Single-Nucleotide Polymorphisms Inside MicroRNA Target Sites Influence Tumor Susceptibility

Milena S. Nicoloso[1], Hao Sun[3,4], Riccardo Spizzo[1], Hyunsoo Kim[4], Priyankara Wickramasinghe[4], Masayoshi Shimizu[1], Sylwia E. Wojcik[5], Jana Ferdin[1,6], Tanja Kunej[6], Lianchun Xiao[2], Siranoush Manoukian[7], Giorgio Secreto[8], Fernando Ravagnani[9], Xuemei Wang[2], Paolo Radice[10,11], Carlo M. Croce[5], Ramana V. Davuluri[4], and George A. Calin[1]

Abstract

Single-nucleotide polymorphisms (SNP) associated with polygenetic disorders, such as breast cancer (BC), can create, destroy, or modify microRNA (miRNA) binding sites; however, the extent to which SNPs interfere with miRNA gene regulation and affect cancer susceptibility remains largely unknown. We hypothesize that disruption of miRNA target binding by SNPs is a widespread mechanism relevant to cancer susceptibility. To test this, we analyzed SNPs known to be associated with BC risk, *in silico* and *in vitro*, for their ability to modify miRNA binding sites and miRNA gene regulation and referred to these as target SNPs. We identified rs1982073-TGFB1 and rs1799782-XRCC1 as target SNPs, whose alleles could modulate gene expression by differential interaction with *miR-187* and *miR-138*, respectively. Genome-wide bioinformatics analysis predicted ~64% of transcribed SNPs as target SNPs that can modify (increase/decrease) the binding energy of putative miRNA::mRNA duplexes by >90%. To assess whether target SNPs are implicated in BC susceptibility, we conducted a case-control population study and observed that germline occurrence of rs799917-BRCA1 and rs334348-TGFR1 significantly varies among populations with different risks of developing BC. Luciferase activity of target SNPs, allelic variants, and protein levels in cancer cell lines with different genotypes showed differential regulation of target genes following overexpression of the two interacting miRNAs (*miR-638* and *miR-628-5p*). Therefore, we propose that transcribed target SNPs alter miRNA gene regulation and, consequently, protein expression, contributing to the likelihood of cancer susceptibility, by a novel mechanism of subtle gene regulation. *Cancer Res; 70(7); 2789–98. ©2010 AACR.*

Authors' Affiliations: [1]Department of Experimental Therapeutics and the Center for RNA Interference and Non-coding RNAs and [2]Division of Quantitative Science, The University of Texas M.D. Anderson Cancer Center, Houston, Texas; [3]Li Ka Shing Institute of Health Sciences, Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China; [4]Center for Systems and Computational Biology, Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, Pennsylvania; [5]Comprehensive Cancer Center, Ohio State University, Columbus, Ohio; [6]University of Ljubljana, Biotechnical Faculty Department of Animal Science, Domzale, Slovenia; and [7]Unit of Medical Genetics and [8]Hormone Research Laboratory, Department of Preventive and Predictive Medicine, [9]Immunohematology and Transfusion Medicine Service, Department of Diagnostic Pathology and Laboratories, [10]Unit of Genetic Susceptibility to Cancer, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale Tumori; [11]IFOM, FIRC Institute of Molecular Oncology Foundation, Milan, Italy

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

M.S. Nicoloso and H. Sun contributed equally to this work.

Corresponding Author: George A. Calin, Department of Experimental Therapeutics, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030. Phone: 713-792-5461; Fax: 713-745-4528; E-mail: gcalin@mdanderson.org or Ramana V. Davuluri, Center for Systems and Computational Biology, Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, PA 19104. Phone: 215-495-6903; E-mail: rdavuluri@wistar.org.

doi: 10.1158/0008-5472.CAN-09-3541

©2010 American Association for Cancer Research.

Introduction

MicroRNAs (miRNA) are a family of endogenous, short noncoding RNAs that modulate posttranscriptional gene regulation. They exert their regulatory role on protein-coding gene (PCG) expression by binding to either full or partial complementary sequences primarily in the 3′ untranslated region (UTR), but also inside the coding sequence (CDS) and the 5′ UTR, of the corresponding mRNAs. Eventually, this affects mRNA stability and translation (1–3). The role of miRNAs in human cancer pathogenesis has been well established by the identification of genetic alterations in miRNA loci, miRNA expression signatures that define different neoplastic phenotypes, and numerous oncogenes and tumor suppressor genes as miRNA targets (4).

The discovery of germline variants that influence cancer susceptibility is currently the subject of intense research (5). The most common variations are single-nucleotide polymorphisms (SNP), which occur approximately once every 100 to 300 bp in the human genome and constitute an immense source of information for unraveling the heterogeneity of human cancer pathogenesis, clinical course, and response to treatment. Several research groups have studied cancer risk association with SNPs located in candidate genes, and it was estimated that around 50,000 to 250,000 SNPs, mostly located

in or around the 25,000 PCGs, have a putative biological effect (6). SNPs can affect protein function by changing the amino acid sequences (nonsynonymous SNP) or by perturbing their regulation [e.g., affecting promoter activity (7), splicing process (8), and DNA and pre-mRNA conformation]. When SNPs occur in 3′ UTRs, they may interfere with mRNA stability and translation by altering polyadenylation, protein::mRNA, and miRNA::mRNA regulatory interactions. Recent studies have found 3′ UTR SNPs that affect PCG expression via miRNA gene regulation in different diseases (9–14). These experimental evidences led us to hypothesize that the disruption of miRNA target binding by SNPs located in the 5′ UTRs, CDS, or 3′ UTRs is a widespread mechanism leading to cancer susceptibility and initiation. In support of our hypothesis, it was recently shown by Chin and colleagues (15) that a kRAS variant within the *let-7* target site increased the risk for non–small cell lung carcinoma among moderate smokers. Moreover, because the 3′ UTRs of PCGs were insufficiently screened for mutations/polymorphisms, it is possible that the extent of such abnormalities might be much greater than initially predicted.

