14). Because microRNAs have a profound effect on mRNA, an understanding of genetic variations in microRNA-mRNA interactions and the associated biological endpoints is important to understand their functional and evolutionary significance (15, 16). Single-nucleotide polymorphisms located in microRNA binding sites (microRNA-binding single-nucleotide polymorphisms) have been shown to disrupt microRNA-target interactions, resulting in the deregulation of target gene expression (17, 18). Recent studies have also shown that microRNA-binding single-nucleotide polymorphisms that regulate oncogenes, tumor suppressor genes, or genes in oncogenic pathways contribute to carcinogenesis (19).

Yu et al. (20) did a genomewide analysis of single-nucleotide polymorphisms located within the microRNA binding sites in the 3'-UTRs of various human genes and found that the allele frequencies of some microRNA-binding single-nucleotide polymorphisms reported in the human cancer expressed sequence tag libraries and the National Center for Biotechnology Information single-nucleotide polymorphism database (dbSNP⁵) were significantly different, suggesting that certain single-nucleotide polymorphisms located within microRNA binding sites may be overrepresented in individuals with cancer. Among the single-nucleotide polymorphisms that exhibit this pattern is the one found within the miR-502 binding site in the 3'-

UTR of the *SET8* gene. *SET8* (also known as *PR-SET7*; located on chromosome 12q24.31) encodes a histone H4-Lys-20-specific methyltransferase that is implicated in cell cycle-dependent transcriptional silencing and mitotic regulation (21). It was recently reported that *SET8* monomethylates TP53 at Lys-382 and that depletion of *SET8* augments the proapoptotic and checkpoint activation functions of TP53 (22). A total of 129 variants of the *SET8* gene have been reported on the *SET8* gene in the dbSNP database. Of these, three have been validated (rs3881293, T→A; rs6576, A→C; rs16917496, T→C); two (rs6576, A→C; rs16917496, T→C) are common (i.e., minor allele frequency, ≥0.05), and one of these two (rs16917496, T→C) is located within the miR-502 binding site in the 3'-UTR (Fig. 1A).

We therefore hypothesized that the miR-502-binding site single-nucleotide polymorphism of the *SET8* gene modulates *SET8* expression, thus influencing its methylation of TP53 and contributing to the TP53 pathway-mediated breast cancer development. To test this hypothesis from an epidemiologic perspective, we conducted a large case-control study in a Chinese population. Our study provides evidence that the miR-502-binding site single-nucleotide polymorphism within the *SET8* gene, as well as the *TP53* codon 72 single-nucleotide polymorphism, is associated with an early age of onset of breast cancer.

Materials and Methods

Patients and controls. The cases were recruited from the Breast Cancer Research Center in Tianjin Medical University Cancer Hospital, and the clinical information was acquired from the Tianjin Cancer Registry. This study included patients with newly diagnosed and histologically confirmed breast cancer, who were consecutively recruited between January 2007 and February 2008. The response rate of the eligible incident cases that we approached for recruitment was ~95%. The cancer-free control subjects were recruited during the same period and were genetically unrelated women living in the nearby community who were frequency matched to the cases by age (± 5 y). The response rate of the eligible controls that we approached for recruitment was ~90%. After signing an informed consent form, all subjects enrolled in the study were interviewed to collect demographic data and information about major risk factors, including family history of disease. For cases, we also collected information about tumor features and disease severity, including morphology, tumor size, lymph node metastasis, organ metastasis, tumor stage, and estrogen and progesterone receptor status. Each subject donated 20 mL of blood that was collected into heparinized tubes and used for biomarker assays, including DNA extraction and genotyping. Fifty-one breast cancer cases with the CC or TT genotypes were randomly selected from cases with one of these two genotypes. Tumor tissues for these selected cases were obtained from the Tissue Bank Facility of Tianjin Medical University Cancer Institute and Hospital, which collects all solid tumor tissues from surgical patients when possible and upon the approval of the Institutional Review Board.

Genotyping. A leukocyte cell pellet obtained from the buffy coat by centrifugation of 1 mL of each whole blood sample was used for DNA extraction. Genomic DNA was isolated using the QIAGEN DNA Blood Mini Kit (QIAGEN, Inc.), according to the manufacturer's instructions. RFLP PCR was used to identify the genotypes of the selected *SET8* (rs16917496 C/T) polymorphisms within the 3'-UTR. Each PCR was done in a 25- μ L reaction mixture containing ~50 ng of genomic DNA template, 12.5 pmol of each primer, 0.1 mmol/L of each deoxynucleotide triphosphate, 1 \times PCR buffer (50 mmol/L KCl, 10 mmol/L Tris-HCl, and 0.1% Triton X-100), 1.5 mmol/L MgCl₂, and 1.5 units of Taq polymerase (Promega Corp.). The PCR profile consisted of an

⁵ <http://www.ncbi.nlm.nih.gov/SNP/>

initial melting step of 95°C for 5 min, 35 cycles of 95°C for 45 s, 66°C for 40 s, and 72°C for 30 s, and a final extension step of 72°C for 10 min. In this process, the primers to the miR-502 binding site (5'-GGCCTCAGCAGGGTGTAC-3' and 5'-GTTCCCAGGAGGATGCTTAC-3') produced a 308-bp DNA fragment, which was digested by SWAI (New England BioLabs, Inc.) overnight at 25°C. The digested DNA product was separated on a 2.0% NuSieve 3:1 agarose gel (FMC BioProducts), which was stained with ethidium bromide and photographed with Polaroid film. Allele C lacks the SwI restriction site and therefore produces a single 308-bp band, whereas allele T produces two bands (149 and 159 bp). Therefore, the TC heterozygote produces three bands of 149, 159, and 308 bp (Fig. 1B). The results of the RFLP-PCR analysis were evaluated without knowledge of the subjects' case-

control status. More than 10% of the samples were randomly selected for repeat analysis, and the results were in agreement with the original results. The CC, TC, and TT genotypes were also validated by direct sequencing of the PCR products (Fig. 1C). We first genotyped the T→C single-nucleotide polymorphism (rs16917496) in part of individuals in our Chinese study population and found that the C allele was the minor one, with a minor allele frequency of 0.31, which was quite different from the frequency reported in a White population in the dbSNP database (minor allele frequency of 0.22 for the T allele). Because of this seemingly big difference, we genotyped some DNA samples of different ethnic groups available at M.D. Anderson Cancer Center for a validation. We confirmed that the minor allele frequency of the C allele was 0.30 in 124 Chinese and 0.34 in 170 non-Hispanic Whites (data

