SINGLE NUCLEOTIDE POLYMORPHISMS INSIDE microRNA TARGET SITES INFLUENCE TUMOR SUSCEPTIBILITY

Milena S. Nicoloso[*,||], **Hao Sun**[$,†,||], **Riccardo Spizzo**[*], **Hyunsoo Kim**[†], **Priyankara Wickramasinghe**[†], **Masayoshi Shimizu**[*], **Sylwia E. Wojcik**[‡], **Jana Ferdin**[*,#], **Tanja Kunej**[#], **Lianchun Xiao**[##], **Siranoush Manoukian**[§], **Giorgio Secreto**[¶], **Fernando Ravagnani**[**], **Xuemei Wang**[##], **Paolo Radice**[††,§§], **Carlo M. Croce**[‡], **Ramana V. Davuluri**[†,‡‡], and **George A. Calin**[*,‡‡]

[*] Department of Experimental Therapeutics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

[$] Li Ka Shing Institute of Health Sciences, Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

[†] Center for Systems and Computational Biology, Molecular and Cellular Oncogenesis Program, The Wistar Institute Philadelphia, PA 19104, USA

[‡] Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210, USA

[#] University of Ljubljana, Biotechnical Faculty Department of Animal Science, 1230 Domzale Slovenia

[##] Division of Quantitative Science, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

[§] Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS, Istituto Nazionale Tumori, 20133, Milan, Italy

[¶] Hormone Research Laboratory, Department of Preventive and Predictive Medicine, Fondazione IRCCS, Istituto Nazionale Tumori, 20133, Milan, Italy

[**] Immunohematology and Transfusion Medicine Service, Department of Diagnostic Pathology and Laboratories, Fondazione IRCCS, Istituto Nazionale Tumori, 20133, Milan, Italy

[††] Unit of Genetic Susceptibility to Cancer, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS, Istituto Nazionale Tumori, 20133, Milan, Italy

[§§] IFOM, FIRC Institute of Molecular Oncology Foundation, 20139, Milan, Italy

Abstract

Single nucleotide polymorphisms (SNPs) 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. In order 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

[‡‡]Corresponding authors: George A. Calin, Tel: 713-792-5461, Fax: 713-745-4528, gcalin@mdanderson.org, Ramana V. Davuluri, Tel: 215-495-6903, rdavuluri@wistar.org.
[||]Milena S. Nicoloso and Hao Sun contributed equally to this work

interaction with *miR-187* and *miR-138*, respectively. Genome-wide bioinformatics analysis predicted approximately 64% of transcribed SNPs as target SNPs that can modify (increase/decrease) the binding energy of putative miRNA::mRNA duplexes by over 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 over-expression 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.

Keywords

Single nucleotide polymorphism; microRNAs; mRNA; BC; tumor susceptibility

INTRODUCTION

MiRNAs are a family of endogenous, short non-coding RNAs that modulate post-transcriptional 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 messenger RNAs (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 (SNPs), which occur approximately once every 100 to 300 base pairs 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 50,000–250,000 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 (non synonymous 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 Slack group that a kRAS variant within the *let-7* target site increased the risk for non-small cell lung carcinoma among moderate smokers (15). Moreover, since 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 one 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 (for review see

(17)); however, the multiplicity of variants identified so 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 ((18,19) and Supplemental Table S1). However, none of the reports investigating the pathogenetic implications of genomic variation on BC susceptibility have 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.

METHODS

Data sources

We obtained mature human miRNA sequences (total of 885) from NCBI build 36 database (version13.0). We downloaded SNP data from HapMap (Release 21A). We retrieved 5′ UTR, 3′ UTR, and CDS mRNA genomic locations from the UCSC (http://genome.ucsc.edu/) known gene table (human genome assembly, NCBI build 36, hg18).

miRNA::SNP interaction analyses

For each SNP we retrieved 2 sequences, centered on each allele with 25 nucleotides flanking both sides. We used the miRNA target prediction program miRanda (2) with two different cut-offs (score $\geq$80, MFE $\leq -$16Kcal/mol and score $\geq$50, MFE $\leq -$5Kcal/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 non-active 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 (INT), Milan Italy with their informed consent. Total genomic DNA was purified from peripheral blood leukocytes using the QIAamp® DNA mini kit (Qiagen).

