

MicroRNA Polymorphisms and Risk of Colorectal Cancer

Stephanie L. Schmit^{1,2}, Jeremy Gollub³, Michael H. Shapero³, Shu-Chen Huang², Hedy S. Rennert⁴, Andrea Finn³, Gad Rennert^{4,5,6}, and Stephen B. Gruber^{1,2}

Abstract

Background: miRNAs act as post-transcriptional regulators of gene expression. Genetic variation in miRNA-encoding sequences or their corresponding binding sites may affect the fidelity of the miRNA–mRNA interaction and subsequently alter the risk of cancer development.

Methods: This study expanded the search for miRNA-related polymorphisms contributing to the etiology of colorectal cancer across the genome using a novel platform, the Axiom miRNA Target Site Genotyping Array (237,858 markers). After quality control, the study included 596 cases and 429 controls from the Molecular Epidemiology of Colorectal Cancer study, a population-based case–control study of colorectal cancer in northern Israel. The association between each marker and colorectal cancer status was examined assuming a log-additive genetic model using logistic regression adjusted for sex, age, and two principal components.

Results: Twenty-three markers had *P* values less than 5.0E–04, and the most statistically significant association

involved rs2985 (chr6:34845648; intronic of *UHRF1BP1*; OR = 0.66; *P* = 3.7E–05). Furthermore, this study replicated a previously published risk locus, rs1051690, in the 3′-untranslated region of the insulin receptor gene *INSR* (OR = 1.38; *P* = 0.03), with strong evidence of differences in *INSR* gene expression by genotype.

Conclusions: This study is the first to examine associations between genetic variation in miRNA target sites and colorectal cancer using a genome-wide approach. Functional studies to identify allele-specific effects on miRNA binding are needed to confirm the regulatory capacity of genetic variation to influence risk of colorectal cancer.

Impact: This study demonstrates the potential for an miRNA-targeted genome-wide association study to identify candidate susceptibility loci and prioritize them for functional characterization. *Cancer Epidemiol Biomarkers Prev*; 24(1); 65–72. ©2014 AACR.

Introduction

In addition to protein-coding messenger RNAs (mRNAs), other classes of small RNA molecules exist with specialized regulatory and processing functions. Among these types of regulatory RNAs are microRNAs (miRNAs), short (18–24 nucleotide) non-protein-coding molecules that act as post-transcriptional regulators of gene expression (1). The biogenesis of a miRNA begins with transcription from a small, stand-alone gene or an intron or exon of a known protein-coding gene and transitions through a series of conversion steps from hairpin precursors to duplexed pre-miRNA intermediates to mature, single-stranded miRNAs (2, 3). miRNAs

exert their regulatory effects via binding to complementary approximately 6–8 nucleotide target seed sites in the 3′ untranslated regions (3′-UTR) of one or more mRNAs. Depending on the fidelity and context of the interaction, this binding acts to repress translation of the messenger into protein or to signal for degradation of the targeted mRNA (1, 4, 5). Each miRNA typically binds multiple, even thousands, of messenger targets, offering the potential for widespread downstream effects (6, 7).

Deregulated miRNA profiles have been described across a range of cancers including colorectal cancer (8, 9). Furthermore, some suggest that miRNA biology can be integrated into the molecular subtyping of colorectal tumors and into the traditional model of genetic alterations accompanying progression from normal mucosa to carcinoma, particularly among tumors that develop through the chromosomal instability pathway (7, 10–13). As an extension of this work, several miRNAs have been proposed as biomarkers for colorectal cancer early detection, prognosis, and progression (14–16).

Despite extensive miRNA profiling in colorectal tumors, the factors driving aberrations in miRNA expression and their impact on colorectal cancer development and progression are not yet well defined. One hypothesis proposes that SNPs in genes encoding the miRNA sequence or 3′-UTR regions of the corresponding binding sites affect miRNA transcription, miRNA processing, and/or the fidelity of the miRNA–mRNA interaction. Any of these alterations could plausibly impact target mRNA translation into proteins critical for cellular differentiation and proliferation. Evidence from studies of candidate miRNA-related genetic

¹Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California. ²USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California. ³Affymetrix, Incorporated, Santa Clara, California. ⁴Department of Community Medicine and Epidemiology, Carmel Medical Center, Haifa, Israel. ⁵Clalit Health Services National Cancer Control Center, Haifa, Israel. ⁶Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Stephen B. Gruber, USC Norris Comprehensive Cancer Center, 1441 Eastlake Avenue, 8302L Ezralow Tower, Los Angeles, CA 90089. Phone: 323-865-0816; Fax: 323-865-0102; E-mail: sgruber@usc.edu

doi: 10.1158/1055-9965.EPI-14-0219

©2014 American Association for Cancer Research.

alterations supports this hypothesis and suggests that such SNPs may alter expression of some miRNAs in colorectal cancer (17) and increase or decrease the risk of tumor development (18). Target site polymorphisms that confer risk in specific populations have been identified in *INSR* (18), *CD86*(18), *IL16* (19), *RPA2* (20), and *GTF2H1*(20); however, replication of these findings has been limited with the exception of rs1051690 in *INSR* and rs17281995 in *CD86* (21). To date, published studies have not comprehensively investigated polymorphisms implicated in the miRNA regulatory pathway across the genome.

