GENES, CHROMOSOMES & CANCER 56:177–184 (2017)

RESEARCH ARTICLES

Germline miRNA DNA Variants and the Risk of Colorectal Cancer by Subtype

Noralane M. Lindor,[1]* Melissa C. Larson,[2] Melissa S. DeRycke,[2,3] Shannon K. McDonnell,[2] Saurabh Baheti,[2] Zachary C. Fogarty,[2] Aung Ko Win,[4] John D. Potter,[5,6,7] Daniel D. Buchanan,[4,7] Mark Clendenning,[7] Polly A. Newcomb,[5] Graham Casey,[8] Steven Gallinger,[9] Loïc Le Marchand,[10] John L. Hopper,[4,11] Mark A. Jenkins,[4] Ellen L. Goode,[2] and Stephen N. Thibodeau[3]

[1]Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ
[2]Department of Health Sciences Research, Mayo Clinic, Rochester, MN
[3]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN
[4]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, Victoria, Australia
[5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA
[6]School of Public Health, University of Washington, Seattle, WA
[7]Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Parkville, Victoria, Australia
[8]Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA
[9]Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
[10]University of Hawaii Cancer Center, Honolulu, HI
[11]Department of Epidemiology and Institute of Health and Environment, School of Public Health, Seoul National University, Seoul, South Korea

MicroRNAs (miRNAs) regulate up to one-third of all protein-coding genes including genes relevant to cancer. Variants within miRNAs have been reported to be associated with prognosis, survival, response to chemotherapy across cancer types, *in vitro* parameters of cell growth, and altered risks for development of cancer. Five miRNA variants have been reported to be associated with risk for development of colorectal cancer (CRC). In this study, we evaluated germline genetic variation in 1,123 miRNAs in 899 individuals with CRCs categorized by clinical subtypes and in 204 controls. The role of common miRNA variation in CRC was investigated using single variant and miRNA-level association tests. Twenty-nine miRNAs and 30 variants exhibited some marginal association with CRC in at least one subtype of CRC. Previously

Additional Supporting Information may be found in the online version of this article.

Supported by: National Cancer Institute and through cooperative agreements with the following CCFR centers, Grant number: UM1 CA167551; Australasian Colorectal Cancer Family Registry, Grant numbers: U01 CA074778 and U01/U24 CA097735; Mayo Clinic Cooperative Family Registry for Colon Cancer Studies, Grant number: U01/U24 CA074800; Ontario Familial Colorectal Cancer Registry, Grant number: U01/U24 CA074783; Seattle Colorectal Cancer Family Registry, Grant number: U01/U24 CA074794; University of Hawaii Colorectal Cancer Family Registry, Grant number: U01/U24 CA074806; USC Consortium Colorectal Cancer Family Registry, Grant number: U01/U24 CA074799; Cancer Surveillance System of the Fred Hutchinson Cancer Research Center (Seattle CCFR research), Grant numbers: N01-CN-67009 (1996-2003) and N01-PC-35142 (2003-2010); from the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute with additional support from the Fred Hutchinson Cancer Research Center; Grant numbers: HHSN2612013000121 (2010-2017); Hawai'i Department of Health as part of the statewide cancer reporting program mandated by Hawai'i Revised Statutes (the collection of cancer incidence data for the State of Hawai'i); National Cancer Institute's Surveillance, Epidemiology and End Results Program (SEER), Grant numbers: N01-PC-67001 (1996-2003) and N01-PC-35137 (2003-2010); University of Hawai'i, Grant numbers: HHSN26120100037C (2010-2013) and HHSN261201300009I (2010-current); State of Hawai'i, Department of Health, the National Cancer Institute; SEER Program or their Contractors and Subcontractors is not intended nor should be inferred;. California Department of Public Health (California Health and Safety Code Section 103885); National Cancer Institute's Surveillance, Epidemiology and End Results Program (Cancer Prevention Institute of California), Grant number: HHSN261201000140C; University of Southern California, Grant number: HHSN261201000035C; Public Health Institute, HHSN261201000034C; Centers for Disease Control and Prevention's National Program of Cancer Registries (California Department of Public Health), Grant number: U58DP003862-01; State of California, Department of Public Health the National Cancer Institute; Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred; National Cancer Institute or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CCFR.