Selection of SNPs associated with breast cancer

We performed a literature search by use of the MedLine/PubMed database (20) covering the time between January 2006 and December 2008 in order to retrieve papers reporting association studies between SNPs and BC risk. We selected SNPs that were confirmed to be associated with BC risk in one or more papers 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 upon 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 NON-active alleles were XbaI cloned into the 3′ UTR of the pGL3-control vector (Promega).

Cell cultures, Transfection and Immunoblotting

All cell lines were obtained by ATCC. 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; for BRCA1 from Calbiochem (clone MS110) and for Vinculin from Santa Cruz (clone N-19). Bands were quantified using GelDoc XR software (Biorad).

Luciferase assay

MCF7 cells (200×10^5 cells per 24-well) were co-transfected with 0.4ug of pGL3 (luciferase), 0.08ug of pRLTK (renilla) (Promega) and 50nM of precursor miRNA molecules or scrambled negative control (Ambion). Thirty hours after transfection, cells were lysed in 100ul of passive lysis buffer according to the dual luciferase reporter assay protocol (Promega) and luciferase activity 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 tend 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 Realtime PCR are available in Supplemental Materials.

RESULTS

Analysis of BC susceptibility alleles for their ability to affect miRNA binding

In order 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 Methods and Supplemental 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 (Supplemental 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 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 Supplemental Table S2) that were expressed in BC samples and cell lines (Supplemental Fig. S1 and data not shown) and selected 3 candidate target SNPs for functional validation, for a total of 16 miRNA::SNP interactions (Supplemental Table S3). Two out 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. 2b) carrying either the active or non-active allele, and co-transfected 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 (Fig. 2a and Fig. 2c). TGFB1's rs1982073 [C] active allele 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] non-active allele (40% decrease in luciferase activity, p=0.006) (Fig. 2c, left panel). This is in agreement with the major line of evidence assigning an inhibitory role of miRNAs on gene expression. Conversely, according to miRanda prediction, the rs1799782-XRCC1 [T] was the active allele increasing the binding energy of *miR-138* compared to 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 to the [C] allele (34% increase, p=0.0003)(Fig. 2c, right panel). 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 luciferase results, preferentially down-regulated 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. 2d, left panel). Similarly, *miR-138* showed a stabilizing effect when over-expressed in a cell line that was heterozygous for rs1799782-XRCC1. While [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. 2d, right panel). 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 (21). Taken together, these results suggest that for approximately 15% of the investigated transcribed SNPs (2 out of 14 target SNP interactions predicted to be allele-dependent and tested in our analysis, Fig. 2, Supplemental Table S3 and 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. While miRanda program is known to produce more false positive predictions than other existing programs that give preference to seed complementarity or inter-species target conservation, it has a higher sensitivity. We, therefore, choose miRanda to identify non-canonical miRNA targets (i.e. with low seed complementarity) like 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% up 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 cut off, through a genome-wide target SNP analyses (see below), we noticed a clear peak in the MFE change distribution, further supporting a possible biological significance for MFE changes greater than 8% (Supplemental Fig. S2).

Identification and Characterization of Target SNPs that Disrupt miRNA::mRNA Interaction