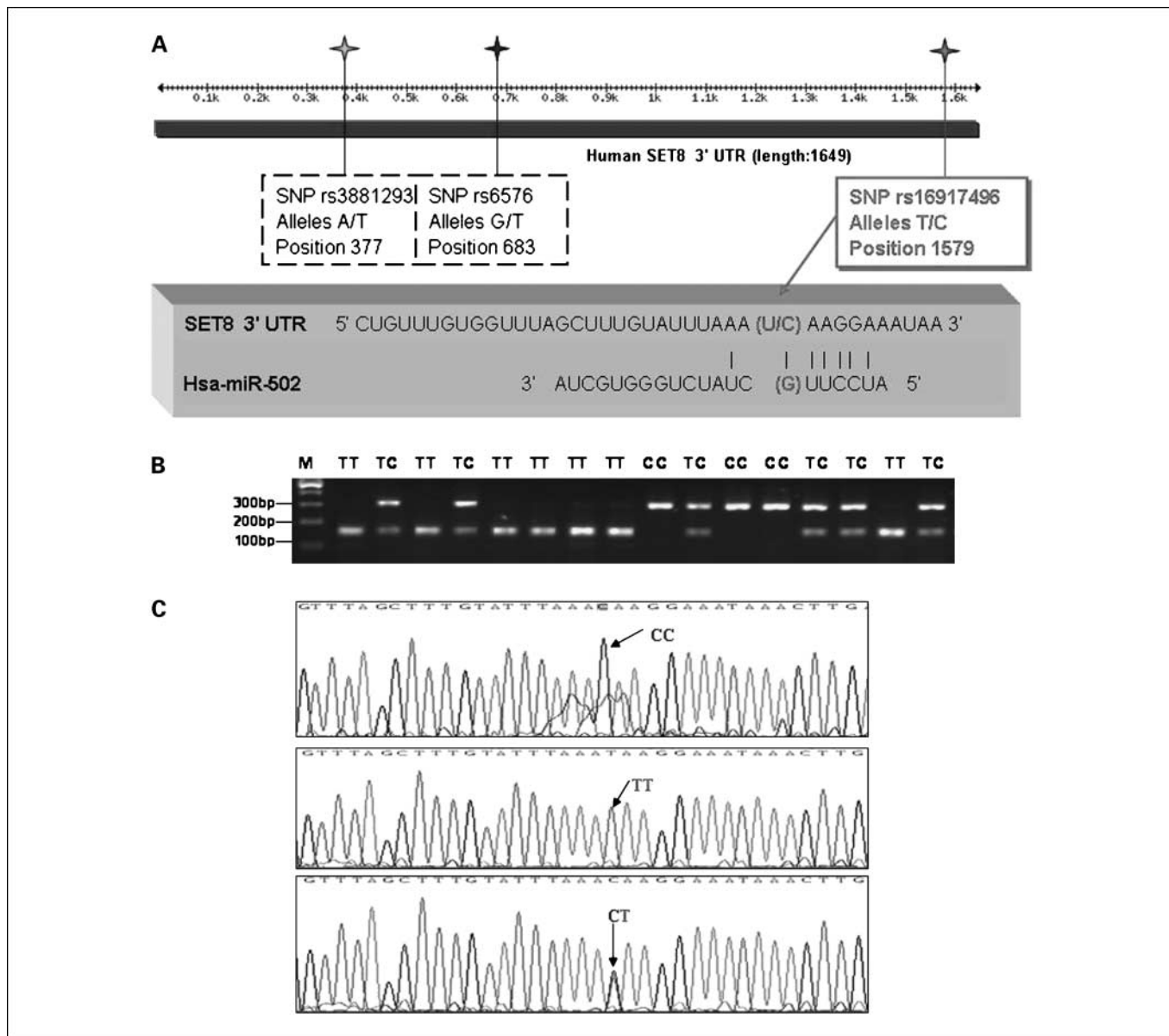


Fig. 1. Genotyping and sequence validation of the *SET8* (rs16917496) polymorphism. **A**, the three validated single-nucleotide polymorphisms and their exact locations within the *SET8* 3'-UTR. The sequence complementarity of *has-miR-502* and *SET8* 3'-UTR is shown here. The single-nucleotide polymorphism rs16917496 T→C is located within the seed region of the binding site with the C allele, perfectly matching the corresponding G allele in miR-502. **B**, RFLP-PCR genotyping method for the *SET8* (rs16917496) polymorphism. M, molecular markers. The CC genotype lacks the SWAI restriction site and produces a single 308-bp band; the TT genotype produces two fragments of 149 and 159 bp, whose sizes are too similar to be separated and appear as one band; the TC heterozygote produces three fragments of 149, 159, and 308 bp, appearing as two bands on the agarose gel. **C**, confirmation of the *SET8* genotypes for single-nucleotide polymorphism rs16917496 (TT, TC, and CC) by direct sequencing.