Breast cancer (BC) is one of the most common female malignancies, with more than 1 million new cases diagnosed every year (16). It is widely accepted that the majority of BC risk might be specified by the combined product of multiple low-penetrance alleles (see ref. 17 for review); however, the multiplicity of variants identified thus far alone cannot account for ~80% of familial BC cases that are unrelated to high-penetrance BC susceptibility genes. The role of polymorphic variants located in BC-relevant genes in tumor susceptibility has been extensively addressed (Supplementary Table S1; refs. 18, 19). However, none of the reports investigating the pathogenetic implications of genomic variation on BC susceptibility has evaluated the role of SNPs in miRNA::mRNA gene regulation and its effect in BC development.

Here, we analyzed SNPs associated with BC susceptibility for their ability to affect miRNA binding sites and miRNA::mRNA gene regulation. We identified SNP-dependent miRNA interactions that might explain the pathogenetic relevance of known BC-associated SNPs. We performed a genome-wide scan for SNPs inside PCG transcribed regions (3′ UTR, CDS, and 5′ UTR) that could potentially alter miRNA binding (referred to as target SNPs). Finally, we illustrated the functional consequences of target SNPs on miRNA regulation of protein expression and their possible clinical implications due to differences in frequency distribution among familial BCs, sporadic BCs, and controls.

Materials and Methods

Data sources. We obtained mature human miRNA sequences (total of 885) from National Center for Biotechnology Information (NCBI) build 36 database (version 13.0). We downloaded SNP data from HapMap (Release 21a). We retrieved 5′ UTR, 3′ UTR, and CDS mRNA genomic locations from the University of California at Santa Cruz

(20) known gene table (human genome assembly, NCBI build 36, hg18).

miRNA::SNP interaction analyses. For each SNP, we retrieved two sequences, centered on each allele with 25 nucleotides flanking both sides. We used the miRNA target prediction program miRanda (2) with two different cutoffs (score ≥80, MFE ≤−16 kcal/mol and score ≥50, MFE ≤−5 kcal/mol) to calculate minimum free energy (MFE) for all the possible miRNA::SNP centered sequences. For all the predicted interactions, we computed the allele-dependent MFE changes. Based on the distribution of MFE changes, we identified the 20 and 80 percentile threshold values for miRNA::target SNP identification (Fig. 1). Based on these calculations, we termed SNP alleles as either active or nonactive alleles, the active allele being the variant that induces a decrease in MFE and a stronger miRNA binding with the target.

Patient sample collection. Blood samples from patients with familial and sporadic BC and from control subjects were collected at the Istituto Nazionale Tumori (Milan, Italy) with
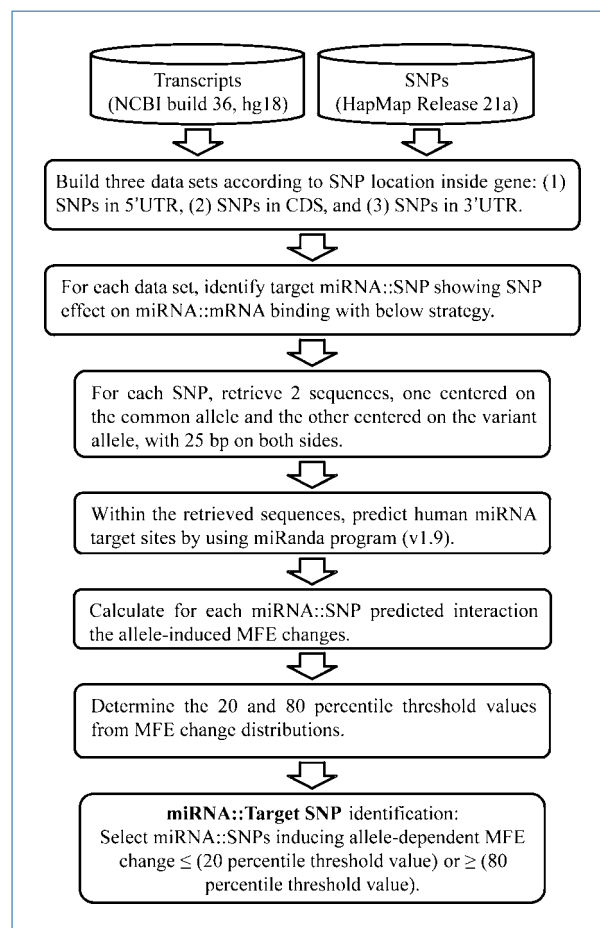

Figure 1. Genome-wide identification of miRNA::target SNP interactions. Schematic overview of the integrated bioinformatics strategy used for miRNA and SNP interaction analyses. The 8% MFE absolute change corresponds to the 20 and 80 percentile threshold values obtained from the MFE change distributions (Supplementary Fig. S2).

their informed consent. Total genomic DNA was purified from peripheral blood leukocytes using the QIAamp DNA Mini kit (Qiagen).

Selection of SNPs associated with BC. We performed a literature search by use of the MedLine/PubMed database (21) covering the time between January 2006 and December 2008 to retrieve articles reporting association studies between SNPs and BC risk. We selected SNPs that were confirmed to be associated with BC risk in one or more articles and/or from large-scale genome-wide association studies. Supplementary Table S1 shows the initial list of SNPs use for subsequent miRanda analysis.

DNA sequencing and SNP cloning. PCR amplification of the transcribed SNP-containing region was performed on genomic DNA (Platinum Taq High Fidelity, Invitrogen). Primers are available on request. Sequences were performed using the BigDye Terminator Reaction Chemistry v3.1 on Applied Biosystems 3730 DNA analyzers (Applied Biosystems). All sequence analyses and alignments were performed with the SeqmanPro program, Lasergene version 7.1 (DNASTAR). PCR-amplified SNP-containing regions carrying either the active or the nonactive alleles were *Xba*I cloned into the 3′ UTR of the pGL3-control vector (Promega).

Cell cultures, transfection, and immunoblotting. All cell lines were obtained from the American Type Culture Collection. Cells were transfected using Lipofectamine 2000 (Invitrogen). Total protein extracts were prepared in 0.5% NP40 lysis buffer. For immunoblotting, protein extracts were separated in 4% to 20% SDS-PAGE (Criterion Precast Gel, Bio-Rad) and transferred to nitrocellulose membranes (Bio-Rad). Antibodies for TGFB, TGFBR1, and XRCC1 were from Cell Signaling; antibodies for BRCA1 were from Calbiochem (clone MS110); and antibodies for vinculin were from Santa Cruz Biotechnology (clone N-19). Bands were quantified using GelDoc XR software (Bio-Rad).