Out 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 and Methods) that includes the predictions of miRanda (2). In supplemental Tables S4, S5 and 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 upon request). We observed highly similar bi-modal distribution of MFE change for 3′ UTR, 5′ UTR, and CDS (Supplemental 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 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% cut-off, we found that approximately 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 et al. identified inside experimentally verified miRNA target sites with the corresponding interacting miRNAs (22) (Supplemental 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 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 Supplemental Materials for patient information) for the germline presence of selected target SNPs located inside cancer relevant genes (Table 2). Supplemental Table S8 summarizes the distribution of target SNPs 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 vs controls: univariate analysis odds ratio (OR) for [CT] carriers 1.57, p=0.03 and still significant in multivariate analysis with OR =1.86 and p=0.046, and univariate analysis OR for [TT] carriers 1.95, p=0.03) and in particular with sporadic BC (sporadic cases vs controls: univariate analysis OR for CT carriers 1.98, p=0.007 and OR for [TT] carriers 2.81, p=0.003) (Table 3). Furthermore, the analysis suggested that rs334348-TGFBR1 [AG] carriers were more likely to have BC (all cases vs controls: univariate analysis OR 1.69, p=0.048) and in particular familiar BC (familiar vs control: univariate analysis OR 2.2, p=0.005; and familiar vs sporadic: univariate analysis OR 1.99, p=0.01) and this association was also significant in the multivariate analysis (familiar vs control: multivariate analysis OR 2.67, p=0.002; and familiar vs sporadic: multivariate analysis OR 2.17, p=0.02) (Table 3). 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, first we carried out *in vitro* luciferase reporter assays. The identified target SNPs that are located in *BRCA1*, *TGFBR1* and *XIAP* (Table 2) affected significantly (p<0.05) the pGL3-SNP luciferase activity by the predicted interacting miRNAs (Fig. 3a and Supplemental 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* over-expression, 7 out of 9 cell lines (in which BRCA1 was clearly detected in order to perform correct band quantification) showed a reduction in BRCA1 protein levels (Fig. 3b left panel and data not shown). By sub-grouping 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 vs. 79%, p=0.047; Fig. 3c left panel). 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 over-expression experiments (Supplemental 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 over-expressing *miR-628-5p* in different cell lines, we observed that *miR-628-5p* behaves as a true repressor of *TGFBR1* (Fig. 3b right panel, and 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% vs 50% reduction with the [AA] and the [CC] genotypes respectively, p=0.006; Fig. 3c right panel). The observed significant effects on protein modulation by miRNA in the presence of the active or non-active alleles (Fig. 2, Fig. 3 and Supplemental 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 post-transcriptional 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 (23). Two out of the 11 *cis* SNPs identified by Spielman et al. (23) 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 greater than 8% with multiple putative interacting miRNAs (Supplemental 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 (24). This phenotype could be explained, for example, by an altered miRNA interaction, and therefore strongly supports our hypothesis that SNPs affecting miRNAs function contribute to allele specific protein expression and consequently play a role in tumor susceptibility. A recent study by Kim and Bartel (25) 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 vary 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, while higher expression levels were found in more advanced metastatic BC stages (26,27). This switch in TFGB1 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 pro -or anti-tumorigenic 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 (28,29). Furthermore, by searching the complete PCG transcripts (5′ UTR, CDS, 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 so 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 coworkers also proved the existence and abundance of miRNA targets located inside the CDS of PCGs (3), 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 (30,31), its association with tumor susceptibility has been recently highlighted in a SNP analysis of DNA repair genes in glioblastomas multiforme samples (32). In the present study, we found that rs7799917 [T] allele is associated with a weaker *miR-638* dependent BRCA1 reduction, and is linked with BC (Table 3). Further investigations are underway 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 non synonymous, data not shown) might exert on miRNA gene regulation effects similar to target SNPs (33). 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 microRNA 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 (30,31)), 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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. [PubMed: 18197166]

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 22 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. [PubMed: 18806776]

4. Spizzo R, Nicoloso MS, Croce CM, Calin GA. SnapShot: MicroRNAs in Cancer. Cell 2009;137:586e1. [PubMed: 19410551]

5. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004;4:850–60. [PubMed: 15516958]

6. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56. [PubMed: 10866211]

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. [PubMed: 16960146]

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. [PubMed: 1427786]

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. [PubMed: 16224024]

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. [PubMed: 17668390]

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. [PubMed: 17686970]

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. [PubMed: 16365291]

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. [PubMed: 18474871]

14. Adams BD, Furneaux H, White BA. The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. Mol Endocrinol 2007;21:1132–47. [PubMed: 17312270]

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. [PubMed: 18922928]

16. Parkin DM. International variation. Oncogene 2004;23:6329–40. [PubMed: 15322508]

17. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet 2008;40:17–22. [PubMed: 18163131]

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;16:1051–1057. [PubMed: 17341484]

19. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93. [PubMed: 17529967]

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

21. 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. [PubMed: 16030121]

22. 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. [PubMed: 17360642]

23. 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. [PubMed: 17206142]

24. 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. [PubMed: 18703712]

25. Kim J, Bartel DP. Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. Nat Biotechnol 2009;27:472–7. [PubMed: 19396161]