*Correspondence to: Noralane M. Lindor; Mayo Clinic, 13400 E. Shea Blvd., Scottsdale, AZ 85259, USA. E-mail: nlindor@mayo.edu

Received 6 June 2016; Revised 28 July 2016; Accepted 4 August 2016

DOI 10.1002/gcc.22420

Published online 30 November 2016 in
Wiley Online Library (wileyonlinelibrary.com).

reported associations were not confirmed ($n = 4$) or could not be evaluated ($n = 1$). The variants noted for the CRCs with deficient mismatch repair showed little overlap with the variants noted for CRCs with proficient mismatch repair, consistent with our evolving understanding of the distinct biology underlying these two groups.

INTRODUCTION

MicroRNAs (miRNAs) are a large group of noncoding RNAs, discovered in 1993 in *Caenorhabditis elegans*, now documented in many other organisms. They act in *trans*- upon *cis*-regulatory elements in target messenger RNAs, affecting protein translation (Lee et al., 1993; Wightman et al., 1993). MiRNAs are located in introns of other genes and intergenic regions of the genome. Genes encoding miRNAs account for approximately 2% of the coding genes, and regulate up to 30% of all protein-coding genes, notably regulating pathways relevant to cancer such as cell growth, differentiation, and apoptosis. MiRNAs can function as tumor-suppressor genes or oncogenes, depending on whether deleted or overexpressed (Hayashita et al., 2005; Esquela-Kerscher and Slack, 2006; Croce, 2009; Medina et al., 2010; Salzman and Weidhaas, 2013). Approximately 50% of miRNAs are in fragile regions of the genome that are often deleted, amplified, or misexpressed in cancers (Calin and Croce, 2006). MiRNAs are collated in a publicly available database, miRbase (http://www.mirbase.org/index.shtml). The most recent version (21; accessed May 12, 2015) contains 28,645 entries representing hairpin precursor miRNAs, expressing 35,828 mature miRNA products, in 223 different species. For *Homo sapiens*, 1,881 unique miRNAs are listed.

The sequence of an individual miRNA, on average 20–24 nucleotides long, determines its target, complementary to a portion of the 3′ UTR of the target gene's mRNA. Nucleotides 2–7 from the 5′ end of the miRNA are the major determinants of target selection for inhibition of expression. MiRNAs and target sites are highly conserved through evolution (Chen and Rajewsky, 2006) and single nucleotide polymorphisms (SNPs) located within miRNA are uncommon but do exist. Recent studies report associations between specific SNPs and prognosis, survival, response to chemotherapy across cancer types, *in vitro* parameters of cell growth, and risks for the development of cancer (Srivastava and Srivastava, 2012).

Exploration of an association between miRNA variants and colorectal cancer (CRC) has been limited. In this study, we estimated the frequency of germline SNPs and small insertions or deletions (indels) in miRNA using a targeted sequencing procedure. Association with CRC was evaluated for five categories of individuals diagnosed with CRC compared with controls. We also compared our findings with published literature on CRC associated with miRNA variants.

MATERIALS AND METHODS

Study Samples

Study samples were from the Colon Cancer Family Registry (Colon CFR), described in detail elsewhere (Newcomb et al., 2007) and at http://coloncfr.org. Between 1997 and 2012, the Colon CFR recruited families via both population-based probands, recently diagnosed CRC cases from state or regional cancer registries in Australia, the USA, and Canada as well as clinic-based probands enrolled from multiple-case families referred to family-cancer clinics in the same countries. Samples in this study were collected from the Australasian Colorectal Cancer Family Registry (Melbourne, Victoria, Australia), Hawaii Family Registry of Colon Cancer (Honolulu, HI), Mayo Colorectal Family Registry (Rochester, MN), Ontario Familial Colorectal Cancer Registry (Toronto, Ontario, Canada), Seattle Familial Colorectal Cancer Registry (Seattle, WA), and University of Southern California Consortium (Los Angeles, CA). Mismatch repair (MMR) status for all tumors was established, as previously described (Ait Ouakrim et al., 2015). All participants provided informed consent. Protocols were approved by the Institutional Review Board at each site.

Sequencing

MiRBase was used to identify 1,424 miRNAs for sequencing of the entire pre-miRNA (http://www.mirbase.org/, build 17; see Supporting information).

Bioinformatics Analysis

Details of bioinformatics analysis are shown in the Supporting information.