Luciferase assay. MCF7 cells (200×10^5 per 24-well plates) were cotransfected with 0.4 μg of pGL3 (luciferase), 0.08 μg of pRLTK (*Renilla*; Promega), and 50 nmol/L of precursor miRNA molecules or scrambled negative control (Ambion). Thirty hours after transfection, cells were lysed in 100 μL of passive lysis buffer according to the dual luciferase reporter assay protocol (Promega), and luciferase activity was measured with Veritas luminometer (Turner BioSytems).

Statistical analysis. Fisher's exact test was used to determine the association of SNPs and patient disease status. Cochran-Armitage trend test was used to investigate whether there was a trend in binomial proportions of cancer patients (familiar + sporadic, familiar, or sporadic, respectively) across levels of the different genotypes for each SNP. The univariate and multivariate logistic regression models were fit to evaluate the association of SNP genotype with patient state status.

Additional methods, including patient information, statistical analysis, RNA extraction, retrotranscription, and real-time PCR, are available in Supplementary Materials and Methods.

Results

Analysis of BC susceptibility alleles for their ability to affect miRNA binding. To test our hypothesis that SNPs can affect BC susceptibility by altering miRNA::mRNA binding, we took advantage of genetic variants that are already known to be associated with BC (see Materials and Methods and Supplementary Table S1). For each of these SNPs, we retrieved two sequences 51 nucleotides long, one centered on the common allele and the other centered on the variant allele. We scanned the retrieved sequences with miRanda (2) in search of miRNA::mRNA target sites and calculated the MFE changes induced by the allelic variants (Supplementary Table S2). We observed that the allelic variants can either increase or decrease the MFE of the corresponding RNA duplexes, leading to either a stronger or a weaker miRNA::mRNA binding, respectively. This mechanism can also lead to either creation of a new binding site or destruction of an existing target site. We focused only on those miRNAs (included in Supplementary Table S2) that were expressed in BC samples and cell lines (Supplementary Fig. S1; data not shown) and selected three candidate target SNPs for functional validation, for a total of 16 miRNA::SNP interactions (Supplementary Table S3). Two of the 16 interactions chosen were negative controls, in which the interacting miRNA binding was not modified by the SNP allelic variants. To verify whether the selected BC-associated SNPs can actually affect miRNA binding, we performed *in vitro* luciferase reporter assays. For each SNP, we produced two pGL3-SNP constructs (Fig. 2A) carrying either the active or the nonactive allele and cotransfected them in parallel with the predicted interacting miRNAs or the scrambled negative control. rs1982073 inside TGFB1 and rs1799782 inside XRCC1 (Table 1) showed statistically significant effect on *miR-187* and *miR-138* activity, respectively (Table 2; Fig. 2B). The rs1982073 [C] active allele of TGFB1 was predicted by miRanda analysis to create a new target site for *miR-187*, which was absent with the more common [T] variant when stringent settings were used. Luciferase assay showed statistically significant suppressive effect of *miR-187* on the construct carrying the [C] active allele, which is completely absent with the [T] nonactive allele (40% decrease in luciferase activity, $P = 0.006$; Fig. 2B, left). This agrees with the major line of evidence assigning an inhibitory role of miRNAs on gene expression. Conversely, according to miRanda prediction, the rs1799782-XRCC1 [T] variant was the active allele increasing the binding energy of *miR-138* compared with the [C] variant. When *miR-138* was assayed with pGL3-rs1799782-XRCC1 constructs, we observed a significant increase in the luciferase activity with the [T] allele compared with the [C] allele (34% increase, $P = 0.0003$; Fig. 2B, right). This clearly indicates a stabilizing role, rather than an inhibitory role, for *miR-138* stronger binding with the rs1799782-XRCC1 [T] variant.