26. Gonzalez-Zuloeta Ladd AM, Arias-Vasquez A, Siemes C, et al. Transforming-growth factor beta1 Leu10Pro polymorphism and breast cancer morbidity. Eur J Cancer 2007;43:371–4. [PubMed: 17035001]

27. Dunning AM, Ellis PD, McBride S, et al. A transforming growth factorbeta1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. Cancer Res 2003;63:2610–15. [PubMed: 12750287]

28. 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. [PubMed: 18498749]

29. Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. Science 2007;318:1931–4. [PubMed: 18048652]

30. 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. [PubMed: 16103107]

31. 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. [PubMed: 15743496]

32. 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. [PubMed: 18559551]

33. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108–13. [PubMed: 17932254]

34. 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. [PubMed: 14679022]

35. Patel AV, Calle EE, Pavluck AL, Feigelson HS, Thun MJ, Rodriguez C. A prospective study of XRCC1 (X-ray cross-complementing group 1) polymorphisms and breast cancer risk. Breast Cancer Res 2005;7:R1168–73. [PubMed: 16457697]
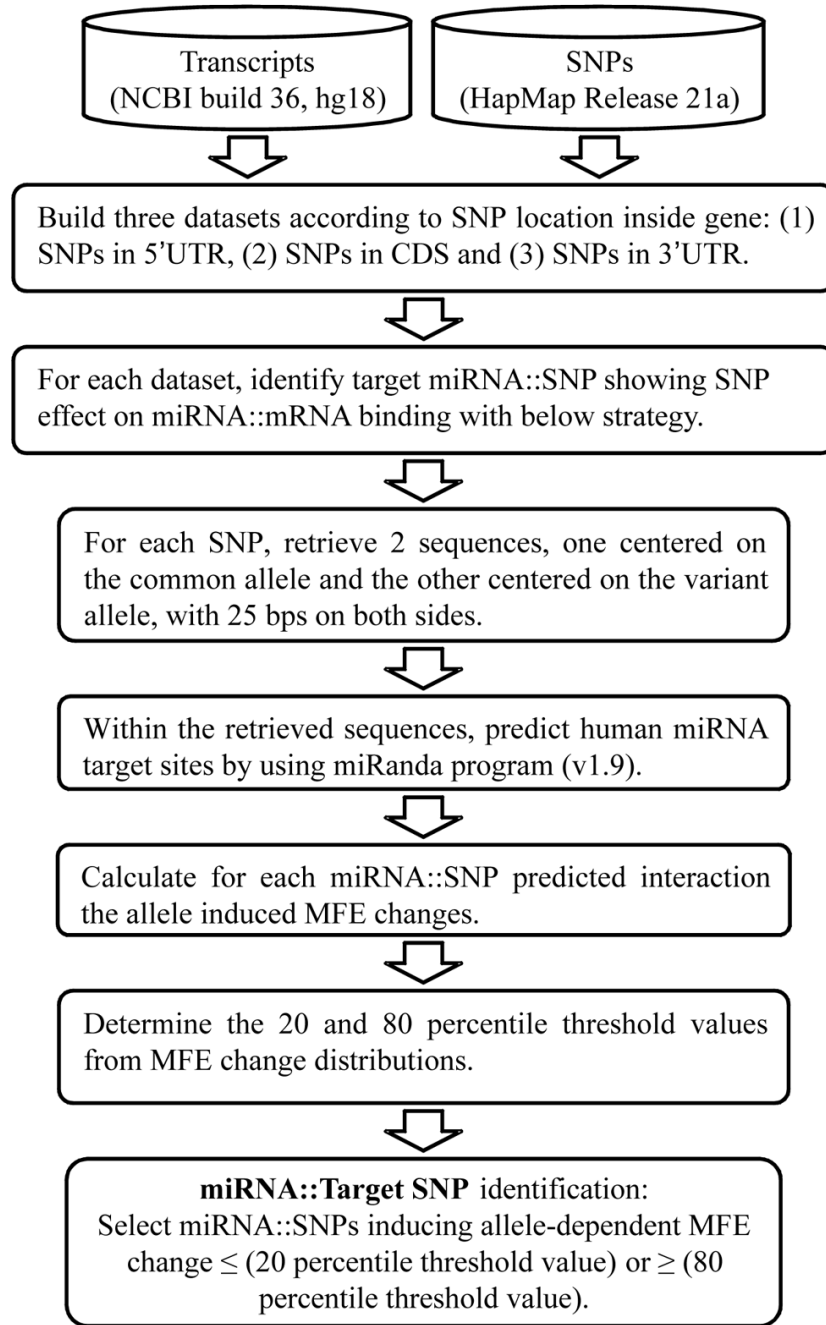
Transcripts (NCBI build 36, hg18)