Quality Control

Comprehensive quality control (QC) identified poor quality samples and potential sequencing

TABLE 1. Numbers of Individuals Within Different Categories Used in Analyses

Group	Original, *n*	Passed quality control, *n*	Quality control and European[a], *n*
Controls			204
Mismatch repair carrier control	165	163	113
Noncarrier spousal control	95	91	91
dMMR, no mutation	147	147	129
FCCTX/pMMR linkage	288	285	229
Young Onset	234	234	206
Unselected	602	602	335
Likely pMMR[b]	1,076	1,070	734
Total assigned cases	1,271	1,265	899

[a]European subset defined as samples with >80% European ancestry based on STRUCTURE.
[b]Combination of FCCTX/pMMR linkage, Young Onset, and unselected cases.

batch effects. We investigated per-sample percent duplicated reads, coverage of the capture region, variant calling quality and depth, variant call-rate in the capture region, heterozygosity rate, transition:transversion ratio, and sex verification using PLINK/SEQ v0.10 (https://atgu.mgh.harvard.edu/plinkseq/index.shtml) and PLINK v1.9 (https://www.cog-genomics.org/plink2). Sample contamination was visually inspected by plotting the fraction of ALT reads at common variant positions against the 1000 Genomes Project allele frequency; samples displaying more than three bands or a "shotgun pattern" indicate probable contamination or poor quality DNA. Pedigree Relationship Statistical Test-Plus was used to identify related samples and population stratification was evaluated using STRUCTURE software (Patterson et al., 2006; Price et al., 2006). Samples with <90% of the capture region covered at 10X, call rate within the capture region <95%, suspected sample contamination, unexpected familial relationships, and <80% European ancestry were excluded from analysis. For variant quality filtering, we included GATK VQSR filtering tranche 99.0 and above. Polymorphic variants mapping to five or more locations in the genome, those with call rate <95%, monomorphic, and those with Hardy–Weinberg equilibrium P value <1E−5 in our controls were excluded.

Analysis

Six case–control analyses were performed, comparing five CRC case sets and all cases combined to the combined group of controls. For CRC cases for whom tumor testing had been conducted, each was categorized as having deficient or proficient DNA mismatch repair tumors (dMMR and pMMR, respectively) (Lindor et al., 2002). We defined five categories of cases: those with dMMR tumors for which no germline mutation could be identified (dMMR); familial colorectal cancer type X cases (Lindor et al., 2006) combined with those from other pMMR multi-case-CRC families from a prior linkage study, not otherwise specified (FCCTX/pMMR linkage); pMMR CRC diagnosed before age 50 years (Young Onset); those for which no tumor had been available for MMR characterization and no causal MMR gene mutation had been found by sequencing ("unselected" [referring to tumor MMR status which was unknown as no tumor was available for testing]); and a combined group (Likely pMMR) that included nonoverlapping cases from the FCCTX/pMMR linkage, Young Onset, and unselected cases (Table 1). To increase power, "controls: included non-carrier spouses and MMR carriers, the cause of whose CRC is considered known. We used principal components analysis as implemented in the *SNPRelate* R package (Zheng et al., 2012) to evaluate sample eigenvectors as covariates to adjust for possible population stratification. The SNPs used for principal components analysis included approximately 2,000 common (MAF >5%), independent (linkage disequilibrium $r^2 < 0.4$), autosomal SNPs. None of the top eigenvectors was associated with case-control status indicating that no population stratification adjustment was necessary.

We performed both single-SNP- and miRNA-level analyses using an extension to commonly used gene-based statistics to allow for known pedigree relationships (Schaid et al., 2013). For miRNA-level tests, analyses were conducted using both a burden test (most powerful if variants in a gene have effects in the same direction) and kernel statistic (most powerful if variants have effects in opposite directions). Variants were weighted using beta density weights of (1, 25), with rare variants receiving a higher weight. False discovery rate was calculated using the R package Q-value

(Storey et al., 2015) and considering all case–control comparisons.

Our approach allows for both pedigree data, for example, multiple cases from a family as well as unrelated subjects and takes a retrospective view treating the trait as fixed and genotypes as random, allowing complex and undefined ascertainment of pedigrees as is typical for many of the pedigrees included in our study.