In this study, we expanded the search for miRNA-related genetic variants important in the etiology of colorectal cancer across the genome and investigated the association between thousands of genetic variants in miRNA target sites in 3'-UTR regions and miRNA-encoding genes and colorectal cancer risk using a novel genotyping platform. In contrast to a classical genome-wide association study (GWAS) approach that relies on haplotype-tagging SNPs, we leveraged genotyping of SNPs bioinformatically predicted to have functional implications specific to the miRNA regulatory pathway. We then characterized the predicted miRNA-binding consequences of our most significantly associated SNPs and further explored these associations with expression quantitative trait loci (eQTL) analyses. This study was designed to evaluate the feasibility of a targeted GWAS approach for identifying lead candidates and prioritizing them for functional characterization based on biologically relevant hypotheses. The genetically homogeneous founder population of Ashkenazi Jewish individuals experiences a high burden of colorectal cancer and served as the focus of this study (22).

Materials and Methods

Study population: Molecular Epidemiology of Colorectal Cancer study

The Molecular Epidemiology of Colorectal Cancer (MECC) study is a population-based, case-control study of pathologically confirmed, incident cases of colorectal cancer recruited from a geographically defined region of northern Israel (23). Subject recruitment began in 1998 and remains ongoing. Individually matched controls with no prior history of colorectal cancer were selected from the same source population that gave rise to cases based on the Clalit Health Services database. Matching factors include age, sex, Jewish ethnicity (Jewish versus non-Jewish), and primary clinic site. Subjects are interviewed to obtain demographic data, clinical information, family history, and dietary habits. Biospecimens including blood, paraffin blocks, and snap frozen tumors are collected. On the basis of resource limitations, 1,266 cases and controls (approximately 15% of all MECC participants) were initially selected for genotyping. Following sample quality control (see Supplementary Fig. S1 for details), genome-wide analysis was performed on 596 cases and 429 controls enriched for Ashkenazi Jewish ancestry (Table 1). Informed consent was obtained according to Institutional Review Board-approved protocols at Carmel Medical Center (Haifa, Israel) and the University of Southern California (Los Angeles, CA).

Genotyping and quality control

Germline DNA was extracted from peripheral blood samples, purified, quantified by nanodrop and Qubit fluorometric quantitation, and genotyped by Affymetrix on a novel Axiom miRNA Target Site Genotyping array with 237,858 SNPs and indels

Table 1. Demographic and clinical characteristics of MECC participants ($n = 1,025$) genotyped on the Axiom miRNA Target Site Genotyping Array

	Cases ($n = 596$)	Controls ($n = 429$)
Age, mean (SD), y	70.9 (10.7)	74.3 (10.6)
Sex (%)		
Male	291 (48.8)	221 (51.5)
Female	305 (51.2)	208 (48.5)
Self-reported race/ethnicity (%)		
Ashkenazi	595 (99.8)	413 (96.3)
Ashkenazi/Sephardi	1 (0.2)	4 (0.9)
Sephardi	0	7 (1.6)
Ashkenazi/non-Jewish	0	2 (0.5)
Missing	0	3 (0.7)

(Supplementary Table S1). Markers were selected for the microarray from four online bioinformatic databases: PolymiRTS (86,340), dPORE (10,400), Patrocles (1,200), and microRNA.org (158,400). These databases were leveraged to select polymorphic loci that overlap genes encoding miRNAs, miRNA gene regulatory regions, proteins important for miRNA processing, and/or target seed sites (24–28). For microRNA.org, Affymetrix used the database's high-quality predictions of miRNA-binding sites (both conserved and non-conserved) and intersected microRNA.org's predicted sites with the 1000 Genomes Phase 1 (March 2012) release to identify markers. In addition, the array included a panel of $n = 4,470$ ancestry informative markers (AIM) and loci with known complex trait associations from the August 16, 2011 National Human Genome Research Institute (NHGRI) GWAS Catalog (29). In an ancillary study, MECC samples were also genotyped on a custom Affymetrix Axiom platform with approximately 1.3 million SNPs and indels as part of the ColoRectal Transdisciplinary (CORECT) Study, and concordance was compared across the genotyping platforms. IMPUTE2 v2.2.2 was used to impute missing genotypes based on reference haplotypes from phase I of the 1000 Genomes Project (March 2012 release; $n = 1,092$; refs. 30, 31).

MECC genotype data was filtered on the basis of quality control metrics at the individual subject and SNP levels (Supplementary Fig. S1). Samples with >5% missing genotypes, sex mismatches (between self-reported and genotypic predicted sex), and duplicate samples were identified and subsequently removed. Monomorphic markers and markers with <95% call rate were excluded. Furthermore, SNPs that were not consistent with Hardy-Weinberg Equilibrium in controls were excluded. Principal components analysis (PCA) was conducted using a panel of AIMs and the *pcaMethods* Bioconductor package (32) in R to identify ethnic outliers for removal and to later adjust for population stratification. Principal component definitions of ancestry were used to exclude ethnic outliers and non-Ashkenazi Jewish individuals from the analysis. Thus, all individuals in the final analysis dataset (described in Table 1) were genetically of Ashkenazi Jewish descent, regardless of their self-reported ethnicity.