SNPs (HapMap Release 21a)

Build three datasets according to SNP location inside gene: (1) SNPs in 5'UTR, (2) SNPs in CDS and (3) SNPs in 3'UTR.

For each dataset, identify target miRNA::SNP showing SNP effect on miRNA::mRNA binding with below strategy.

For each SNP, retrieve 2 sequences, one centered on the common allele and the other centered on the variant allele, with 25 bps on both sides.

Within the retrieved sequences, predict human miRNA target sites by using miRanda program (v1.9).

Calculate for each miRNA::SNP predicted interaction the allele induced MFE changes.

Determine the 20 and 80 percentile threshold values from MFE change distributions.

miRNA::Target SNP identification: Select miRNA::SNPs inducing allele-dependent MFE change ≤ (20 percentile threshold value) or ≥ (80 percentile threshold value).
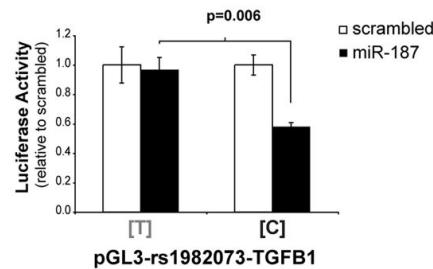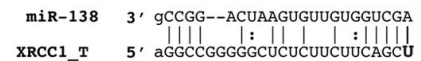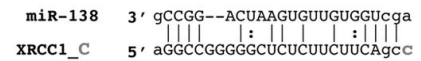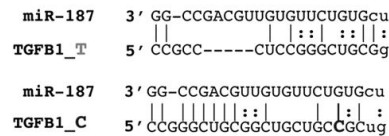
Fig. 1. Genome-wide identification of miRNA::target SNPs interactions
Schematic overview of the integrated bioinformatics strategy used for miRNA and SNP interaction analyses. The 8% MFE absolute change corresponds to the 20 and 80 percentile threshold values obtained form the MFE change distributions (Supplemental Fig. S2).
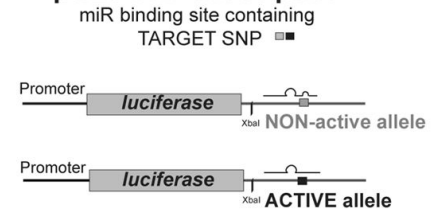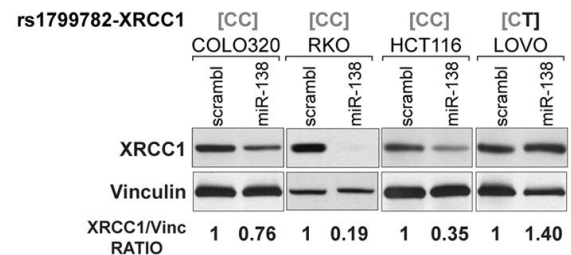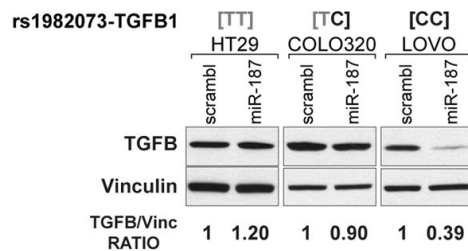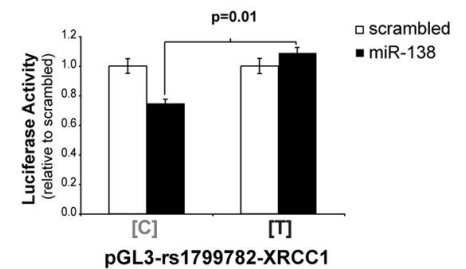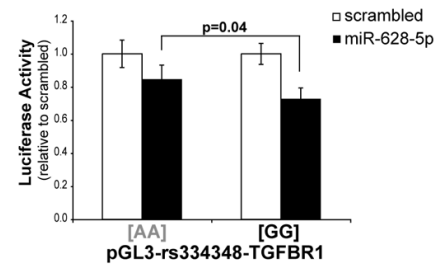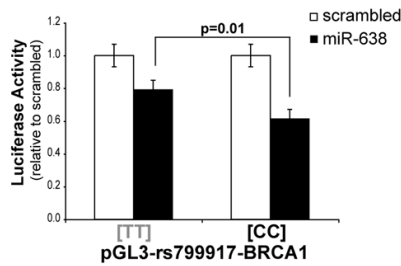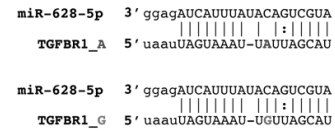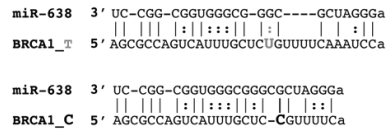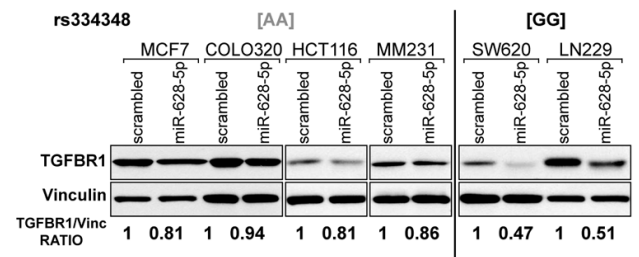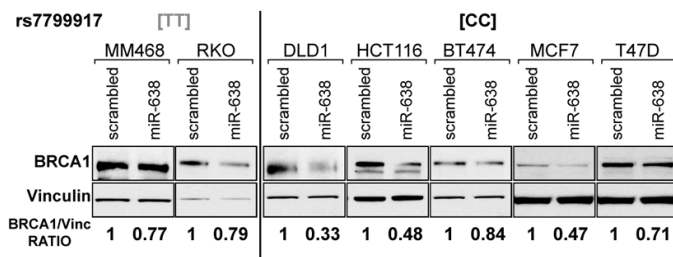