Subsequently, we tested the effect of these two target SNPs on endogenous TGFB1 and XRCC1 protein levels in cancer cell lines carrying different SNP genotypes by transfecting with the interacting miRNAs. *miR-187*, consistent with the
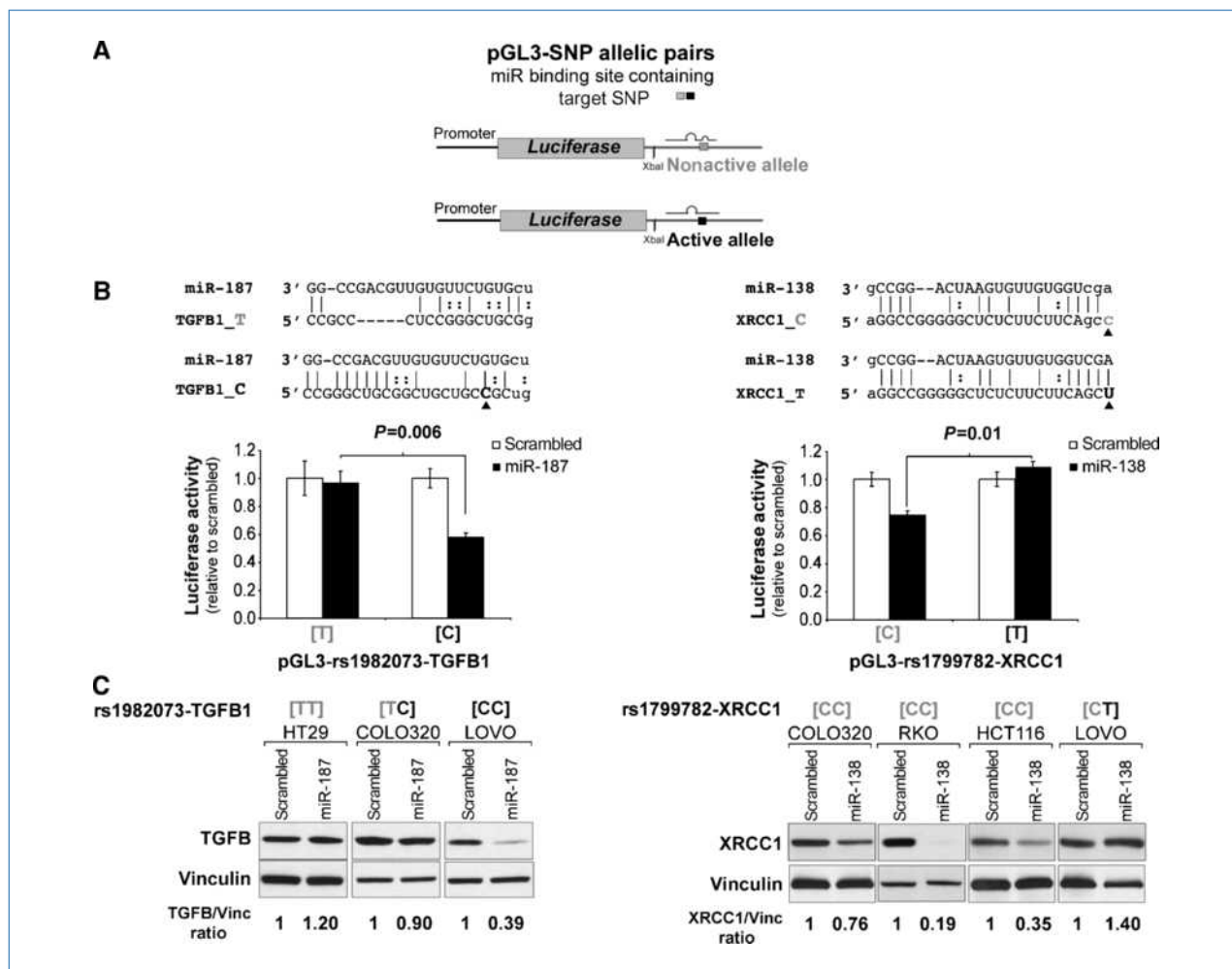
Figure 2. BC-associated target SNPs affect miRNA gene regulation. A, schematic representation of luciferase reporter constructs containing target SNP inside the miRNA binding site and their effect on miRNA interaction. B, top, for the selected SNPs, the miRNA::mRNA interaction is shown, with the active alleles or the nonactive alleles highlighted. Bottom, luciferase activity for pGL3-SNP constructs of BC-associated SNP cotransfected with the predicted interacting miRNA or the scrambled negative control (=1); values represent the average ± SD of three to five independent experiments performed in six replicates. C, Western blot for TGFB1 (left) and XRCC1 (right) after *miR-187* and *miR-138* transfection, respectively, in cell lines carrying different rs1982073-TGFB1 and rs1799782-XRCC1 variants. Genotypes are reported on the top of the immunoblots; below are reported TGFB1 and XRCC1 expression levels normalized for vinculin protein levels and compared with the scrambled negative control transfection (=1) per each cell line.

luciferase results, preferentially downregulated protein levels of the TGFB1 gene carrying the [CC] rs1982073 active alleles, whereas *miR-187* had an intermediate and opposite effect on the [TC] and [TT] rs1982073 genotypes, respectively (Fig. 2C, left). Similarly, *miR-138* showed a stabilizing effect when overexpressed in a cell line that was heterozygous for rs1799782-XRCC1. Whereas [CC] carriers displayed decreased XRCC1 protein levels, a cell line carrying the [CT] genotype displayed increased XRCC1 levels, in agreement with the luciferase results (Fig. 2C, right). Among the 25 different cancer cell lines we screened, we did not find any rs1799782-XRCC1 [TT] homozygous carrier, supporting the notion that [TT] carriers are protected from developing cancer (22). Taken together, these results suggest that for ∼15%

of the investigated transcribed SNPs (2 of 14 target SNP interactions predicted to be allele dependent and tested in our analysis; Fig. 2; Supplementary Table S3; data not shown), their known association with BC can be biologically due to interference with miRNA binding and therefore miRNA gene regulation. The fact that only one seventh of the target SNP interactions could be biologically confirmed is not surprising, given the high rate of false positives of all miRNA target prediction programs. Although miRanda program is known to produce more false-positive predictions than other existing programs that give preference to seed complementarity or interspecies target conservation, it has a higher sensitivity. We, therefore, choose miRanda to identify noncanonical miRNA targets (i.e., with low seed complementarity) such

Table 1. Established common BC susceptibility SNPs proven to affect miRNA binding

SNP	Gene	Location	Chromosome	Population	No. cases/ controls	Reference	Notes	Genotype and BC risk	OR (95% CI; P)
rs1982073 (T+29C)	TGFB1	Exonic (missense; Leu^{10}Pro)	19(q13.2)	European (Netherlands)	143/ 3,646	(28)		CC carriers have increased risk of BC	1.4 (1.1-2.0; P = 0.04)
				European	3,987/ 3,867	(29)	TGFB1 Pro variant is more highly secreted by cells	CC vs T carriers (Pro/Pro vs Leu carriers) have increased BC risk	1.21 (1.05-1.37; P = 0.01)
rs1799782 (580C > T)	XRCC1	Exonic (missense; Arg^{194}Trp)	19(q13.31)	Mixed ethnicity (Meta-analysis)	4,933/ 6,775	(23)	BC together with other cancers	TT + CT vs CC carriers (Trp/ Trp + Arg/ Trp vs Arg/Arg carriers) have lower BC risk	0.89 (0.81-0.98)
				Mixed ethnicity (Nurses' Health Study)	1,004/ 1,385	(35)	Association with BC risk and carotenoid levels	Trp carriers have a marginal lower risk of BC and a definitely lower risk in association with high carotenoid levels	0.79 (0.60-1.04) alone and 0.32 (0.16-0.61; P = 0.003) with high carotenoid levels
				American	502/502	(36)	Postmenopausal women	Trp carriers among postmenopausal women have lower BC risk	0.62 (0.40-0.95; P < 0.05)

as rs1982073-TGFB1 and rs1799782-XRCC1 that would have been otherwise overlooked by using prediction programs focusing only on 3′ UTR and considering extensive 5′ complementarity (data not shown). Overall, from our *in silico* predictions, we validated target SNP interactions that were predicted to have a wide range of MFE change (ranging from 8% to 42%) with no direct correlation between MFE and

probability of a true miRNA::target SNP interaction. Therefore, for subsequent analyses, we filtered the predicted target interactions with a MFE change of at least 8%, at which, in our *in vitro* experiments, a significant biological effect was detected. Moreover, at this cutoff, through a genome-wide target SNP analyses (see below), we noticed a clear peak in the MFE change distribution, further supporting a possible