Gene expression quantification

Gene expression levels from 419,473 probe sets derived from two Affymetrix expression arrays were quantified on RNA isolated from snap frozen tumors of 331 MECC colorectal cancer cases. Of these 331 cases, 135 also had high-throughput germline genotype data available (63 on the Affymetrix Axiom CORECT custom array

and 72 on the Illumina HumanOmni 2.5S-v1 BeadChip). Methods for gene expression quantification via hybridization to GeneChip Human Genome U133A 2.0 and Human Genome U133 Plus 2.0 Arrays have been described elsewhere (33). Briefly, expression was measured in two batches (one for each array) followed by quantile normalization and \log_2 transformation of MAS 5.0-calculated signal intensities. Data from the two batches were aligned after individual batch pre-processing and quality control. These microarray data have been deposited in the Gene Expression Omnibus database (accession number GSE26682) to comply with Minimum Information About a Microarray Gene Experiment guidelines.

Statistical analysis

Logistic regression was employed to examine the marginal association between each marker on the miRNA target site array with minor allele frequency (MAF) $\geq 1\%$ ($n_{\text{marker}} = 55,208$) and colorectal cancer risk assuming a log-additive genetic model. Here, each additional copy of the minor allele was assumed to confer the same magnitude of risk or protection. Each model was run both unadjusted and adjusted for sex, age, and the first two principal components. We calculated beta coefficients, SEs, ORs with associated 95% confidence intervals (95% CI), and P values from unconditional logistic regression. The Bonferroni-corrected alpha level was set at 9.0×10^{-7} (0.05/55,406 SNPs). After taking this genome-wide approach, we then examined previously published SNPs from three studies in the candidate miRNA-related polymorphism literature to assess our ability to replicate purported risk loci (18–20).

To begin the bioinformatic characterization of functional consequences of our most significantly associated SNPs, we investigated predicted changes in miRNA binding using a combination of databases: microrna.org, miRBase, PolymiRTS, and dPORE (24, 25, 28, 34–37). MicroSNiPer was also used to identify the potential disruption or creation of miRNA-binding sites for the following 3'-UTR SNPs in Table 2: rs3180466, rs1972820, and rs2985 (38). A seed site of a minimum of either 7 or 8 bases was specified for each of these SNPs. In addition, we conducted ANOVA to compare differences in gene expression by genotype for all SNPs with association P values less than 5×10^{-4} as well as for a previously published risk locus, where expression and genotype data permitted. Expression of the gene nearest to each SNP was considered.

Results

Targeted GWAS

Plots of the first 3 eigenvalues from PCA indicated that the original samples selected for analysis included some non-Ashkenazi Jewish individuals (almost exclusively among subjects without colorectal cancer) that inhibited our ability to control for confounding due to population stratification through principal component adjustment (data not shown). However, following the removal of 5 ethnic outliers and 211 non-Ashkenazis identified on the basis of genetic definitions, the first two principal components were sufficient to control for the remaining population stratification, as indicated by genomic control lambda [genomic control (GC) λ] values shown below. The distributions of demographic and clinical characteristics of the final analysis dataset were comparable across case and control groups (Table 1).

Quantile–quantile (Q–Q) and Manhattan plots visually display $-\log_{10}(P$ values) resulting from the logistic regression models

adjusted for age, sex, and two principal components (Fig. 1). The Q–Q plot in the left panel plots the rank-ordered observed $-\log_{10}(P$ value) against the rank-ordered expected $-\log_{10}(P$ value). It demonstrates that, on average, we did not observe SNPs with associations more statistically significant than expected under a uniform distribution of P values. The GC λ value of 1 suggests that principal components 1 and 2 were sufficient to control for population stratification in our ethnically homogenous study sample. The Manhattan plot displays the summary results by ordered chromosomal position and shows that our smallest P values are in the 10^{-5} range with none reaching genome-wide statistical significance after correction for multiple testing.

Although none of the individual SNPs achieved genome-wide significance, our top findings are detailed in Table 2. Interestingly, 7 of our 9 most statistically significant SNPs yield a predicted change in miRNA binding in an allele-specific manner. Each of these seven most statistically significant variants predict either a change from no miRNA binding to one or more miRNAs binding or from one set of miRNAs to a different set. None of the most significant miRNA-related SNPs has been previously reported as significantly associated with risk of colorectal cancer.

Gene expression analysis for top association findings

Of the top 25 SNPs that met our P value threshold of 5×10^{-4} from the association analysis, 11 corresponding nearest genes had at least 1 matching probe in our gene expression dataset. Among the 21 total probes quantifying gene expression from these 11 genes (some genes had multiple probes), 13 probes for 6 genes had a corresponding genotype measured in MECC cases from the custom Affymetrix CORECT Axiom and/or Illumina Omni platforms. ANOVA results for gene expression [$\log_2(\text{normalized intensity})$] by genotype for the 6 represented nearest genes with appropriate data availability revealed only one statistically significant SNP, intergenic rs6827968 that falls downstream of the RAP guanine nucleotide exchange factor 2 (*RAPGEF2*) gene ($F = 5.71$; $P = 0.02$). *RAPGEF2* expression levels for two probes plotted against the number of copies of the minor allele at this SNP locus in our study sample can be visualized in Fig. 2, which provides evidence of an eQTL (probe 215992_s_at: $F = 3.6$, $P = 0.06$; probe 203096_s_at: $F = 5.7$, $P = 0.02$).