Table 1. Frequency distributions of selected variables in breast cancer cases and cancer-free controls

Variables	No. of subjects (%)		OR* (95% CI)	P [†]
	Cases (n = 1,110)	Controls (n = 1,097)		
Age (y)				
≤50	546 (49.19)	547 (49.86)		0.752
>50	564 (50.81)	550 (50.14)		
No. of pregnancies				
≤2	523 (47.12)	488 (44.48)	1.00	0.215
>2	587 (52.88)	609 (55.52)	0.98 (0.81-1.20)	
Duration of breast feeding (mo)				
≤12	536 (48.29)	427 (38.92)	1.00	<0.001
>12	574 (51.71)	670 (61.08)	0.62 (0.51-0.76)	
Menopause [‡]				
No	534 (48.37)	516 (47.91)	1.00	0.830
Yes	570 (51.63)	561 (52.09)	0.84 (0.63-1.12)	
Oral contraception [‡]				
Never	867 (82.81)	870 (83.25)	1.00	0.786
Ever	180 (17.19)	175 (16.75)	1.03 (0.80-1.33)	
Smoking status [‡]				
Never	941 (87.70)	1,007 (93.07)	1.00	<0.001
Ever	132 (12.30)	75 (6.93)	2.22 (1.60-3.10)	
Benign breast disease [‡]				
Never	808 (73.32)	982 (94.15)	1.00	<0.001
Ever	294 (26.68)	61 (5.85)	5.67 (4.18-7.70)	
Family history of any cancer ^{‡,§}				
No	761 (68.62)	965 (88.78)	1.00	<0.001
Yes	348 (31.38)	122 (11.22)	3.71 (2.90-4.76)	

Abbreviation: OR, odds ratio.

*Odds ratios are adjusted for age, duration of breast feeding, menopause, oral contraception, smoking status, benign breast disease, and family history of cancer.

[†]Two-sided χ^2 test.

[‡]Due to missing information, cases are n < 1,110 and controls are n < 1,097.

[§]First and second degree of relatives.

not shown). The TP53 codon 72 single-nucleotide polymorphism (rs1042522) was genotyped according to the published method (23).

Quantitative measurement of SET8 and miR-502 expression. Total RNA was isolated from the 51 frozen breast cancer tissues with known SET8 genotypes (CC or TT). The extraction and purification of total RNA were done using the Trizol reagent (Invitrogen) and ethanol precipitation, according to the manufacturer's instructions. RNA quality and concentration were determined using an Agilent 2100 Bioanalyzer (Agilent Technologies). The quantitative real-time PCR was done in a 96-well reaction plate (MicroAmp Optical 96-Well Reaction Plate, Applied Biosystems) on an ABI PRISM 7500 Sequence Detector System (Applied Biosystems), according to the manufacturer's instructions.

Reverse transcription-PCR for SET8 expression was done using Taqman one-step reverse transcription-PCR master mix reagent kit (Applied Biosystems). The Taqman probes (6-FAM dye-labeled probe) and primers for SET8 were designed using the Primer Express 3.0 Software (Applied Biosystems). Glyceraldehyde-3-phosphate dehydrogenase (GADPH) RNA was measured as an endogenous control to normalize for differences in the amount of total RNA used in each reaction (24). The probe and primer sequences used were as follows: probe for SET8, 5'-FAM-CCCTGTCCGAAGGAGCTCCAGGAAGA-TAMRA-3'; forward and reverse primers for SET8, 5'-CGCAAACCTACGGATTTC-3' and 5'-CGATGAGGTCAATCTTCATT-3', respectively; probe for GADPH, 5'-GAAGATGGTGATGGGATTTTC-3'; forward primer for GADPH, 5'-CGCAAACCTACGGATTTC-3'; and reverse primer for GADPH, 5'-GAA-GATGGTGATGGGATTTTC-3'. All primers and probes were synthesized by Sangon Corp. One-step reverse transcription-PCR was done in a 20 μ L reaction volume containing 100 ng total RNA, primer at a final concentration of 200 nmol/L, and the Taqman probes for both target genes and endogenous controls at a final concentration of 250 nmol/L

each. The PCR cycling parameters were as follows: reverse transcription at 48°C for 30 min, AmpliTaq activation at 95°C for 10 min, 40 cycles of denaturation at 95°C for 15 s, and annealing/extension at 52°C for 40 s. Each sample was analyzed in duplicate, and if the coefficient of variation of all reactions was <5%, the mean values of the duplicates were used. The expression level of SET8 relative to GADPH was calculated using the equation ratio = Ct_{SET8} / Ct_{GADPH} * 100%. The PCR products of the SET8 genotypes were validated by sequencing, and the results were compared with the GenBank sequence (accession number NM 020382).

MiR-502 expression was measured by reverse transcription-PCR according to the Taqman microRNA Assay protocol (Applied Biosystems). The microRNA cDNA was prepared from 10 ng of total RNA using the Taqman microRNA Reverse Transcription kit (Applied Biosystems). All reactions were done in a 15- μ L reaction volume. The mixtures were incubated at 16°C for 30 min, followed by a cycle of 42°C for 30 min and 85°C for 5 min. The cDNA samples were stored at -20°C until analysis. Each PCR was done in triplicate and contained 1.33 μ L of reverse transcription product, Taqman Universal PCR Master Mix, No AmpErase UNG, and probe mix from the Taqman microRNA Assay (both products from Applied Biosystems). The 20- μ L reaction mixtures were incubated in a 96-well optical plate at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and annealing/extension at 60°C for 60 s. The amount of miR-502 relative to U6 small nuclear RNA was calculated using the equation ratio = Ct_{miR-502} / Ct_{U6} * 100% to normalize the results to the endogenous U6 small nuclear RNA expression level.