Table 2. miRanda predictions

SNP	Gene	Interacting miRNA	Alleles	MFE	Active alleles	MFE change
rs1982073	TGFB1	miR-187	T	−649*	C	42%
			C	−21.93		
rs1799782	XRCC1	miR-138	C	−18.79	T	10%
			T	−20.59		

*No specific miRNA::SNP interaction was found with default thresholds (MFE <−16, score >80).

Table 3. List of target SNPs inside cancer-relevant genes analyzed by direct sequencing in BC and control samples together with an example of interacting miRNA

Gene	SNP	Allelic variants	Genomic position	Interacting miRNA	Active allele	MFE	MFE change
BRCA1	rs799917	C/T	Exonic	miR-638[*,†]	C	−27.52	−8%
MDM2	rs769412	A/G	Exonic	miR-296-5p	G	−18.15	66%
TGFBR1	rs334348	A/G	3′ UTR	miR-628-5p[*,†]	G	−20.95	29%
TP53BP2	rs17739	C/T	3′ UTR	miR-129-5p	T	−19.31	121%
TP53INP1	rs7760	G/T	3′ UTR	miR-330-5p	G	−17.92	−92%
XIAP	rs28382751	A/C	3′ UTR	miR-542-5p[*]	C	−21.25	28%

NOTE: For each target SNP, an example of computational prediction of MFE percentage change with a selected miRNA is shown.
[*]SNP::miRNA interactions tested by luciferase assay (Fig. 3A; Supplementary Fig. S3).
[†]SNP::miRNA interactions tested by Western blot (Fig. 3B).

biological significance for MFE changes >8% (Supplementary Fig. S2).

Identification and characterization of target SNPs that disrupt miRNA::mRNA interaction. Of 3,839,363 SNPs comprised in the human HapMap database (Release 21a), we identified 90,985 SNPs that are located in mRNA regions (i.e., transcribed SNPs) and grouped them according to their genomic locations (5′ UTRs, CDS, and 3′ UTRs). We identified candidate target SNPs that can potentially create, destroy, or modify miRNA binding sites due to allele-specific MFE changes through an integrated bioinformatics approach (Fig. 1; Materials and Methods) that includes the predictions of miRanda (2). In Supplementary Tables S4 to S6, we present the top 100 miRNA::SNP allele-specific MFE changes for SNPs located in 3′ UTR, CDS, and 5′ UTR, respectively (additional target SNP interactions are available on request). We observed highly similar bimodal distribution of MFE change for 3′ UTR, 5′ UTR, and CDS (Supplementary Fig. S2) and considered all SNPs that induce at least an 8% MFE change as target SNPs. Using this strategy (Fig. 1) and stringent miRanda settings (see Materials and Methods), we found that 41%, 6%, and 53% of target SNPs were located inside 5′ UTR, CDS, and 3′ UTR, respectively. In the MFE change distribution of all transcribed SNPs, the lower 2.5 and the higher 97.5 percentile corresponded to minimum 94% MFE change. By using 94% cutoff, we found that ~64% of transcribed SNPs are predicted to disrupt (increase/decrease) the MFE of putative miRNA::mRNA duplexes. In summary, through this genome-wide bioinformatics approach, we identified a set of putative target SNPs that modify the MFE of miRNA::mRNA binding and eventually influence the interaction and the regulatory function of miRNAs with their targets.

Additionally, we performed a similar analysis on the 17 SNPs that Saunders and colleagues (23) identified inside experimentally verified miRNA target sites with the corresponding interacting miRNAs (Supplementary Table S7). We observed that 78% of these SNPs were competent to change the miRNA MFE by at least 8%. Therefore, variant alleles inside experimentally verified miRNA target sites can change the status of miRNA::mRNA interactions and affect

Table 4. Target SNP association with disease status

SNP (gene)		(Sporadic + familiar) vs control		Familiar vs control		Sporadic vs control		Familiar vs sporadic	
		OR (95% CI)[*]	P	OR (95% CI)[*]	P	OR (95% CI)[*]	P	OR (95% CI)[*]	P
rs334348	AG vs AA	1.69 (1.01-2.83)	0.048	2.2 (1.29-4.07)	0.005	1.15 (0.63-2.11)	0.65	1.99 (1.15-3.46)	0.01
(*TGFBR1*)				2.67 (1.19-6.03)	0.002	2.17 (1.15-4.07)			0.02
	GG vs AA	0.71 (0.26-1.99)	0.52	0.46 (0.11-1.86)	0.27	0.94 (0.31-2.88)	0.92	0.48 (0.12-1.97)	0.31
rs799917	CT vs CC	1.57 (1.04-2.37)	0.03	1.25 (0.76-2.05)	0.38	1.98 (1.20-3.26)	0.007	0.63 (0.36-1.10)	0.1
(*BRCA1*)		1.86 (1.01-3.42)	0.046						
	TT vs CC	1.95 (1.06-3.6)	0.03	1.26 (0.59-2.71)	0.55	2.81 (1.40-5.64)	0.003	0.45 (0.21-0.97)	0.04

[*]ORs according to univariate and multivariate (adjusted for age) logistic regression model (first and second row, respectively). Only ORs from statistically significant multivariate logistic regressions are shown.

gene expression, consistent with our genome-wide bioinformatics predictions.