Replication of previously published risk SNPs

We also examined the colorectal cancer association with 19 candidate miRNA SNPs previously presented in the literature [8 from Landi and colleagues (18), 5 from Azimzadeh and colleagues (19), and 6 from Naccarati and colleagues (20)], of which 6 were statistically significant in the original report. In our dataset, we replicated only one of the previously reported findings (Table 3). The replicated variant (rs1051690), originally reported in Landi and colleagues (18), is located in the 3'-UTR region of the insulin receptor gene *INSR* (Table 3; OR = 1.38; $P = 0.03$) and has predicted miRNA-binding consequences. Our eQTL analysis demonstrated a statistically significant association between the rs1051690 variant and expression of the *INSR* gene, with increasing expression tracking with each additional copy of the minor allele (Supplementary Fig. S2; $F = 21.3$; $P = 8.98 \times 10^{-6}$).

Genotype concordance: miRNA targeted array versus custom GWAS array

Only 14,436 markers were directly measured on both the Affymetrix miRNA Target Site Genotyping Array and the CORECT

Table 2. Summary of variants from the Axiom miRNA Target Site Genotyping Array with the most statistically significant associations from age-, sex-, and principal component 1–2-adjusted logistic regression models

rsID	CHR	BP	A1 ^a	MAF	OR	SE ^b	P	Gene	A1 miRNA	A2 miRNA
rs2985	6	34845648	C	0.30	0.66	0.10	3.7E–05	<i>UHRF1BP1</i> (intron)	miR4529-5p	miR885-5p
rs1139139	10	5020625	T	0.27	1.54	0.11	6.0E–05	<i>AKR1C1</i> (downstream)	NA	miR451b miR556-5p
rs6827968	4	161399652	A	0.34	1.48	0.10	1.0E–04	<i>RAPGEF2</i> (downstream) <i>FSTL5</i> (downstream)	AIM	AIM
rs12130051	1	19545053	T	0.06	2.27	0.21	1.2E–04	<i>KIAA0090</i> (3'-UTR)	NA	miR222 miR1244 miR3129
rs80350662	1	178819064	A	0.04	2.63	0.26	2.1E–04	<i>RALGPS2</i> (intron)	miR32-3p miR4775	miR1277-5p miR889
rs1834481	11	112023827	G	0.12	0.60	0.14	2.4E–04	<i>IL18</i> (intron)	NA	miR637 miR5009-5p miR541-3p miR3978
rs1044724	6	125412231	C	0.08	0.54	0.17	2.5E–04	<i>RNF217</i> (downstream) <i>TPD52L1</i> (upstream)	miR3978	
rs4766991	12	113137384	T	0.16	1.58	0.13	2.7E–04	<i>PTPN11</i> (downstream) <i>RPH3A</i> (upstream)	In promoter for miR1302-1	In promoter for miR1302-1
rs7746892	6	125408263	G	0.08	0.54	0.17	2.9E–04	<i>RNF217</i> (downstream) <i>TPD52L1</i> (upstream)	miR545	miR1252 miR4476 miR4533 miR873
rs7746860	6	125408221	G	0.08	0.54	0.17	2.9E–04	<i>RNF217</i> (downstream) <i>TPD52L1</i> (upstream)	miR2681	miR1295 miR4747-3p
rs2489495	10	38502333	T	0.20	0.66	0.12	3.6E–04	<i>LOC100129055</i> (exon)	miR635	NA
rs853158	5	142605172	C	0.36	0.71	0.10	3.7E–04	<i>ARHGAP26</i> (3'-UTR)	miR3926	miR4480
rs9374072	6	109591586	G	0.31	0.71	0.10	3.7E–04	<i>CEP57L1</i> (downstream) <i>CCDC162</i> (upstream)	miR605	NA
rs471429	6	125409031	G	0.09	0.57	0.16	3.9E–04	<i>RNF217</i> (downstream) <i>TPD52L1</i> (upstream)	miR3126-5p miR3174 miR3591-5p miR3606 miR4419a miR4510 miR921	miR4270 miR4441
rs12268559	10	32856746	C	0.07	0.54	0.18	4.3E–04	<i>CCDC7</i> (coding)	miR578	NA
rs142004998	23	119760042	C	0.08	0.51	0.19	4.4E–04	<i>C1GALT1C1</i> (3'-UTR)	miR1284 miR337-3p miR520d-5p	NA NA
rs3180466	2	129023866	G	0.03	0.38	0.28	4.5E–04	<i>HS6ST1</i> (3'-UTR)	miR4758-5p miR574-5p miR615-3p miR1238-5p miR4745-5p miR3677-5p	miR4669 miR3659
rs79029362	1	178516408	G	0.03	2.91	0.30	4.6E–04	<i>C1orf220</i> (exon)	miR455-3p	NA
rs4766992	12	113137531	A	0.26	1.45	0.11	4.7E–04	<i>PTPN11</i> (downstream) <i>RPH3A</i> (upstream)	Upstream of miR1302-encoding gene	Upstream of miR1302-encoding gene
rs56391924	10	32745248	C	0.07	0.54	0.18	4.7E–04	<i>CCDC7</i> (missense)	miR4273	NA
rs117299563	19	52095600	C	0.02	0.28	0.37	4.9E–04	<i>FLJ30403</i> (exon, 3'-UTR)	NA	miR3684
rs6072275	20	39743905	A	0.15	1.59	0.13	4.9E–04	<i>TOPI</i> (intron)	GWAS	NA
rs107321	22	18512282	T	0.41	0.72	0.09	4.9E–04	<i>FLJ41941</i> (exon)	miR1284 miR337-3p miR374a-5p	NA
rs1972820	2	212243422	G	0.36	1.41	0.10	4.9E–04	<i>ERBB4</i> (3'-UTR)	miR4633-5p miR532-5p	miR3144-3p miR875-5p
rs12247495	10	32802829	C	0.07	0.54	0.18	4.9E–04	<i>CCDC7</i> (intron)	miR539-5p	NA

NOTE: A1 miRNA and A2 miRNA refer to miRNAs predicted to bind in the presence of each alternative allele. Gray indicates the availability of high-throughput genotype data and gene expression data for the nearest gene.