Fig. 2. BC-associated target SNPs affect miRNA gene regulation
(**a**) miRanda predictions. # indicates that no specific miRNA::SNP interaction was found with default thresholds (MFE< −16, score >80). (**b**) Schematic representation of luciferase reporter constructs containing target SNP inside the miRNA binding site and their effect on miRNA interaction. (**c**) Top, for the selected SNPs the miRNA::mRNA interaction is shown with the active alleles (green) or the NON-active alleles (red) highlighted. Bottom, luciferase activity for pGL3-SNP constructs of BC-associated SNP co-transfected with the predicted interacting miRNA or the scrambled negative control (=1); values represent the average +/− standard deviation of 3 to 5 independent experiments performed in six replicates. (**d**) WB 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 to the scrambled negative control transfection (=1) per each cell line.

a

miR-638   3′ UC-CGG-CGGUGGGCG-GGC----GCUAGGGa
BRCA1_T   5′ AGCGCCAGUCAUUUGCUCUGUUUUCAAAUCCa

miR-638   3′ UC-CGG-CGGUGGGCGGGCGCUAGGGa
BRCA1_C   5′ AGCGCCAGUCAUUUGCUC-CGUUUUCa

miR-628-5p   3′ ggagAUCAUUUAUACAGUCGUA
TGFBR1_A   5′ uaauUAGUAAAU-UAUUAGCAU

miR-628-5p   3′ ggagAUCAUUUAUACAGUCGUA
TGFBR1_G   5′ uaauUAGUAAAU-UGUUAGCAU

b

rs7799917 [TT] [CC]

| | MM468 | RKO | DLD1 | HCT116 | BT474 | MCF7 | T47D |
| BRCA1/Vinc RATIO | 1 0.77 | 1 0.79 | 1 0.33 | 1 0.48 | 1 0.84 | 1 0.47 | 1 0.71 |