Target SNP distribution in BC and control populations. We sequenced a panel of BC patients (166 sporadic and 169 familial BRCA1- and BRCA2-negative probands) and controls (186; see Supplementary Materials and Methods for patient information) for the germline presence of selected target SNPs located inside cancer-relevant genes (Table 3). Supplementary Table S8 summarizes the distribution of target SNP genotypes and their association with BC risk in Caucasian cases (sporadic and familial BC) and controls. The logistic regression analysis indicated that rs799917-BRCA1 [T] carriers were more likely associated with BC [all cases versus controls: univariate analysis odds ratio (OR) for [CT] carriers of 1.57 (P = 0.03), and still significant in multivariate analysis with OR of 1.86 (P = 0.046), and univariate analysis OR for [TT] carriers of 1.95 (P = 0.03)] and particularly with sporadic BC [sporadic cases versus controls: univariate analysis OR for CT carriers of 1.98 (P = 0.007) and OR for [TT] carriers of 2.81 (P = 0.003); Table 4]. Furthermore, the analysis suggested that rs334348-TGFBR1 [AG] carriers were more likely to have BC (all cases versus controls: univariate analysis OR, 1.69; P = 0.048) and particularly familiar BC (familiar versus control: univariate analysis OR, 2.2; P = 0.005; familiar versus sporadic: univariate analysis OR, 1.99; P = 0.01), and this association was also significant in the multivariate analysis (familiar versus control: multivariate analysis OR, 2.67; P = 0.002; familiar versus sporadic: multivariate analysis OR, 2.17; P = 0.02; Table 4). Despite the small size of the case-control study population, this evidence suggests that specific target SNPs have differential distribution among familial BCs, sporadic BCs, and controls.

Validation of target SNP predictions. To validate the computational predictions and the biological relevance of target SNPs, we first carried out *in vitro* luciferase reporter assays. The identified target SNPs that are located in *BRCA1, TGFBR1*, and *XIAP* (Table 3) affected significantly (P < 0.05) the pGL3-SNP luciferase activity by the predicted interacting miRNAs (Fig. 3A; Supplementary Fig. S3). In all three cases tested, miRNAs displayed a higher repressive effect when interacting with the active allele. Next, we tested the effect of the two target SNPs that we found to be associated with BC, rs799917-*BRCA1* and rs334348-*TGFBR1*, on endogenous BRCA1 and TGFBR1 protein levels. We transfected the interacting miRNAs (*miR-638* and *miR-628-5p*) in cancer cell lines carrying different target SNP genotypes. Following *miR-638* overexpression, seven of nine cell lines (in which BRCA1 was clearly detected to perform correct band quantification) showed a reduction in BRCA1 protein levels (Fig. 3B, left; data not shown). By subgrouping cell lines responsive to *miR-638*–induced BRCA1 reduction according to rs799917 genotype, we observed that the presence of the [CC] genotype was responsible for a stronger reduction of BRCA1 protein levels (61% reduction versus 79%, P = 0.047; Fig. 3C, left). Additionally, we excluded the possibility that *miR-638*–dependent BRCA1 reduction was due to a conserved target site inside BRCA1 3′ UTR predicted by TargetScan

via luciferase and BRCA1 overexpression experiments (Supplementary Fig. S4), further confirming the functional significance of the target SNP interaction site. Following similar approach, we evaluated the effect of rs334348 on *miR-628-5p* regulation of *TGFBR1*. By overexpressing *miR-628-5p* in different cell lines, we observed that *miR-628-5p* behaves as a true repressor of *TGFBR1* (Fig. 3B, right; data not shown) in a cell-specific manner and that its repressor activity on TGFBR1 protein levels is dependent on the rs334348 variant inside the *TGFBR1* 3′ UTR miRNA target sequence (80% versus 50% reduction with the [AA] and the [CC] genotypes, respectively, P = 0.006; Fig. 3C, right). The observed significant effects on protein modulation by miRNA in the presence of the active or nonactive alleles (Figs. 2 and 3; Supplementary Fig. S3) support our hypothesis that genetic variants inside miRNA targets can actually disrupt miRNA gene regulation and affect protein expression, eventually leading to tumor susceptibility.

Discussion

In the present study, we identified a new pathogenetic mechanism to explain the association of certain SNPs with BC susceptibility by using the latest knowledge on posttranscriptional gene regulation by miRNAs. Evidence of the effect of SNPs on miRNA binding derived from our integrative analysis delineates a novel role for SNPs in gene expression modulation. Supporting this theory, a recent work suggested that differences in SNP allele frequency among ethnic groups account for differences in gene expression (24). Two of the 11 *cis* SNPs identified by Spielman and colleagues (24) were actually transcribed SNPs located in 3′ UTRs, and by bioinformatics analysis, we found that both of the SNPs were qualified to induce MFE absolute changes >8% with multiple putative interacting miRNAs (Supplementary Table S9). Therefore, we speculate that *cis* SNPs modulate phenotypic gene expression diversities, at least in part, through alteration of miRNA target binding capability, ultimately leading to differences in the susceptibility to complex genetic diseases, such as BC. Notably, among the target SNPs identified as differentially expressed in BC, we found rs334348 located in the 3′ UTR of *TGFBR1*. The association of this SNP with germline allele-specific expression of TGFBR1 was recently found, and it was shown to confer an increased risk of colorectal cancer (25). This phenotype could be explained, for example, by an altered miRNA interaction and therefore strongly supports our hypothesis that SNPs affecting miRNA function contribute to allele-specific protein expression and consequently play a role in tumor susceptibility. A recent study by Kim and Bartel (26) based on transcriptome analyses of heterozygous mouse strains showed that polymorphic miRNA target sites determine differences in gene expression, which further supports our hypothesis.