Abbreviation: NA, no prediction found.

^aA1, effect allele for the corresponding odds ratios.

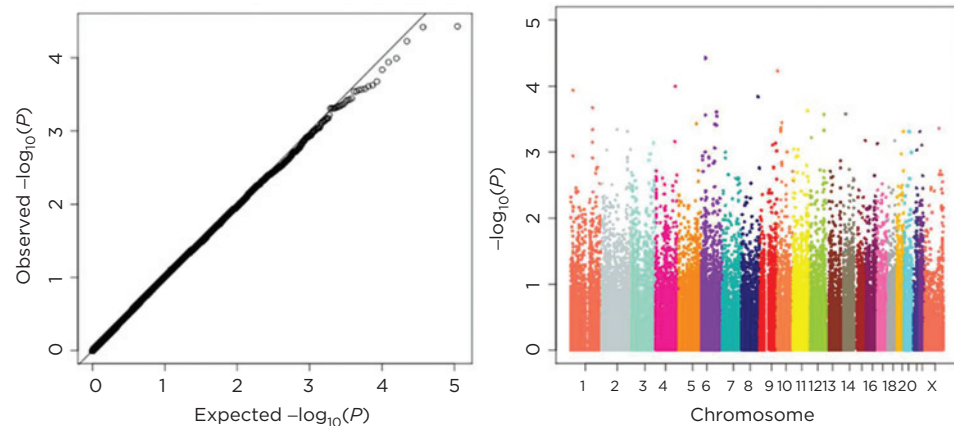
^bSE, standard error of the beta estimate.

Axiom 1.3M custom array in the same set of samples. For the 14,436 overlapping markers, there was a 99.89% overall genotype concordance across arrays. However, because the number of

directly measured markers shared by both arrays was low, we then compared genotypes for directly measured markers on the miRNA targeted array with 1000 Genomes imputed genotypes

Figure 1.

Q-Q and Manhattan plots of P values for association between each genetic marker and colorectal cancer from logistic regression models adjusted for sex, age, and two principal components for ancestry. $MAF \geq 1\%$; $n_{\text{marker}} = 55,406$. $GC \lambda = 1$.



from the Axiom 1.3M custom array. Of the 88,205 directly genotyped, post-quality control markers on the miRNA-targeted array, 63,407 were imputed with high quality from the Axiom 1.3M custom array. A comparison of the 63,407 miRNA-targeted array genotypes and the corresponding best call genotypes derived from the Axiom 1.3M custom array imputation showed that heterozygote genotype concordance was severely depressed for SNPs with $MAF \leq 5\%$. That is, consistent with prior GWAS studies using imputed genotypes, imputation did not perform well for the less common alleles important in the miRNA pathway and did not accurately reflect the directly measured genotypes.

Discussion

To our knowledge, this is the first study to examine associations between genetic variation in miRNA-encoding genes or target seed sites and risk of colorectal cancer using a genome-wide approach informed by bioinformatic miRNA prediction algorithms. While we did not identify any genome-wide significant associations meeting the traditional threshold of 5×10^{-8} , this study did highlight suggestive variants with predicted miRNA-binding implications. These findings led us to replicate a previously reported association between rs1051690 in *INSR* and colorectal

cancer risk and to demonstrate variability in *INSR* gene expression by genotype at this locus. While limited with respect to power, our initial study of only 596 cases and 429 controls demonstrated the potential for a targeted miRNA GWAS approach to identify candidate susceptibility loci and to prioritize them based on biologic insights for further functional characterization.

Alterations of expression from miRNA targets may be mediated by seed site polymorphisms that strengthen or weaken the miRNA-mRNA interaction. We illustrated a relevant example in this study from an association finding through eQTL analysis, and more generally, demonstrated our novel approach (Fig. 3). The *INSR* 3'-UTR variant (rs1051690) association with colorectal cancer had previously been detected in both Czech Republic and Spanish case-control studies assuming a codominant model (18, 21). We were able to replicate this risk locus based on a log-additive genetic model assumption. To date, few studies have examined the functional consequences miRNA-related SNPs. However, *INSR* is a notable exception. The same group that originally identified the *INSR* association later conducted *in vitro* luciferase reporter assays to show that the minor allele differentially regulates reporter gene expression (21). Evidence from our eQTL analysis corroborates this finding and provides an example of how such a target site polymorphism could influence that same gene's expression in a

Figure 2.

RAPGEF2 gene expression [$\log_2(\text{normalized intensity})$] measured by two separate probes in 72 MECC colorectal cancer cases by rs6827968 genotype.