We conducted a literature search for miRNA variants reported to be associated with altered risk for CRC. The frequency of these variants was determined in all our subtypes and controls.

RESULTS

A total of 1,436 individuals with CRC and 95 unaffected spouse controls were selected (Table 1). Those with CRC included 165 CRCs in individuals with known MMR germline mutations (which were used as mutation-positive controls). After removing samples failing QC, unexpected duplicate results, cryptically related individuals, and non-European ancestry, 1,103 subjects were included in the analysis. Final comparison groups were All cases (n = 899), with case subsets including dMMR (n = 129), FCCTX/pMMR Linkage (n = 229), Young Onset (n = 206), unselected (n = 335), and Likely pMMR (n = 734) cases. Spouses and affected MMR carriers served as controls (n = 204) for case–control comparisons.

A total of 1,316 variants in 689 miRNA passed QC filters and variants in 575 miRNAs were polymorphic in European samples and were included in analysis. Three hundred eighty miRNA had more than a single variant available for analysis and were included in miRNA-level analyses, while 242 variants in 195 miRNA with MAF >1% were analyzed for single-variant association. The average number of variants per miRNA included in miRNA-level analysis was 3.1 (range 2–24) with over half (n = 210, 55%) of miRNA including only two variants. Considering multiple testing, none of the miRNA-level tests was statistically significant (minimum miRNA-level P value = 0.003, false discovery rate = 0.41). MiRNA exhibiting a marginal positive association in at least one CRC subtype are presented in Table 2 where "marginal association" is defined as a higher frequency of rare variants in cases (positive-burden statistic) and a kernel statistic P value <0.10. Single variants meeting these same criteria are presented in Table 3. The miRNA location of the variant and the frequency of that variant in public databases is included in Table S1. The frequency of variants with minor allele frequency >1% in our controls matched that of the 1000 Genomes Project well (Fig. S1).

Analysis of the same CRC subtypes using only the spousal controls was conducted but did not substantively change results (results not shown). The decision to combine the MMR positive DNAs with the spouse controls was based on the reasonable hypothesis that CRC in individuals with known MMR deficiency was explained by the germline mutation and the probability of other contributing factors approximates that of the general population.

Five miRNA variants were found in the literature reporting altered risks for CRC including rs11614913 (miRNA196a2), rs2910164 (miRNA146a), rs4938723 (miRNA34b/c), rs2292832 (miRNA149), and rs3746444 (miRNA499; Table S2). Our results did not confirm these associations in the four variants we could evaluate. The fifth variant, rs4938723 in miRNA34b/c, had no coverage in our dataset, perhaps indicating a failure in the sequencing capture.

DISCUSSION

In this observational case-control study, we sought to evaluate whether variants that occur in miRNA genes were associated with CRC. Of the 1,424 different miRNAs studied, 29 miRNAs and 30 variants exhibited some marginal association in at least one subtype of CRC. No variant was associated with all subtypes of CRC (Tables 2 and 3). The miRNAs of interest (albeit marginally significant) were not found in previous studies. It is notable that our subgroup with definite dMMR tumors exhibited association with miRNAs that had little overlap with the miRNAs associated with the other predominantly pMMR groups. This is not unexpected based upon knowledge of the fundamentally different underlying biology of the dMMR group, recently reaffirmed by new definitions of consensus molecular subtypes (Guinney et al., 2015). We acknowledge that we had limited power to detect overlap. Therefore, although lack of overlap may be consistent with different biology, it does not confirm it. The results of the present study do support the importance of conducting research that does not ignore the well-defined molecular heterogeneity of CRC.

Five miRNA variants have been associated with altered risk for CRC in some but not all studies

TABLE 2. MiRNA Genes Exhibiting Marginal Increased Rare-Variant Frequency[a] in Cases Versus Controls[b]