Statistical analysis. We used the χ^2 test to compare the differences in the frequency distributions of demographic variables, other known risk factors, and the alleles and genotypes of the SET8 and

TP53 polymorphisms between the cases and controls. We also tested the genotype distributions of the controls for Hardy-Weinberg equilibrium. In addition, we used unconditional univariate and multivariate logistic regression analyses to examine the association between the single-nucleotide polymorphisms and breast cancer risk by estimating odds ratios and 95% confidence intervals (95% CI), with and without adjustment for age and other known risk factors. We further stratified the genotype distributions by age for age of onset of breast cancer in the case-only analysis. We also did a stratified analysis of the genotype data from breast cancer patients by family history and clinical variables, including morphology, tumor size, lymph node metastasis, organ metastasis, tumor stage, and estrogen receptor and progesterone receptor status. The TTEST procedure was used to compare the expression levels of miR-502 and *SET8* between cases with different *SET8* genotypes. All statistical tests were two-sided, and a *P* value of 0.05 was considered significant. We used the SAS software version 9.0 (SAS Institute) for all statistical analyses.

Results

Case-control analysis of *SET8* single-nucleotide polymorphisms. The single-nucleotide polymorphism analysis included 1,110 breast cancer cases and 1,097 controls, and the distributions of known risk factors in cases and controls are shown in Table 1. Age was adequately matched between cases and controls (*P* = 0.752).

The frequencies of the *SET8* CC, CT, and TT genotypes in controls were 9.48%, 43.30%, and 47.22%, respectively, and the genotype distribution was in Hardy-Weinberg equilibrium (*P* = 0.745). We found that the C allele was the minor one (minor allele frequency, 0.31), which was consistent with the results from two independent samples of Chinese and non-Hispanic Whites (the minor allele frequency of the C allele was 0.30 in 124 Chinese and 0.34 in 170 non-Hispanic Whites) studied at M.D. Anderson. The *TP53* single-nucleotide polymor-

phism distributions in both ethnic groups were similar to those previously reported. Overall, there was no statistically significant difference in the frequency distributions of either the *SET8* or *TP53* genotypes between the cases and controls (*P* = 0.627 and 0.519, respectively).

Epidemiologic studies have shown a unique pattern of breast cancer incidence in China, in which there are two incidence peaks, with the first peak coming at a relatively early age. Thus, we investigated whether the *SET8* and *TP53* genotypes were associated with breast cancer development at an early age. For premenopausal women, a significantly higher risk for breast cancer was associated with the *SET8* CC genotype (odds ratio, 1.66; 95% CI, 1.06-2.61) than with the *SET8* TT genotype. For postmenopausal women, however, we did not find a significant association between the *SET8* CC and TT genotypes and breast cancer risk (odds ratio, 0.76; 95% CI, 0.47-1.22). The interaction between the *SET8* genotypes and menopause states (pre versus post) was not statistically significant (*P* = 0.118; Table 2).

Case-only analysis of *SET8* genotypes by the selected known risk factors. Further case-only analysis revealed an association between *SET8* genotype and mean age of diagnosis of breast cancer in an allele-dependent manner (i.e., a younger age of onset with increasing number of C alleles): 52.2 years (range, 25-88 years) for TT, 51.4 years (range, 23-87 years) for TC, and 49.5 years (range, 29-78 years) for CC (trend test; *P* = 0.022). Likewise, the *TP53* genotypes were also associated with age of onset of breast cancer in an allele-dependent manner (i.e., a younger age of onset with an increasing number of G alleles): 53.1 years (range, 28-88 years) for CC, 51.5 years (range, 23-85 years) for GC, 50.7 years (range, 24-83 years) for GG (trend test; *P* = 0.016). Among these single genotype groups, individuals carrying the *SET8* CC genotype had the youngest age of onset (median, 49.5 years). When the data on the *SET8* CC and *TP53* GG genotypes were combined, the early age of onset of breast

Table 2. Multivariate-adjusted odds ratios for risk for breast cancer associated with *SET8* and *TP53* polymorphisms by menopause status