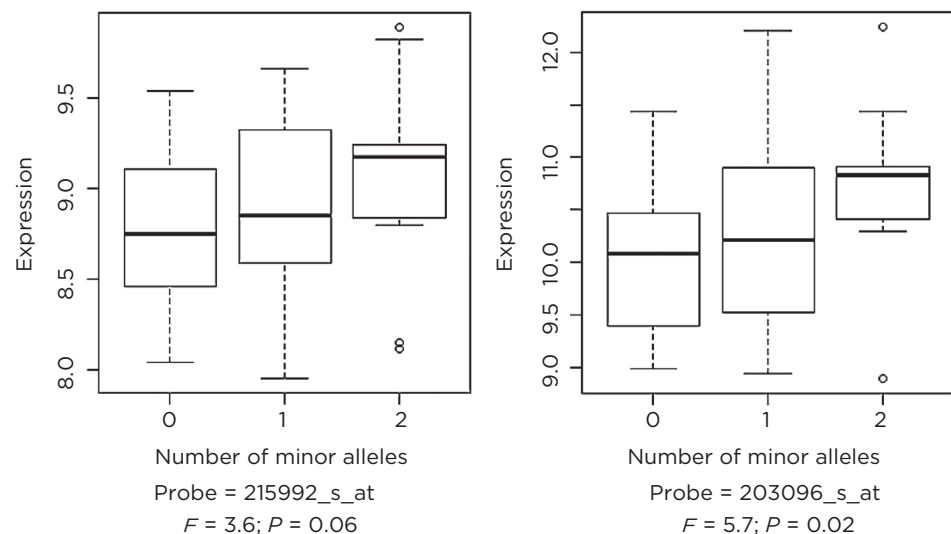


Table 3. MECC association results from logistic regression adjusted for age, sex, and principal components 1-2 for Axiom miRNA Target Site Genotyping Array markers previously studied as miRNA-related variants

SNP	CHR	Position	Gene	A1 ^a	Published OR	Published P	MECC OR	MECC SE ^b	MECC P
rs1051690 ^c	19	7116963	<i>INSR</i>	T	1.86	0.05	1.38	0.15	0.03
rs1368439 ^c	5	158742014	<i>IL12B</i>	G	1.17	0.65	0.80	0.12	0.06
rs11515 ^d	9	21968199	<i>CDKN2A</i>	G	1.16	0.71	1.14	0.12	0.30
rs1126547 ^e	3	14186757	<i>XPC</i>	C	1.13	0.73	1.14	0.13	0.32
rs3135500 ^c	16	50766886	<i>NOD2</i>	A	1.22	0.07	1.09	0.09	0.37
rs1131445 ^c	15	81601782	<i>IL16</i>	C	0.99	1.00	0.93	0.09	0.41
rs1131445 ^d	15	81601782	<i>IL16</i>	C	2.21	0.004	0.93	0.09	0.41
rs1051208 ^d	3	12625747	<i>RAF1</i>	T	1.11	0.85	0.90	0.14	0.44
rs4596 ^e	11	18388128	<i>GTF2H1</i>	C	0.79	0.03	0.93	0.10	0.44
rs2229090 ^e	3	14187345	<i>XPC</i>	C	0.91	0.38	1.08	0.12	0.54
rs17281995 ^c	3	121839641	<i>CD86</i>	C	2.93	0.01	1.05	0.12	0.66
rs11677 ^c	1	20301964	<i>PLA2G2A</i>	T	1.02	0.97	1.06	0.14	0.70
rs7356 ^e	1	28218100	<i>RPA2</i>	C	1.33	0.04	1.02	0.09	0.81
rs1803541 ^e	2	128014913	<i>ERCC3</i>	T	0.96	0.70	0.97	0.14	0.86
rs16870224 ^c	5	40692940	<i>PTGER4</i>	A	2.31	0.14	0.99	0.12	0.94
rs16870224 ^d	5	40692940	<i>PTGER4</i>	A	0.29	0.11	0.99	0.12	0.94
rs4781563 ^e	16	14045399	<i>ERCC4</i>	A	0.68	0.09	1.00	0.10	0.97
rs916055 ^c	17	4534834	<i>ALOX15</i>	C	0.98	0.91	1.00	0.10	0.98
rs743554 ^d	17	73754248	<i>ITGB4</i>	A	0.76	0.36	NA	NA	NA

NOTE: Data row 1 shows replication of a variant in *INSR* that was suggestively associated with colorectal cancer risk in prior publications by Landi et al. in 2008 and 2011 [OR (95% CI), 1.86 (0.99–3.50); $P = 0.052$].

^aA1, effect allele for the corresponding ORs.

^bSE, standard error of the beta estimate.

^cLandi et al. 2008 (18).

^dAzimzadeh et al. 2012 (19).

^eNaccarati et al. 2012 (20).

dose-response manner (Supplementary Fig. S2). A link between insulin resistance and colorectal cancer has long been recognized (39). It is possible that each additional copy of the minor/risk allele reduces miRNA–mRNA binding to the point of inhibiting mRNA degradation, which is what may lead to the increased *INSR* gene expression observation. It is also possible that the SNP exerts an effect analogous to haploinsufficiency, such that one copy of the major allele is not sufficient to appropriately repress *INSR* protein expression. Further functional work is necessary to elucidate this particular SNP's mechanism of action.

Another illustrative example for the success of our approach lies with rs6827968, our third most statistically significant finding from the targeted GWAS that also showed evidence as a *cis*-eQTL. Although rs6827968 is highly unlikely to exert a direct regulatory influence via the miRNA pathway on the nearest gene as it is an AIM, *RAPGEF2* encodes a protein that could plausibly be linked to colorectal cancer etiology. *RAPGEF2* activates RAS through promotion of the active GTP-bound state in a GTP/GDP-regulated signal transduction switch (40).