				All cases (n = 899)		dMMR cases (n = 129)		FCCTX/pMMR linkage cases (n = 229)		Young Onset cases (n = 206)		Unselected cases (n = 335)		pMMR cases (n = 734)		
miRNA	Chromosome	Start	Stop	Number of variants	Signed kernel P value[c]	Number of variants	Signed kernel P value[c]	Number of variants	Signed kernel P value[c]	Number of variants	Signed kernel P value[c]	Number of variants	Signed kernel P value[c]	Number of variants	Signed kernel P value[c]	Minimum FDR Q-value[d]
MIR1262	chr1	68649200	68649293	2	**0.0802**	–	–	–	–	–	–	2	**0.0421**	2	**0.0735**	0.410
MIR216A	chr2	56216084	56216194	2	0.1108	–	–	2	0.3373	–	–	–	–	2	**0.0985**	0.410
MIR3679	chr2	134884695	134884763	8	0.1861	5	**0.0255**	7	0.1898	5	0.1346	6	0.6089	8	0.2638	0.410
MIR3138	chr4	10080234	10080316	4	**0.0640**	4	0.3276	4	−0.3837	3	0.1170	3	**0.0258**	4	**0.0965**	0.410
MIR1289-2	chr5	132763287	132763398	3	−0.5971	3	**0.0049**	3	0.7297	2	−0.6346	2	−0.0254	3	−0.3347	0.410
hsa-mir-1294	chr5	153726665	153726807	4	0.2487	2	0.2606	4	0.7498	2	**0.0913**	2	0.4335	4	0.2404	0.410
MIR4465	chr6	141004950	141005020	6	0.4849	6	**0.0766**	–	–	–	–	–	–	–	–	0.410
MIR4470	chr8	62627346	62627418	5	0.3635	2	**0.0748**	2	0.4219	4	0.4114	2	0.2112	4	0.3937	0.410
MIR4289	chr9	91360750	91360820	5	0.1443	4	0.8337	4	**0.0462**	3	0.1564	4	0.1093	5	**0.0887**	0.410
MIR199B	chr9	131006999	131007109	3	0.1228	2	**0.0648**	2	0.1440	2	−0.3654	3	**0.0699**	3	0.1675	0.410
MIR129-2	chr11	43602943	43603033	2	**0.0606**	–	–	–	–	–	–	2	0.1349	–	–	0.410
MIR612	chr11	65211928	65212028	2	0.9413	2	0.7416	2	−0.1764	2	**0.0432**	2	−0.5867	2	−0.9554	0.410
MIR381HG	chr14	101511493	101518132	14	**0.0684**	7	0.2151	10	0.1382	9	**0.0058**	9	0.3512	14	**0.0826**	0.410
MIR656	chr14	101533060	101533138	3	0.1277	2	**0.0852**	–	–	2	0.1137	–	–	2	0.1260	0.410
MIR1225	chr16	2140195	2140285	3	0.4868	2	**0.0518**	2	−0.7932	2	−0.8113	3	−0.7588	3	−0.6683	0.410
MIR4520A	chr17	6558758	6558828	3	0.1125	3	0.2301	2	0.2811	2	−0.7596	3	**0.0132**	3	0.1349	0.410
MIR4520B	chr17	6558767	6558821	3	0.1125	3	0.2301	2	0.2811	2	−0.7596	3	**0.0132**	3	0.1349	0.410
MIR4743	chr18	46196970	46197039	2	0.2919	2	**0.0477**	2	0.4868	2	0.4817	2	0.1682	2	0.4333	0.410
MIR3190	chr19	47730198	47730278	2	0.2308	–	–	–	–	–	–	2	**0.0107**	2	0.1658	0.410
MIR3192	chr20	18451258	18451335	2	**0.0385**	–	–	–	–	–	–	2	**0.0094**	2	**0.0387**	0.410
MIR499A	chr20	33578178	33578300	6	0.2572	2	**0.0770**	3	0.4153	2	**0.0476**	5	0.6552	6	0.3162	0.410
MIR941-3	chr20	62550833	62550947	4	**0.0884**	–	–	3	**0.0731**	–	–	2	0.1349	4	0.1187	0.410
MIR941-4	chr20	62550836	62550950	5	**0.0981**	4	**0.0283**	4	0.1720	4	0.1425	5	0.4872	5	0.2206	0.410
MIR941-3	chr20	62550889	62551003	4	**0.0884**	–	–	3	**0.0731**	–	–	2	0.1349	4	0.1187	0.410
MIR941-4	chr20	62550892	62551006	5	**0.0981**	4	**0.0283**	4	0.1720	4	0.1425	5	0.4872	5	0.2206	0.410
MIR941-3	chr20	62551084	62551201	4	**0.0884**	–	–	3	**0.0731**	–	–	2	0.1349	4	0.1187	0.410
MIR941-4	chr20	62551196	62551313	5	**0.0981**	4	**0.0283**	4	0.1720	4	0.1425	5	0.4872	5	0.2206	0.410
MIR3687	chr21	9826202	9826263	5	0.1122	3	0.1398	3	0.2864	5	0.3865	3	**0.0980**	5	0.1230	0.410
MIR658	chr22	38240278	38240378	3	0.7210	2	0.7436	2	0.9663	3	**0.0778**	2	−0.8359	3	0.5887	0.410