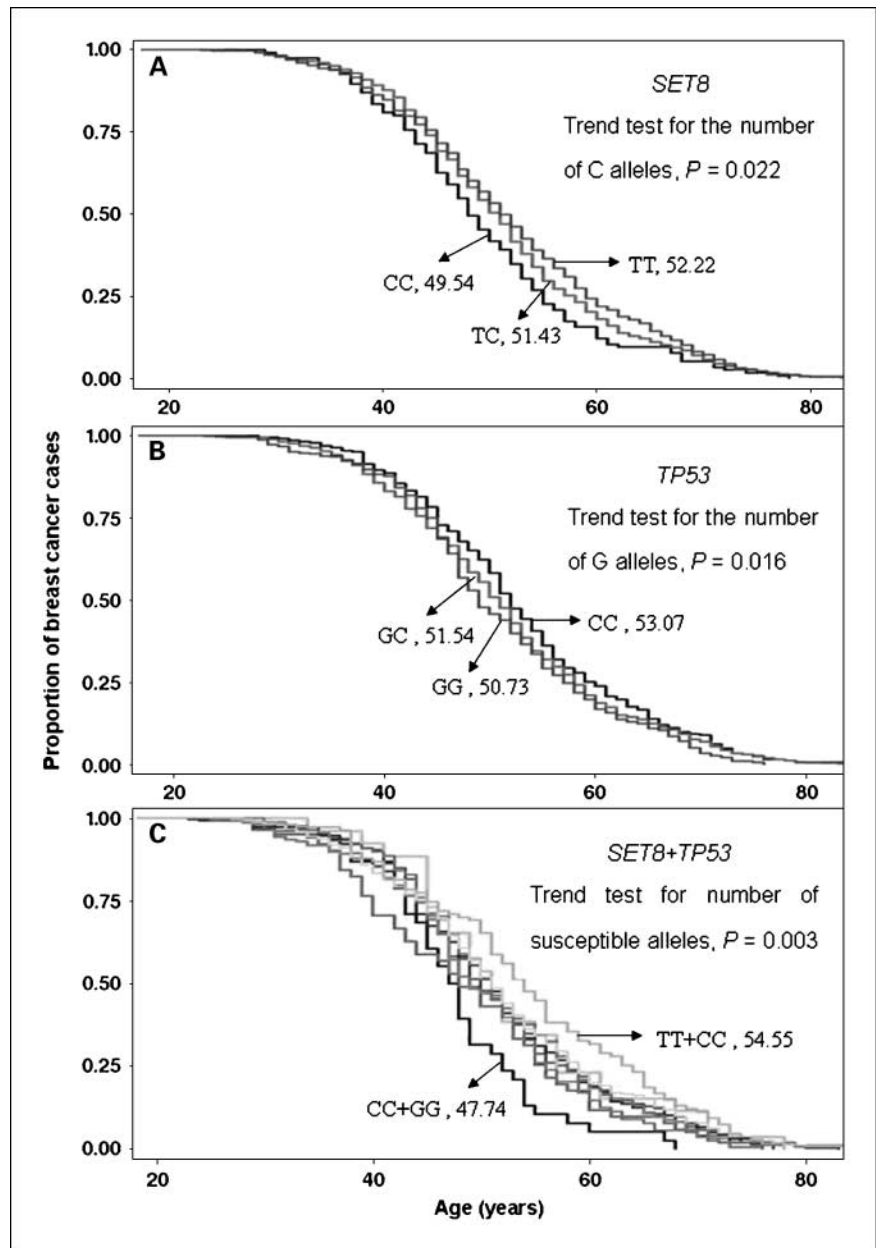
Genotype	All subjects			Premenopause			Postmenopause		
	Cases (n = 1,110)	Controls (n = 1,097)	OR* (95% CI)	Cases (n = 534)	Controls (n = 516)	OR* (95% CI)	Cases (n = 570)	Controls (n = 561)	OR* (95% CI)
<i>SET8</i> (T/C) [†]									
TT	504 (45.41)	518 (47.22)	1.00	235 (44.01)	241 (46.71)	1.00	268 (57.02)	267 (47.59)	1.00
CT	491 (44.23)	475 (43.30)	0.96 (0.79-1.16)	229 (42.88)	229 (44.38)	0.94 (0.71-1.24)	259 (45.44)	237 (42.25)	0.96 (0.73-1.26)
CC	115 (10.36)	104 (9.48)	1.18 (0.86-1.62)	70 (13.11)	46 (8.91)	1.66 (1.06-2.61)	43 (7.54)	57 (10.16)	0.76 (0.47-1.22)
Trend test <i>P</i> [‡]			0.335			0.095			0.594
<i>TP53</i> (G/C) [†]									
GG	341 (30.72)	355 (32.36)	1.00	175 (32.77)	174 (33.72)	1.00	164 (28.77)	175 (31.19)	1.00
GC	547 (49.28)	514 (46.86)	1.12 (0.91-1.38)	258 (48.31)	239 (46.32)	1.16 (0.86-1.56)	286 (50.18)	269 (47.95)	1.07 (0.79-1.44)
CC	222 (20.00)	228 (20.78)	1.03 (0.80-1.34)	101 (18.91)	103 (19.96)	0.93 (0.64-1.36)	120 (21.05)	117 (20.86)	1.15 (0.80-1.65)
Trend test <i>P</i> [‡]			0.778			0.982			0.534

*Odds ratios are adjusted for age, duration of breast feeding, menopause, oral contraception, smoking status, benign breast disease, and family history of cancer.

[†]Minor allele frequency: *SET8* C allele, 0.325 for cases and 0.311 for controls; *TP53* C allele, 0.446 for cases and 0.442 for controls.

[‡]*P* value of the test for interaction is 0.118 for menopause and *SET8* genotypes and 0.634 for menopause and *TP53* genotypes.

Fig. 2. Comparison of age onset of breast cancer by *SET8* genotypes, *TP53* genotypes, and their combination in the case-only analysis of 1,110 cases. *A*, age of onset of breast cancer decreases with increasing number of *SET8* C alleles. *B*, age of onset of breast cancer decreases with increasing number of *TP53* G alleles. *C*, age of onset of breast cancer decreases with increasing number of *SET8* C and *TP53* G alleles.



cancer also exhibited a genotype dose-dependent pattern ($P = 0.003$); in particular, individuals with *SET8* CC and *TP53* GG genotypes developed cancer at a median age of 47.7 years as compared with 54.6 years for individuals with the *SET8* TT and *TP53* CC genotypes (Fig. 2).

Impact of *SET8* genotypes on *SET8* expression and miR-502 levels. To further evaluate the biological relevance of the *SET8* single-nucleotide polymorphisms, we measured the expression levels of *SET8* mRNA in breast cancer tissues from 51 cases with different *SET8* genotypes. We found that the cases with the *SET8* CC genotype ($n = 24$), which has the perfect complementary seed sequence for miR-502, had lower levels of *SET8* expression than did cases with the TT genotype ($n = 27$; $P = 0.018$; Fig. 3A). In contrast, the *SET8* expression levels were not significantly different between cases with different *TP53* genotypes (data not shown).

To exclude the possibility that patients with different *SET8* genotypes may have different miR-502 levels, which would contribute to the differential *SET8* expression levels, we used reverse transcription-PCR to measure the miR-502 expression levels in the same 51 tumor tissue samples. We did not find a significant difference in miR-502 expression levels among patients with the CC ($n = 24$) or TT ($n = 27$) genotypes ($P = 0.944$; Fig. 3B).