Sethupathy and Collins suggested in 2008 that studies elucidating the role of miRNA-related polymorphisms in complex diseases such as colorectal cancer should focus on three domains: genetic, functional (testing altered miRNA targeting mediated by genetic variation), and mechanistic (testing the mechanism by which altered miRNA leads to tumor development; 41). We provided evidence with respect to genetic and functional studies. The next step is to expand our genotyped dataset to increase power for detecting novel risk loci. With respect to genotyping platform for future studies, this study highlighted the advantage of this novel genotyping array over a traditional GWAS array with imputation based on the 1000 Genomes Project haplotypes, particularly when rare variants are of interest. A comparison of concordance between the Affymetrix miRNA Target Site Genotyping Array genotypes and imputed best call

genotypes from the CORECT Axiom 1.3M custom array showed that the targeted miRNA array has added value over the GWAS array for the thousands of markers from this regulatory pathway with $MAF \leq 5\%$. Also, functional studies are underway to identify SNP effects on miRNA-binding fidelity (for rs1051690 as well as other top association findings) and to find the best *in vitro* model for allele-specific effects. Furthermore, replication and fine-mapping will strengthen our confidence in both novel and previously published findings. Finally, this study suggests the benefit of reexamining previously published colorectal cancer susceptibility regions identified through GWAS for potential functional SNPs in miRNA-binding sites or other miRNA pathway-related sequences. Given that most GWAS risk loci identified to date have $MAF \geq 5\%$, reanalysis of existing, imputed GWAS datasets using the bioinformatic approaches described here has a high probability of yielding insights into the functional relevance of these regions.

This study has limitations with respect to power and modeling assumptions. Our sample size is limited to 1,025 samples. However, this analysis, which was able to replicate a previously identified miRNA risk locus and characterize preliminary functionality, provides justification for study in a larger sample. Our lack of genome-wide significant findings is likely attributable to a lack of power, and our sample size did not permit the investigation of effects for rare variants with $MAF < 1\%$. Also, not all SNPs exert their effects according to the assumed log-additive genetic model, and this choice made to restrict multiple testing could inhibit our ability to identify risk loci that are consistent with a recessive, dominant, or codominant model. Furthermore, we did not consider interactions between these potentially risk-conferring variants or variant effects in the context of environmental risk factors. Finally, our ability to examine gene expression was limited by data availability and restriction to studying the SNP's nearest gene.

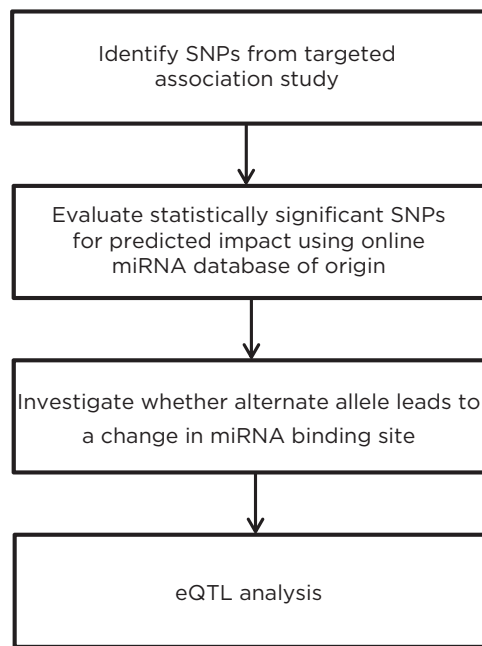


Figure 3.

Bioinformatic pipeline for the predicted impact of statistically significant SNPs on miRNA target seed sites and *cis* gene expression. SNPs identified from the targeted association study were evaluated using PolymiRTS, Patrocles, and microRNA.org (in conjunction with markers identified by the 1000 Genomes Project) for indications of target seed sites that would be disrupted by the presence of the alternate allele for each genetic variant. Markers from the PolymiRTS database were further annotated for potential creation of new seed sites. Statistically significant markers were further investigated with MicroSNiPER to identify additional target seed sites both disrupted and potentially created by the presence of the alternate allele. The final step was *cis*-eQTL analysis where gene expression data were available.

Despite these limitations, we provide evidence that a targeted genome-wide approach for studying germline susceptibility can be extended beyond known or purported cancer biology pathways to the exploration of a regulatory pathway with widespread post-transcriptional effects. A better understanding of the mechanisms by which aberrations in miRNA expression

and binding impact colorectal cancer development and progression may offer insights for prevention and targeted therapeutics. Specifically, the *INSR* variant warrants further investigation in a functional setting to elucidate its role in the alteration of colorectal cancer risk.

Disclosure of Potential Conflicts of Interest

A. Finn is a genotyping specialist at Affymetrix Inc. No potential conflicts of interest were disclosed by the other authors.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' Contributions

Conception and design: S.L. Schmit, A. Finn, S.B. Gruber
Development of methodology: S.L. Schmit, J. Gollub, H.S. Rennert, S.B. Gruber
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Gollub, H.S. Rennert, G. Rennert, S.B. Gruber
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.L. Schmit, H.S. Rennert, G. Rennert, S.B. Gruber
Writing, review, and/or revision of the manuscript: S.L. Schmit, M.H. Shaperro, H.S. Rennert, A. Finn
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Gollub, S.-C. Huang, S.B. Gruber
Study supervision: G. Rennert, S.B. Gruber

Acknowledgments

The authors acknowledge the ColoRectal Transdisciplinary (CORECT) Study and its investigators for contributing genotypes for a comparative analysis of the miRNA-targeted array versus a custom GWAS array.