[a]Marginal evidence defined as sign of the burden statistic multiplied by kernel statistic P value in [0, 0.10].
[b]The control group consists of 91 spousal controls and 113 participants with known mismatch repair gene mutations (n = 204 total).
[c]Highlighted (bold) boxes = p value <0.10.
[d]FDR, false discovery rate.

TABLE 3. Individual miRNA Variants Exhibiting Marginal Positive Association[a] in Cases Versus Controls[b]

miRNA	Chromosome	POS[c]	REF[d]	ALT[e]	Signed kernel *P* value						Minimum FDR *Q*-value[g]
					All cases ($n = 899$)[f]	dMMR cases ($n = 129$)[f]	FCCTX/pMMR linkage cases ($n = 229$)[f]	Young Onset cases ($n = 206$)[f]	Unselected cases ($n = 335$)[f]	pMMR cases ($n = 734$)[f]	
MIR216A	chr2	56216090	A	T	0.102	0.142	0.320	0.138	0.117	**0.090**	0.271
MIR663B	chr2	133014587	C	T	0.488	**0.016**	−0.630	−0.665	0.331	0.955	0.271
MIR1258	chr2	180725568	T	C	0.519	**0.027**	0.821	−0.780	0.759	0.787	0.271
MIR4268	chr2	220771223	C	T	0.222	0.407	**0.064**	0.745	0.403	0.230	0.271
MIR4789	chr3	175087408	C	T	**0.099**	−0.525	**0.016**	**0.091**	0.224	**0.046**	0.271
hsa-mir-1294	chr5	153726769	A	G	0.265	0.241	0.789	**0.088**	0.459	0.250	0.271
hsa-mir-3144	chr6	120336327	C	A	0.166	0.643	**0.033**	0.852	0.191	0.123	0.271
MIR4467	chr7	102111936	G	A	**0.062**	0.289	**0.034**	**0.076**	0.289	**0.062**	0.271
MIR3622A:MIR3622B	chr8	27559214	G	A	0.104	**0.066**	0.278	0.117	0.271	0.138	0.271
hsa-mir-1302-7	chr8	142867668	ATGT	A	**0.022**	**0.012**	0.107	**0.026**	**0.009**	**0.021**	0.271
MIR4669	chr9	137271318	C	A	**0.008**	0.173	0.109	**0.003**	**0.030**	**0.007**	0.271
MIR3689A	chr9	137742206	C	T	**0.050**	0.354	**0.049**	0.248	**0.038**	**0.043**	0.271
MIR1908	chr11	61582708	T	C	0.841	**0.016**	−0.636	−0.424	0.885	−0.713	0.271
MIR612	chr11	65211940	C	A	0.930	0.745	−0.174	**0.043**	−0.590	−0.942	0.271
MIR492	chr12	95228286	G	C	0.199	**0.078**	**0.077**	0.905	0.435	0.284	0.271
hsa-mir-300:MIR300	chr14	101507727	C	T	0.235	0.778	0.213	**0.074**	0.619	0.198	0.271
MIR381HG	chr14	101513795	C	T	**0.042**	0.138	**0.095**	**0.003**	0.305	**0.048**	0.271
MIR656	chr14	101533093	C	T	0.118	**0.087**	0.115	0.110	0.369	0.119	0.271
MIR4513	chr15	75081078	G	A	0.189	**0.087**	0.314	−0.798	**0.095**	0.324	0.271
MIR184	chr15	79502168	G	T	0.251	0.511	0.479	0.749	**0.097**	0.336	0.271
MIR1225	chr16	2140262	T	TC	0.567	**0.047**	0.913	−0.988	0.868	0.831	0.271
MIR4520A:MIR4520B	chr17	6558808	G	A	0.104	0.241	0.279	−0.766	**0.012**	0.127	0.271
MIR423	chr17	28444183	A	C	**0.059**	0.168	0.833	0.203	**0.006**	**0.088**	0.271
MIR4745	chr19	804959	C	T	0.718	−0.759	−0.614	**0.079**	0.941	0.605	0.271
MIR3190:MIR3191	chr19	47730272	A	C	0.186	0.811	−0.398	*−0.043*	**0.00008**	0.126	0.155
MIR4751	chr19	50436371	G	A	0.121	**0.092**	0.148	0.382	0.243	0.139	0.271
MIR4754	chr19	58898193	C	T	**0.001**	0.164	**0.003**	**0.018**	**0.00100**	**0.0005**	0.271
MIR3192	chr20	18451325	T	C	**0.036**	0.135	**0.085**	0.182	**0.009**	**0.036**	0.271
hsa-mir-941-3:MIR941-2:MIR941-4	chr20	62551298	C	T	0.264	**0.071**	0.567	0.174	0.672	0.445	0.271
MIRLET7BHG	chr22	46487011	G	A	0.104	**0.034**	0.159	0.259	0.313	0.123	0.271