Discussion

We show that a single-nucleotide polymorphism in the miR-502 binding site modulates *SET8* gene expression and is associated with the age of onset of breast cancer. In this study, we investigated for the first time the role of a common single-nucleotide polymorphism in the *SET8* gene in the etiology of

breast cancer. This single-nucleotide polymorphism is particularly interesting because it is located in the *SET8* 3'-UTR seed region for miR-502 binding. In addition, *SET8* encodes a key modifier of TP53, which controls a pathway that is critical for genome stability and cancer risk (25). Our analyses reveal that this single-nucleotide polymorphism is significantly involved in breast cancer etiology. In our case-control analysis, we found that, compared with the *SET8* TT genotype, the *SET8* CC genotype is associated with an increased risk for breast cancer in premenopausal women, and the C allele is associated with an increased risk for breast cancer in this age group in an allele-dose-dependent manner. In comparison, we did not observe such an association between breast cancer risk and the *TP53* codon 72 single-nucleotide polymorphism. Interestingly however, this *SET8* C allele interacts synergistically with the *TP53* G (Pro) allele in a dose-dependent manner to effect age of onset of breast cancer. Specifically, individuals with the *SET8* CC and *TP53* GG genotypes developed cancer at 47.7 years of age as compared with 54.6 years of age for individuals with the *SET8* TT and *TP53* CC genotypes. Thus, our study reveals a novel single-nucleotide polymorphism in the TP53 pathway gene *SET8* that is critical for the early age of onset of sporadic breast cancer. Furthermore, our study provides another example of single-nucleotide polymorphism interactions among TP53 pathway genes, which include *MDM2* and *TP53BP1*, in conferring a joint effect on cancer risk.

TP53 is involved in mediating the cellular response to various stresses, mainly by activating or repressing the transcription of a number of genes involved in cell cycle arrest, cell senescence, apoptosis, DNA repair, and angiogenesis (26). As a tumor suppressor, *TP53* plays an undisputed role in cellular homeostasis (27). Abundant data from molecular, pathologic, and transgenic animal studies support an important role of *TP53* inactivation in mammary carcinogenesis (28, 29). It is not surprising that genes in the TP53 pathway interact with *TP53* at multiple levels (e.g., protein, promoter, and microRNA binding site) to modulate the functions of TP53 in these biological processes.

Previously known cancer-related single-nucleotide polymorphisms in the TP53 pathway either exist in the *TP53* coding region (e.g., *TP53* codon 72), which affects the functions of the protein, or in the promoter regions of *TP53* pathway genes (e.g., *MDM2* and *TP53BP1*), which regulate the levels of the gene products that in turn regulate TP53 (30). *TP53* single-nucleotide polymorphisms also functionally interact with single-nucleotide polymorphisms of its target genes. For example, the *TP53* codon 72 polymorphism was shown to interact with the *BRCA2* -26G→A single-nucleotide polymorphism within the 5'-UTR region to increase susceptibility to sporadic breast cancer (31). What is unique about the *SET8* single-nucleotide polymorphism is that it is located within the 3'-UTR region of the *SET8* gene, wherein one of the single-nucleotide polymorphism variants forms a perfect binding site for the microRNA molecule miR-502. MicroRNAs are a new class of molecules that do not encode proteins but have been shown to play an important role in regulating protein-coding genes. Therefore, this single-nucleotide polymorphism represents a new class of regulatory single-nucleotide polymorphisms that link the noncoding genome with the coding genome in modulating the risk for human diseases.

This class of polymorphisms within microRNA binding sites, as well as in microRNAs or their precursors, is emerging as a key risk factor for disease phenotypes (32, 33). Shen et al. (34) recently reported that a G→C single-nucleotide polymorphism (rs2910164) located within the *miR-146a* precursor leads to a change from a G:U pair to a C:U mismatch in its stem region and that breast cancer patients who had at least one *miR-146a* variant C allele were diagnosed with the disease at an earlier age than those without this variant allele (median age, 45 versus 56 years; $P = 0.029$). In the study of Landi et al. (4), two of eight microRNA-binding site single-nucleotide polymorphisms studied were significantly associated with a higher risk for colorectal cancer. Brendle et al. (35) examined single-nucleotide polymorphisms in the predicted microRNA-binding sites in integrin genes and found no association with breast cancer risk, although they found that *ITGB4* is itself a prognostic marker.