The focus of this study is to address the effect of transcribed SNPs on miRNA::mRNA interactions and thereby the regulation of gene expression by these SNPs with possible
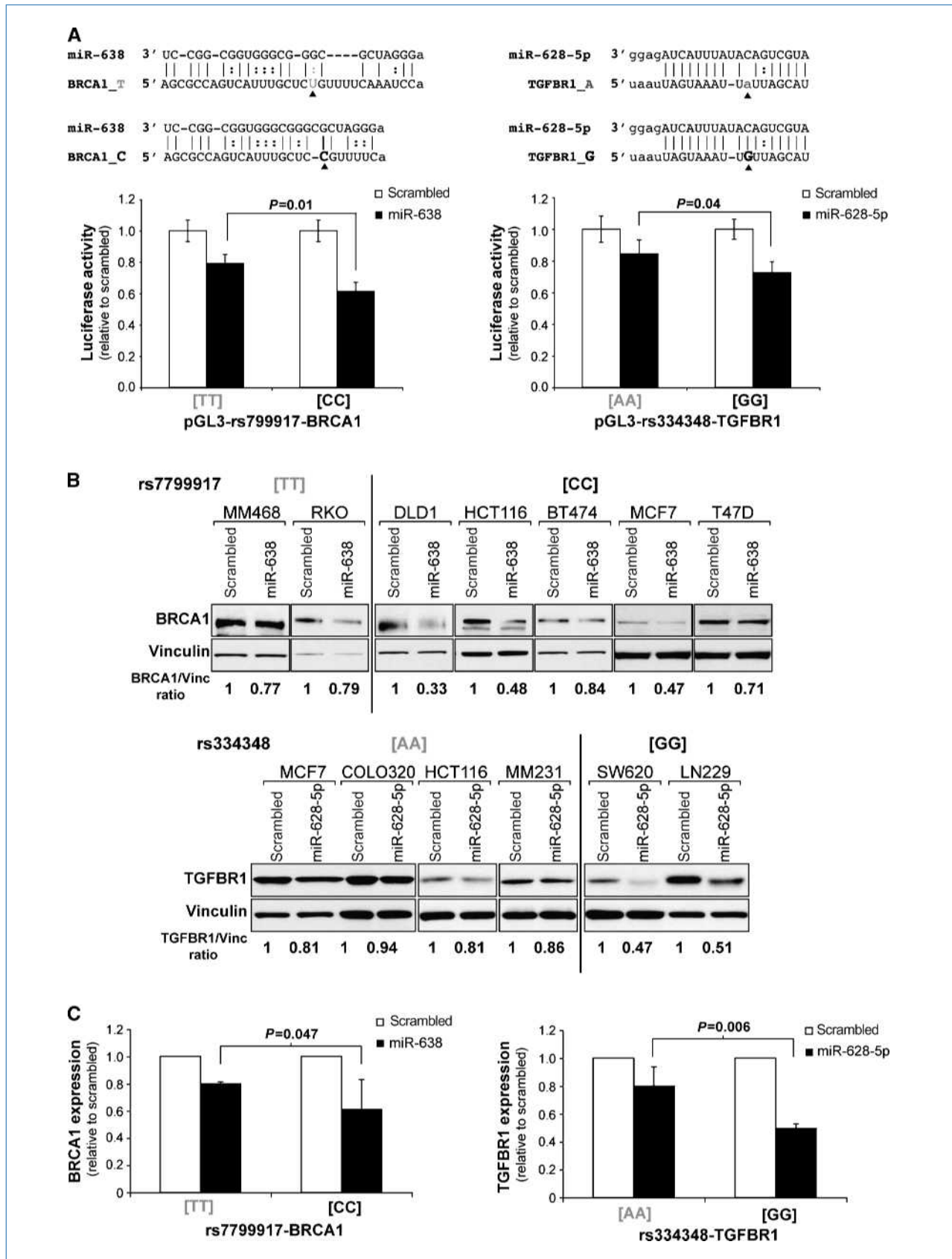
implications in tumorigenesis. We found that numerous target SNPs can effectively interfere with miRNA target recognition, and the distribution of these SNPs significantly varies among BC and control populations, suggesting a potential role in BC susceptibility. It is interesting to note that the BC-associated rs1982073-TGFB1 variant had been previously linked with low expression levels of TGFB1 during the initial phases of tumorigenesis, whereas higher expression levels were found in more advanced metastatic BC stages (27, 28). This switch in TGFB1 expression could be due to concomitant changes of *miR-187* expression levels and could be favored by carrying the [T] variant. Additionally, the described model, together with TGFB1 protumorigenic or antitumorigenic activities in different phases of tumorigenesis, can explain the conflicting results observed for rs1982073-TGFB1 association with BC. Another interesting observation is that the stronger binding occurring between miRNA and active allele does unequivocally mean repression; in fact, as shown for rs1799782-XRCC1, *miR-187* stabilizes XRCC1 expression when carrying the active allele. This finding is in line with other evidence showing that miRNAs can positively modulate the protein expression (29, 30). Furthermore, by searching the complete PCG transcripts (5′ UTR, CDS, and 3′ UTR) for miRNA target SNPs, we have identified *miR-638* as regulator of BRCA1 via binding to its target site inside CDS (to our knowledge, the first miRNA regulating BRCA1 published thus far). Although a conserved interaction site was predicted by TargetScan in the 3′ UTR of BRCA1, we found that it is not a functional site in the conditions we tested. Recently, Tay and colleagues (3) also proved the existence and abundance of miRNA targets located inside the CDS of PCGs, confirming the possibility that *miR-638* regulates BRCA1 in a similar way. Although large case-control studies have not reached a unifying conclusion on the relevance to BC susceptibility of the miRNA target SNP rs7799917 inside *BRCA1* CDS (31, 32), its association with tumor susceptibility has been recently highlighted in a SNP analysis of DNA repair genes in glioblastomas multiforme samples (33). In the present study, we found that the rs7799917 [T] allele is associated with a weaker *miR-638*–dependent BRCA1 reduction and is linked with BC (Table 4). Further investigations are under way to explain this apparent inconsistency. We speculate that this is due to concomitant variations during tumor development of *miR-638* expression levels. Moreover, we observed that low-frequency CDS mutations (both synonymous and nonsynonymous; data not shown)

might exert on miRNA gene regulation effects similar to target SNPs (34). This provides an innovative insight on the importance of synonymous mutations, which in the future should be considered and investigated with renewed attention.

Supported by the data presented here, we highlight the possibility that both genetic variants and a miRNA expression patterns can account for gene expression differences among distinct populations. Therefore, the newly recognized class of target SNPs may contribute to cancer susceptibility not per se (or as tags for specific haplotypes; refs. 31, 32) but in concert with miRNA expression patterns, which are becoming more easily available in this new era of miRNA profiling.