rs334348 [AA] [GG]

| | MCF7 | COLO320 | HCT116 | MM231 | SW620 | LN229 |
| TGFBR1/Vinc RATIO | 1 0.81 | 1 0.94 | 1 0.81 | 1 0.86 | 1 0.47 | 1 0.51 |

c

Fig. 3. Validation of target SNP predictions
(*a*) Top, for the two SNPs the miRNA::mRNA interaction is shown with the active alleles (green) or the NON-active alleles (red) highlighted. Bottom, luciferase reporter assay for pGL3-SNP allelic pairs co-transfected with the predicted interacting miRNA, rs799917-BRCA1::*miR-638* (left panel) and rs334348-TGFBR1::*miR-628-5p* (right panel). Luciferase activity is expressed relative to scrambled negative control (=1); values represent the average +/− standard deviation of at least 3 independent experiments performed in six replicates. Also shown are miRNA::mRNA interactions according to miRanda predictions, with the active alleles (green) or the NON-active alleles (red) highlighted. (*b*) WB for BRCA1 (left panels) and TGFBR1 (right panels) 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 to the scrambled negative control transfection (=1) per each cell line. (**c**) Average of BRCA1 and TGFBR1 protein reduction
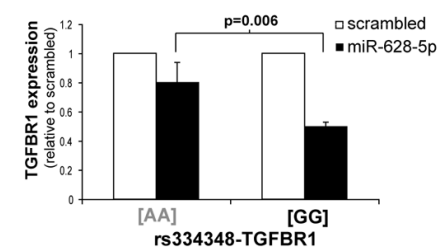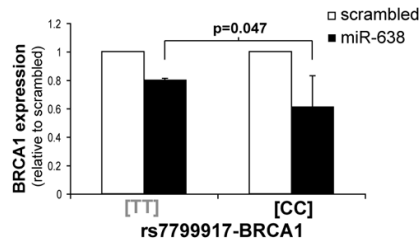
(left and right panel, respectively) in all cells analyzed after miRNA transfection compared to scrambled native control (=1) per each genotype.

Table 1

Established Common Breast-Cancer Susceptibility SNPs proven to affect miRNA binding.

SNP	Gene	Location	Chromosome	Population	No. of cases/controls	Ref	Notes	Genotype and BC risk	Odds Ratio (95% CI; p)
rs1982073 (T+29C)	TGFB1	Exonic (missense; Leu10Pro)	19(q13.2)	European (Netherlands)	143/3,646	[27]		CC-carriers have increased risk of BC	1.4 (1.1–2.0; p=0.04)
				European	3,987/3,867	[28]	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; Arg194Trp)	19(q13.31)	Mixed ethnicity (Meta-analysis)	4,933/6,775	[22]	Breast cancer 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 Breast cancer risk and carotenoids levels	Trp-carriers have a marginal lower risk of BC and a definitely lower risk in association with high carotenoids levels	0.79(0.60–1.04) alone and 0.32 (0.16–0.61; p=0.003) with high carotenoids 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)

Table 2

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

GENE	SNP	Allelic Variants	Genomic Position	INTERACTING miR	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%

For each target SNP, an example of computational prediction of MFE percentage change with a selected miRNA is shown.

[^] SNP::miR interactions tested by luciferase assay (Fig 3a and Supplemental Fig 3), and

[#] SNP::miR interactions tested by Western blot (Fig. 3b).

Table 3

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 value	OR (95% CI)*	p value	OR (95% CI)*	p value	OR (95% CI)*	p value
rs334348 (TGFBR1)	AG vs AA	**1.69** (1.01, 2.83)	**0.048**	**2.2** (1.29, 4.07) **2.67** (1.19, 6.03)	**0.005** **0.002**	1.15 (0.63, 2.11)	0.65	**1.99** (1.15, 3.46) **2.17** (1.15, 4.07)	**0.01** **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 (BRCA1)	CT vs CC	**1.57** (1.04, 2.37) **1.86** (1.01, 3.42)	**0.03** **0.046**	1.25 (0.76, 2.05)	0.38	**1.98** (1.20, 3.26)	**0.007**	0.63 (0.36, 1.10)	0.1
	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**

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