Grant Support

This work was supported by the National Cancer Institute at the NIH (U19 CA148107 to S.B. Gruber; P30 CA014089 to S.B. Gruber; and R01 CA81488 to S.B. Gruber), the National Human Genome Research Institute at the NIH (T32 HG000040 to S.L. Schmit), the National Institute of Environmental Health Sciences at the NIH (T32 ES013678 to S.L. Schmit), and the Rackham Graduate School at the University of Michigan (Rackham Predoctoral Fellowship to S.L. Schmit).

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

Received March 17, 2014; revised September 23, 2014; accepted October 6, 2014; published OnlineFirst October 23, 2014.

References

- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;136:215–33.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.
- Ambros V. The functions of animal microRNAs. *Nature* 2004;431:350–5.
- Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 2005;436:740–4.
- Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 2002;297:2056–60.
- Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol* 2011;6:479–507.
- Vilar E, Taberero J, Gruber SB. Micromanaging the classification of colon cancer: the role of the microRNAome. *Clin Cancer Res* 2011;17:7207–9.
- Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–66.
- Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, et al. The colorectal microRNAome. *Proc Natl Acad Sci U S A* 2006;103:3687–92.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759–67.
- Olson P, Lu J, Zhang H, Shai A, Chun MG, Wang Y, et al. MicroRNA dynamics in the stages of tumorigenesis correlate with hallmark capabilities of cancer. *Genes Dev* 2009;23:2152–65.
- Balaguer F, Moreira L, Lozano JJ, Link A, Ramirez G, Shen Y, et al. Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clin Cancer Res* 2011;17:6239–49.
- Bartley AN, Yao H, Barkoh BA, Ivan C, Mishra BM, Rashid A, et al. Complex patterns of altered MicroRNA expression during the adenoma-adenocarcinoma sequence for microsatellite-stable colorectal cancer. *Clin Cancer Res* 2011;17:7283–93.
- Schee K, Fodstad O, Flatmark K. MicroRNAs as biomarkers in colorectal cancer. *Am J Pathol* 2010;177:1592–9.
- Liu M, Chen H. The role of microRNAs in colorectal cancer. *J Genet Genomics* 2010;37:347–58.
- Ju J. miRNAs as biomarkers in colorectal cancer diagnosis and prognosis. *Bioanalysis* 2010;2:901–6.

17. Vinci S, Gelmini S, Mancini I, Malentacchi F, Pazzagli M, Beltrami C, et al. Genetic and epigenetic factors in regulation of microRNA in colorectal cancers. *Methods* 2013;59:138–46.
18. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, et al. Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis* 2008;29:579–84.
19. Azimzadeh P, Romani S, Mohebbi SR, Mahmoudi T, Vahedi M, Fatemi SR, et al. Association of polymorphisms in microRNA-binding sites and colorectal cancer in an Iranian population. *Cancer Genet* 2012;205:501–7.
20. Naccarati A, Pardini B, Stefano L, Landi D, Slysikova J, Novotny J, et al. Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. *Carcinogenesis* 2012;33:1346–51.
21. Landi D, Moreno V, Guino E, Vodicka P, Pardini B, Naccarati A, et al. Polymorphisms affecting micro-RNA regulation and associated with the risk of dietary-related cancers: a review from the literature and new evidence for a functional role of rs17281995 (CD86) and rs1051690 (INSR), previously associated with colorectal cancer. *Mutat Res* 2011;717:109–15.
22. Fireman Z, Sandler E, Kopelman Y, Segal A, Sternberg A. Ethnic differences in colorectal cancer among Arab and Jewish neighbors in Israel. *Am J Gastroenterol* 2001;96:204–7.
23. Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS, et al. Statins and the risk of colorectal cancer. *N Engl J Med* 2005;352:2184–92.
24. Ziebarth JD, Bhattacharya A, Chen A, Cui Y. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res* 2012;40:D216–21.
25. Schmeier S, Schaefer U, MacPherson CR, Bajic VB. dPORE-miRNA: polymorphic regulation of microRNA genes. *PLoS ONE* 2011;6:e16657.
26. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res* 2010;38:D640–51.
27. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* 2010;11:R90.
28. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res* 2008;36:D149–53.
29. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009;106:9362–7.
30. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
31. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
32. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 2007;23:1164–7.
33. Vilar E, Bartnik CM, Stenzel SL, Raskin L, Ahn J, Moreno V, et al. MRE11 deficiency increases sensitivity to poly(ADP-ribose) polymerase inhibition in microsatellite unstable colorectal cancers. *Cancer Res* 2011;71:2632–42.
34. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;39:D152–7.
35. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;36:D154–8.
36. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006;34:D140–4.
37. Griffiths-Jones S. The microRNA registry. *Nucleic Acids Res* 2004;32:D109–11.
38. Barenboim M, Zoltick BJ, Guo Y, Weinberger DR. MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. *Hum Mutat* 2010;31:1223–32.
39. Giovannucci E. Insulin and colon cancer. *Cancer Causes Control* 1995;6:164–79.
40. Rebhun JF, Castro AF, Quilliam LA. Identification of guanine nucleotide exchange factors (GEFs) for the Rap1 GTPase. Regulation of MR-GEF by M-Ras-GTP interaction. *J Biol Chem* 2000;275:34901–8.
41. Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. *Trends Genet* 2008;24:489–97.