In summary, considering the multifactorial model of BC susceptibility, we propose that transcribed allelic variants can alter miRNA PCG regulation and thus increase miRNA abnormalities. These contribute to tumor susceptibility and tumor general risk by a mechanism of subtle gene regulation still not fully appreciated. As the lists of miRNAs and target SNPs keep growing, the need to confirm the effect of target SNPs on gene expression becomes more pressing and will eventually help us achieve a deeper insight into many variables of the molecular pathways relevant to the development of BC and other human malignancies.

Disclosure of Potential Conflicts of Interest

Acknowledgments

Grant Support

Figure 3. Validation of target SNP predictions. A, top, for the two SNPs, the miRNA::mRNA interaction is shown, with the active alleles or the nonactive alleles highlighted. Bottom, luciferase reporter assay for pGL3-SNP allelic pairs cotransfected with the predicted interacting miRNA, rs799917-BRCA1::*miR-638* (left) and rs334348-TGFBR1::*miR-628-5p* (right). Luciferase activity is expressed relative to scrambled negative control (=1); values represent the average ± SD of at least three independent experiments performed in six replicates. Also shown are miRNA::mRNA interactions according to miRanda predictions, with the active alleles or the nonactive alleles highlighted. B, Western blot for BRCA1 (left) and TGFBR1 (right) after *miR-638* and *miR-628-5p* transfection, respectively, in cell lines carrying different rs799917-BRCA1 or rs334348-TGFBR1 variants. Genotypes are reported on the top of the immunoblots; below are reported BRCA1 and TGFBR1 expression levels normalized for vinculin protein levels and compared with the scrambled negative control transfection (=1) per each cell line. C, average of BRCA1 (left) and TGFBR1 (right) protein reduction in all cells analyzed after miRNA transfection compared with scrambled native control (=1) per each genotype.

References

1. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 2008;9:102–14.

2. Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 2006;126:1203–17.

3. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature 2008;455:1124–8.

4. Spizzo R, Nicoloso MS, Croce CM, Calin GA. SnapShot: microRNAs in cancer. Cell 2009;137:586e1.

5. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004;4:850–60.

6. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56.

7. Nuckel H, Frey UH, Bau M, et al. Association of a novel regulatory polymorphism (−938C>A) in the BCL2 gene promoter with disease progression and survival in chronic lymphocytic leukemia. Blood 2007;109:290–7.

8. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 1992;90:41–54.

9. Abelson JF, Kwan KY, O'Roak BJ, et al. Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 2005; 310:317–20.

10. Sethupathy P, Borel C, Gagnebin M, et al. Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3′ untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. Am J Hum Genet 2007;81:405–13.

11. Mishra PJ, Humeniuk R, Mishra PJ, Longo-Sorbello GS, Banerjee D, Bertino JR. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. Proc Natl Acad Sci U S A 2007;104:13513–8.

12. He H, Jazdzewski K, Li W, et al. The role of microRNA genes in papillary thyroid carcinoma. Proc Natl Acad Sci U S A 2005;102: 19075–80.

13. Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. Proc Natl Acad Sci U S A 2008;105:7269–74.

14. Adams BD, Furneaux H, White BA. The micro-ribonucleic acid (miR-NA) miR-206 targets the human estrogen receptor-α (ERα) and represses ERα messenger RNA and protein expression in breast cancer cell lines. Mol Endocrinol 2007;21:1132–47.

15. 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.

16. Parkin DM. International variation. Oncogene 2004;23:6329–40.

17. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet 2008;40:17–22.

18. Johnson N, Fletcher O, Palles C, et al. Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. Hum Mol Genet 2007;16:1051–7.

19. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007; 447:1087–93.

20. http://genome.ucsc.edu/

21. http://www.ncbi.nlm.nih.gov/pubmed/

22. Hu Z, Ma H, Chen F, Wei Q, Shen H. XRCC1 polymorphisms and cancer risk: a meta-analysis of 38 case-control studies. Cancer Epidemiol Biomarkers Prev 2005;14:1810–8.

23. 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.

24. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. Nat Genet 2007;39:226–31.

25. Valle L, Serena-Acedo T, Liyanarachchi S, et al. Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. Science 2008;321:1361–65.

26. Kim J, Bartel DP. Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. Nat Biotechnol 2009;27:472–7.

27. Gonzalez-Zuloeta Ladd AM, Arias-Vasquez A, Siemes C, et al. Transforming-growth factor β1 Leu10Pro polymorphism and breast cancer morbidity. Eur J Cancer 2007;43:371–4.

28. Dunning AM, Ellis PD, McBride S, et al. A transforming growth factorβ1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. Cancer Res 2003;63:2610–15.

29. Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5′UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell 2008;30:460–71.

30. Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. Science 2007;318: 1931–4.

31. Freedman ML, Penney KL, Stram DO, et al. A haplotype-based case-control study of BRCA1 and sporadic breast cancer risk. Cancer Res 2005;65:7516–22.

32. Cox DG, Kraft P, Hankinson SE, Hunter DJ. Haplotype analysis of common variants in the BRCA1 gene and risk of sporadic breast cancer. Breast Cancer Res 2005;7:R171–5.

33. Chang JS, Yeh RF, Wiencke JK, et al. Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. Cancer Epidemiol Biomarkers Prev 2008;17:1368–73.

34. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108–13.

35. Han J, Hankinson SE, De Vivo I, et al. A prospective study of XRCC1 haplotypes and their interaction with plasma carotenoids on breast cancer risk. Cancer Res 2003;63:8536–41.

36. Patel AV, Calle EE, Pavluck AL, Feigelson HS, Thun MJ, Rodriguez C. A prospective study of XRCC (X-ray cross-complementing group 1) polymorphisms and breast cancer risk. Breast Cancer Res 2005;7: R1168–73.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Single-Nucleotide Polymorphisms Inside MicroRNA Target Sites Influence Tumor Susceptibility

Milena S. Nicoloso, Hao Sun, Riccardo Spizzo, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-09-3541
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2010/03/23/0008-5472.CAN-09-3541.DC1

Cited articles	This article cites 34 articles, 15 of which you can access for free at: http://cancerres.aacrjournals.org/content/70/7/2789.full#ref-list-1
Citing articles	This article has been cited by 13 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/70/7/2789.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.