[a]Marginal evidence defined as sign of the burden statistic multiplied by kernel statistic *P* value in [0, 0.10].
[b]The control group consists of 91 spousal controls and 113 participants with known mismatch repair gene mutations.
[c]POS, genomic position.
[d]REF, reference allele.
[e]ALT, alternate allele.
[f]Highlighted (bold) boxes = *P* value <0.10.
[g]FDR, false discovery rate. Minimum FDR *Q*-value for each variant. FDR *Q*-value was calculated considering all case–control comparisons simultaneously.

(Table S3). We looked specifically at these five variants: one was not well captured in our dataset so could not be evaluated, but the other four were not different between cases and controls. It is notable that the majority of the published studies to date were conducted in Asian populations whereas our study was restricted to Europeans; it is possible these variants are in linkage disequilibrium with an ethnic-specific risk factor or that our study was underpowered to detect a modest association. Other investigators also report non-replication of these miRNAs in CRC cohorts of European ancestry (Hezova et al., 2012; Vinci et al., 2013; Kupcinskas et al., 2014). In addition, expression levels for these miRNAs were not reported to differ in CRC across the newly described consensus molecular subtypes of CRC (Guinney et al., 2015). Larger studies with careful attention to ethnic selection are needed to assess the validity of all observations.

One strength of this study was the ability to evaluate across well-characterized subsets of CRC cases (dMMR, pMMR, Young Onset, etc.) for whom other major germline mutations had been sought but were not found. Second, coverage of miRNAs was broad due to the inclusion of nearly all the miRNAs known at the time the study was initiated Overall, the quality of the sequencing reads was high and 1,123 miRNAs could be evaluated. One weakness was our limited sample size and the number of controls. However, our control allele frequencies matched frequencies in the 1000 Genomes European Project well (Fig. S1). Another weakness is the absence of functional studies to follow-up on our current findings, which is beyond the scope of this short report.

This study identified a list of miRNAs for which there was a suggestion of association with CRC, which varied by molecular subtype. These findings argue for additional testing in a larger study. We have provided an assessment using a European sample of four of the miRNA variants reported by others to be associated with CRC and were not able to confirm those associations even though our numbers were comparable to the discovery reports.

REFERENCES

Ait Ouakrim D, Dashti SG, Chau R, Buchanan DD, Clendenning M, Rosty C, Winship IM, Young JP, Giles GG, Leggett B, Macrae FA, Ahnen DJ, Casey G, Gallinger S, Haile RW, Le Marchand L, Thibodeau SN, Lindor NM, Newcomb PA, Potter JD, Baron JA, Hopper JL, Jenkins MA, Win AK. 2015. Aspirin, ibuprofen, and the risk for colorectal cancer in Lynch Syndrome. J Natl Cancer Inst 107.

Calin GA, Croce CM. 2006. MicroRNA signatures in human cancers. Nat Rev Cancer 6:857–866.

Chen K, Rajewsky N. 2006. Deep conservation of microRNA-target relationships and 3′UTR motifs in vertebrates, flies, and nematodes. Cold Spring Harbor Symp Quant Biol 71:149–156.

Croce CM. 2009. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet 10:704–714.