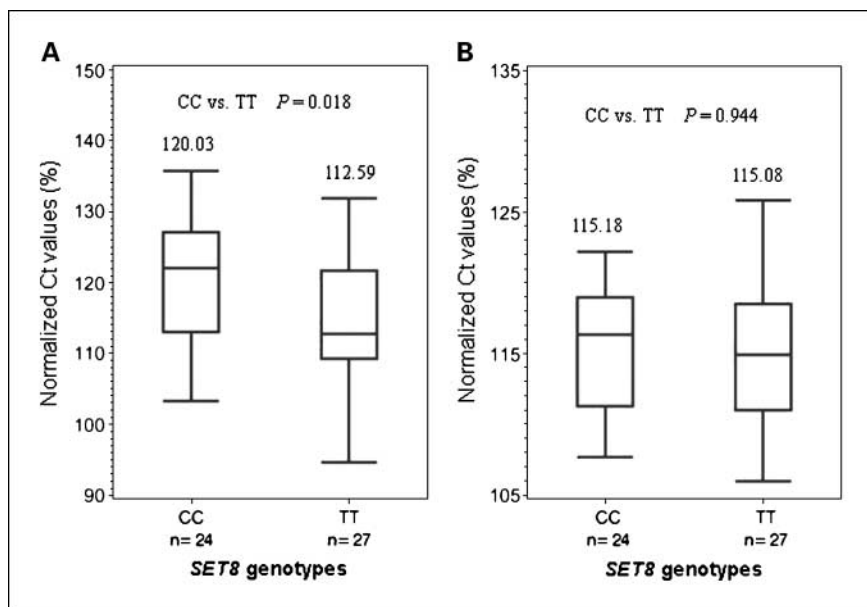


Fig. 3. Relative expression levels of *SET8* mRNA and miR-502 in tumor tissues from 51 breast cancer cases with known *SET8* genotypes. **A**, quantitative real-time reverse transcription-PCR was used to measure the relative expression of *SET8* mRNA. Housekeeping gene *GADPH* was used as the reference. The threshold cycle or Ct value is the PCR cycle at which a significant increase in fluorescent signal is first detected. Ct values form the basis for quantitative comparison of genes in the real-time PCR. The smaller the Ct value, the higher the level of the measured target RNA. We normalized the relative levels of *SET8* to that of *GADPH* using the equation ratio = $Ct_{SET8} / Ct_{GADPH} * 100\%$. A higher normalized Ct value represents less expression. The TT genotype has lower normalized Ct values and, thus, higher *SET8* expression levels, whereas the CC genotype has higher relative Ct values, representing a lower expression level ($P = 0.018$). **B**, the expression level of miR-502 was measured by real-time PCR. The expression levels are similar between the two groups with the *SET8* CC and TT genotypes, $P = 0.944$.

SET8 has a well-defined function in the TP53 pathway by monomethylating TP53 at Lys-382 and suppressing the p53-mediated transcription activation of target genes (22); however, the function of SET8 is likely to be much broader. The most notable function of SET8 is the modulation of chromatin dynamics as a histone-modifying enzyme (36). It is well established that epigenetic alterations of the histone code, which include loss of H3K4 and H4K20 trimethylation, gain of H3K9 methylation, and H3K27 trimethylation and deacetylation of histones H3 and H4K16, contribute to the initiation of multiple cancers, such as lymphomas, colorectal adenocarcinomas, and squamous cell carcinoma (37, 38). These epigenetic changes appear at an early stage of cancer development and accumulate during carcinogenic progression (38). Although direct evidence for the role of SET8 in cancer pathogenesis is still missing, it is conceivable that SET8 may be a key contributing factor to cancer development based on its histone (21, 39)- and nonhistone (22)-modifying functions. Furthermore, recent experiments have shown a connection between SET8 and the cell cycle and DNA replication process (40, 41). SET8 is required for normal S-phase progression (41) and controls the G₁/S transition, possibly by blocking histone Lys acetylation by binding to the histone H4 N-terminal tail (42). SET8 is also required for G₂/M phase transition because a reduction in SET8 has been shown to lead to the inability of cells to progress past G₂ and cause global chromosome condensation failure, aberrant centrosome amplification, and substantial DNA damage (43). EZH2, another member of the Lys methyltransferase family, is highly expressed in a large set of human primary tumors and is linked to breast cancer invasion and progression (44, 45). Our results provide the first evidence that the expression levels of SET8 are regulated by a 3'-UTR genotype that contributes to the differential susceptibility to breast cancer.

MicroRNAs have been shown to regulate epigenetic functions, which include DNA methylation and histone modification. Recently, reports show that microRNAs can control the expression of epigenetic mediators such as DNA methyltransferases and histone deacetylases (46, 47). For example, miR-1,

miR-133, and miR-140 directly target histone deacetylase 4 and regulate muscle and bone differentiation (46, 48). Here, we provide the first epidemiologic evidence showing that a microRNA may regulate histone methyltransferases. In addition, we show a novel mechanism by which microRNAs may differentially regulate target genes through genetic polymorphisms. Whereas the SET8-regulating miR-502 itself had similar expression levels in breast cancers with different SET8 genotypes (i.e., CC, CT, and TT), the observed difference in the expression levels of the target SET8 gene is likely due to the different complementary miR-502 binding efficiencies of the different genotypes.

Although studies on single-nucleotide polymorphisms within microRNA binding sites are at an early stage and the results from reported studies need replication and laboratory-based functional validation, our findings are nevertheless encouraging in the search for cancer-predisposing alleles involving microRNAs. It is increasingly recognized that microRNAs are critical in the gene expression regulatory network, with the predicted proportion of microRNA-regulated protein coding genes being >30% of all protein coding genes. MicroRNAs have also been increasingly emphasized as key molecules that are critical for susceptibility to, development and progression of, and therapeutic response to many complex diseases, including cancer (49). Our findings further suggest that genetic variations in microRNA-related genes are important in identifying individuals at high risk for developing breast cancer, particularly at an early age.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Hongwei Han and Lei Lei from Tianjin Cancer Institute and Hospital for their technical assistance and Dr. Kate J. Newberry at The University of Texas M.D. Anderson Cancer Center Scientific Publication for editing this manuscript.