Esquela-Kerscher A, Slack FJ. 2006. Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer 6:259–269.

Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa EMF, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S. 2015. The consensus molecular subtypes of colorectal cancer. Nat Med 21:1350–1356.

Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, Yatabe Y, Kawahara K, Sekido Y, Takahashi T. 2005. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. Cancer Res 65:9628–9632.

Hezova R, Kovarikova A, Bienertova-Vasku J, Sachlova M, Redova M, Vasku A, Svoboda M, Radova L, Kiss I, Vyzula R, Slaby O. 2012. Evaluation of SNPs in miR-196-a2, miR-27a and miR-146a as risk factors of colorectal cancer. World J Gastroenterol 18:2827–2831.

Kupcinskas J, Bruzaite I, Juzenas S, Gyvyte U, Jonaitis L, Kiudelis G, Skieceviciene J, Leja M, Pauzas H, Tamelis A, Pavalkis D, Kupcinskas L. 2014. Lack of association between miR-27a, miR-146a, miR-196a-2, miR-492 and miR-608 gene polymorphisms and colorectal cancer. Sci Rep 4:5993.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854.

Lindor NM, Burgart LJ, Leontovich O, Goldberg RM, Cunningham JM, Sargent DJ, Walsh-Vockley C, Petersen GM, Walsh MD, Leggett BA, Young JP, Barker MA, Jass JR, Hopper J, Gallinger S, Bapat B, Redston M, Thibodeau SN. 2002. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. J Clin Oncol 20:1043–1048.

Lindor NM, Petersen GM, Hadley DW, Kinney AY, Miesfeldt S, Lu KH, Lynch P, Burke W, Press N. 2006. Recommendations for the care of individuals with an inherited predisposition to Lynch Syndrome: A systematic review. J Am Med Assoc 296: 1507–1517.

Medina PP, Nolde M, Slack FJ. 2010. OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. Nature 467:86–90.

Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R, Hall D, Hopper JL, Jass J, Le Marchand L, Limburg P, Lindor N, Potter JD, Templeton AS, Thibodeau S, Seminara D. 2007. Colon Cancer Family Registry: An international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev 16:2331–2343.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet 2:e190.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909.

Salzman DW, Weidhaas JB. 2013. SNPing cancer in the bud: microRNA and microRNA-target site polymorphisms as diagnostic and prognostic biomarkers in cancer. Pharmacol Therap 137:55–63.

Schaid DJ, McDonnell SK, Sinnwell JP, Thibodeau SN. 2013. Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. Genet Epidemiol 37:409–418.

Srivastava K, Srivastava A. 2012. Comprehensive review of genetic association studies and meta-analyses on miRNA polymorphisms and cancer risk. PLoS One 7:e50966.

Storey JD, Bass A, Dabney A, Robinson D. 2015. qvalue: Q-Value Estimation for False Discovery Rate Control. R Package, Version 2.3.0 ed.

Vinci S, Gelmini S, Mancini I, Malentacchi F, Pazzagli M, Beltrami C, Pinzani P, Orlando C. 2013. Genetic and epigenetic factors in regulation of microRNA in colorectal cancers. Methods 59:138–146.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 75:855–862.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28: 3326–3328.

Author/s:
Lindor, NM; Larson, MC; DeRycke, MS; McDonnell, SK; Baheti, S; Fogarty, ZC; Win, AK; Potter, JD; Buchanan, DD; Clendenning, M; Newcomb, PA; Casey, G; Gallinger, S; Le Marchand, L; Hopper, JL; Jenkins, MA; Goode, EL; Thibodeau, SN

Title:
Germline miRNA DNA Variants and the Risk of Colorectal Cancer by Subtype

Date:
2017-03-01

Citation:
Lindor, N. M., Larson, M. C., DeRycke, M. S., McDonnell, S. K., Baheti, S., Fogarty, Z. C., Win, A. K., Potter, J. D., Buchanan, D. D., Clendenning, M., Newcomb, P. A., Casey, G., Gallinger, S., Le Marchand, L., Hopper, J. L., Jenkins, M. A., Goode, E. L. & Thibodeau, S. N. (2017). Germline miRNA DNA Variants and the Risk of Colorectal Cancer by Subtype. GENES CHROMOSOMES & CANCER, 56 (3), pp.177-184. https://doi.org/10.1002/gcc.22420.