References

- Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. *CA Cancer J Clin* 2006;56:106-30.
- Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell* 2007;11:103-5.
- Suspitsin EN, Due EU, Vu P, Hirvonen A, Borresen-Dale AL, Imyanitov EN. TP53 mutations in synchronous and metachronous bilateral breast carcinomas. *Cancer Genet Cytogenet* 2008;184:119-21.
- Landi D, Gemignani F, Naccarati A, et al. Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis* 2008;29:579-84.
- van Hest LP, Ruijs MW, Wagner A, et al. Two TP53 germline mutations in a classical Li-Fraumeni syndrome family. *Fam Cancer* 2007;6:311-6.
- Tabori U, Nanda S, Druker H, Lees J, Malkin D. Younger age of cancer initiation is associated with shorter telomere length in Li-Fraumeni syndrome. *Cancer Res* 2007;67:1415-8.
- Lum SS, Chua HW, Li H, et al. MDM2 SNP309 G allele increases risk but the T allele is associated with earlier onset age of sporadic breast cancers in the Chinese population. *Carcinogenesis* 2008;29:754-61.
- Santos AM, Sousa H, Catarino R, et al. TP53 codon 72 polymorphism and risk for cervical cancer in Portugal. *Cancer Genet Cytogenet* 2005;159:143-7.
- Chen K, Hu Z, Wang LE, et al. Single-nucleotide polymorphisms at the TP53-binding or responsive promoter regions of *BAX* and *BCL2* genes and risk of squamous cell carcinoma of the head and neck. *Carcinogenesis* 2007;28:2008-12.
- Cox DG, Deer D, Guo Q, et al. The p53 Arg72-Pro and MDM2-309 polymorphisms and risk of breast cancer in the nurses' health studies. *Cancer Causes Control* 2007;18:621-5.
- Hussain SP, Harris CC. p53 Biological network: at the crossroads of the cellular-stress response pathway and molecular carcinogenesis. *J Nippon Med Sch* 2006;73:54-64.
- Rangel-Lopez A, Pina-Sanchez P, Salcedo M. Genetic variations of the tumor suppressor TP53: outstanding and strategies of analysis. *Rev Invest Clin* 2006;58:254-64.
- Fabbri M, Croce CM, Calin GA. MicroRNAs. *Cancer J* 2008;14:1-6.
- 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.
- Chin LJ, Ratner E, Leng S, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res* 2008;68:8535-40.
- Ghosh T, Soni K, Scaria V, Halimani M, Bhattacharjee C, Pillai B. MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic β -actin gene. *Nucleic Acids Res* 2008;36:6318-32.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet* 2007;39:1278-84.
- Wang G, van der Walt JM, Mayhew G, et al. Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of α -synuclein. *Am J Hum Genet* 2008;82:283-9.
- Zhu S, Wu H, Wu F, Nie D, Sheng S, Mo YY. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res* 2008.

20. Yu Z, Li Z, Jolicoeur N, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res* 2007;35:4535–41.
21. Couture JF, Collazo E, Brunzelle JS, Trievel RC. Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev* 2005;19:1455–65.
22. Shi X, Kachirskaia I, Yamaguchi H, et al. Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol Cell* 2007;27:636–46.
23. Chen RH, Chang CT, Wang TY, Huang WL, Tsai CH, Tsai FJ. p53 Codon 72 proline/arginine polymorphism and autoimmune thyroid diseases. *J Clin Lab Anal* 2008;22:321–6.
24. Chuikov S, Kurash JK, Wilson JR, et al. Regulation of p53 activity through lysine methylation. *Nature* 2004;432:353–60.
25. Vazquez A, Bond EE, Levine AJ, Bond GL. The genetics of the p53 pathway, apoptosis and cancer therapy. *Nat Rev Drug Discov* 2008;7:979–87.
26. Lacroix M, Toillon RA, Leclercq G. p53 And breast cancer, an update. *Endocr Relat Cancer* 2006;13:293–325.
27. Papazoglu C, Mills AA. p53: At the crossroad between cancer and ageing. *J Pathol* 2007;211:124–33.
28. Gasco M, Shami S, Crook T. The p53 pathway in breast cancer. *Breast Cancer Res* 2002;4:70–6.
29. Yang YP, Kuo HS, Tsai HD, Peng YC, Lin YL. The p53-dependent apoptotic pathway of breast cancer cells (BC-M1) induced by the bis-type bioreductive compound aziridinylnaphthoquinone. *Breast Cancer Res* 2005;7:R19–27.
30. Harris SL, Gil G, Hu W, et al. Single-nucleotide polymorphisms in the p53 pathway. *Cold Spring Harbor Symp Quant Biol* 2005;70:111–9.
31. Gochhait S, Bukhari SI, Bairwa N, et al. Implication of BRCA2 -26G→A 5' untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53 codon 72 Arg→Pro polymorphism. *Breast Cancer Res* 2007;9:R71.
32. Horikawa Y, Wood CG, Yang H, et al. Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. *Clin Cancer Res* 2008;14:7956–62.
33. Hu Z, Chen J, Tian T, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest* 2008;118:2600–8.
34. Shen J, Ambrosone CB, DiCioccio RA, Odunsi K, Lele SB, Zhao H. A functional polymorphism in the *miR-146a* gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis* 2008;29:1963–6.
35. Brendle A, Lei H, Brandt A, et al. Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker. *Carcinogenesis* 2008;29:1394–9.
36. Yin Y, Liu C, Tsai SN, Zhou B, Ngai SM, Zhu G. SET8 recognizes the sequence RHRK20VLRDN within the N terminus of histone H4 and monomethylates lysine 20. *J Biol Chem* 2005;280:30025–31.
37. Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;128:683–92.
38. Fraga MF, Ballestar E, Villar-Garea A, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* 2005;37:391–400.
39. Fang J, Feng Q, Ketel CS, et al. Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr Biol* 2002;12:1086–99.
40. Huen MS, Sy SM, van Deursen JM, Chen J. Direct interaction between SET8 and proliferating cell nuclear antigen couples H4-20 methylation with DNA replication. *J Biol Chem* 2008;283:11073–7.
41. Jorgensen S, Elvers I, Trelle MB, et al. The histone methyltransferase SET8 is required for S-phase progression. *J Cell Biol* 2007;179:1337–45.
42. Yin Y, Yu VC, Zhu G, Chang DC. SET8 plays a role in controlling G1/S transition by blocking lysine acetylation in histone through binding to H4 N-terminal tail. *Cell Cycle* 2008;7:1423–32.
43. Houston SI, McManus KJ, Adams MM, et al. Catalytic function of the PR-Set7 histone H4 lysine 20 monomethyltransferase is essential for mitotic entry and genomic stability. *J Biol Chem* 2008;283:19478–88.
44. Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 2003;22:5323–35.
45. Kleer CG, Cao Q, Varambally S, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003;100:11606–11.
46. 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.
47. Fabbri M, Garzon R, Cimmino A, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc Natl Acad Sci U S A* 2007;104:15805–10.
48. Tuddenham L, Wheeler G, Ntonia-Fousara S, et al. The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells. *FEBS Lett* 2006;580:4214–7.
49. Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 2005;65:7065–70.