MLH1 –93G>A Promoter Polymorphism and the Risk of Microsatellite-Unstable Colorectal Cancer

Stavroula Raptis, Miralem Mrkonjic, Roger C. Green, Vaijayanti V. Pethe, Neerav Monga, Yuen Man Chan, Darshana Daftary, Elizabeth Dicks, Banfield H. Younghusband, Patrick S. Parfrey, Steven S. Gallinger, John R. McLaughlin, Julia A. Knight, Bharati Bapat

Background Although up to 30% of patients with colorectal cancer have a positive family history of colorectal neoplasia, few colorectal cancers can be explained by mutations in high-penetrance genes. We investigated whether polymorphisms in DNA mismatch repair genes are associated with the risk of colorectal cancer.

Methods We genotyped 929 case patients and 1098 control subjects from Ontario and 430 case patients and 275 control subjects from Newfoundland and Labrador for five polymorphisms in the mismatch repair genes MLH1 and MSH2 with the fluorogenic 5′ nuclease assay. Tumor microsatellite instability (MSI) was determined with a polymerase chain reaction–based method; MSI status was assigned as high (MSI-H, ≥30% unstable markers among all markers tested), low (MSI-L, <30% markers unstable), or stable (MSS, no unstable markers). We used unconditional logistic regression to evaluate the association between each polymorphism and colorectal cancer after adjusting for age and sex. The associations between polymorphisms and tumor clinicopathologic features were evaluated with a Pearson's chi-square or Fisher's exact test. All statistical tests were two-sided.

Results We observed strong associations between the MLH1 –93G>A polymorphism and MSI-H tumors among case patients from Ontario ($P = .001$) and Newfoundland ($P = .003$). When compared with the control populations, homozygosity for the MLH1 –93G>A variant allele was associated with MSI-H tumors among case patients in Ontario (adjusted odds ratio [OR] = 3.23, 95% confidence interval [CI] = 1.65 to 6.30) and in Newfoundland (OR = 8.88, 95% CI = 2.33 to 33.9), as was heterozygosity among case patients in Ontario (OR = 1.84, 95% CI = 1.20 to 2.83) and in Newfoundland (OR = 2.56, 95% CI = 1.14 to 5.75). Genotype frequencies were similar among case patients with MSS and MSI-L tumors and control subjects, and the majority of homozygous variant carriers had MSS tumors. Among case patients from Ontario, an association between the MLH1 –93G>A polymorphism and a strong family history of colorectal cancer (for Amsterdam criteria I and II, $P = .004$ and $P = .02$, respectively) was observed.

Conclusion In two patient populations, the MLH1 –93G>A polymorphism was associated with an increased risk of MSI-H colorectal cancer.

J Natl Cancer Inst 2007;99:463–74

Colorectal cancer is the second leading cause of cancer-related deaths in North America, affecting approximately one in 20 persons in the general population (1,2). Most colorectal cancers progress through mechanisms that involve two distinct molecular pathways—the suppressor pathway, which is characterized by frequent mutations in oncogenes (e.g., KRAS) and tumor suppressor genes (e.g., APC, DCC, TP53), and the mutator pathway, which is characterized by defects in DNA mismatch repair genes (e.g., MLH1, MSH2, MSH6, and PMS2) (3–5). Germline mutations in mismatch repair genes are responsible for Lynch syndrome, or hereditary nonpolyposis colorectal cancer (HNPCC), a colorectal cancer susceptibility syndrome that accounts for 2%–3% of all colorectal cancers (6,7). The loss of normal mismatch repair function leads to a mutator phenotype that is characterized by alterations in the length of microsatellite sequences and results in

Affiliations of authors: Departments of Pathology and Laboratory Medicine (SR, MM, VVP, BB) and Surgery (SSG), Samuel Lunenfeld Research Institute (SR, MM, VVP, SSG, JRM, JAK, BB), and Prosserman Centre for Health Research (JRM, JAK), Mount Sinai Hospital, Toronto, ON, Canada; Departments of Laboratory Medicine and Pathobiology (SR, MM, YMC, BB), Surgery (SSG), and Public Health Sciences (JRM, JAK), University of Toronto, Toronto, ON, Canada; Departments of Genetics (RCG, BHY) and Clinical Epidemiology (PSP), and Faculty of Medicine (ED), Memorial University, St John's, NL, Canada; Ontario Familial Colorectal Cancer Registry, Cancer Care Ontario, Toronto, ON, Canada (NM, DD).

Correspondence to: Bharati Bapat, PhD, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 60 Murray St, L6-304B, Box 30, Toronto, ON, Canada M5T 3L9 (e-mail: bapat@mshri.on.ca).

See "Notes" following "References."

DOI: 10.1093/jnci/djk095

CONTEXT AND CAVEATS

Prior knowledge

Approximately 30% of patients with colorectal cancer have a positive family history of this disease, but few colorectal cancers can be explained by mutations in high-penetrance genes.

Study design

Population-based case–control study in two different populations.

Contribution

In two populations, a polymorphism in a mismatch repair gene, MLH1, was shown to be associated with an increased risk of colorectal cancer with high microsatellite instability.

Implications

A common polymorphism in the MLH1 gene may account for an increased risk of colorectal cancer in some patients.

Limitations

A potential source of bias was self-reported ethnicity. The authors tried to minimize this bias by excluding subjects who did not report ethnicity or reported non-white ethnicity.

genomic instability. Although as many as 30% of all patients with colorectal cancer have a family history of the disease, only a few of these colorectal cancers are caused by mutations in known susceptibility genes, such as a mismatch repair gene or APC, so that many colorectal cancers appear to be sporadic (8). Some of this familial risk may be attributed to high-penetrance mutations in genes that have not yet been identified, and some may result from common variant alleles with low to moderate penetrance of key candidate genes already associated with colorectal cancer (9,10).

MLH1 and MSH2 are the key components of the mismatch repair system, which participates in the recognition of nucleotide mismatches occurring during DNA replication and in the recruitment of additional mismatch repair proteins to the site to correct the replication error (4–6). In addition to germline mutations that have been identified in the MLH1 and MSH2 genes, numerous polymorphisms have also been identified; however, their functional contributions are currently unknown. Five single-nucleotide polymorphisms (SNPs) are of particular interest because of their prevalence and potential to affect mismatch repair functions. These SNPs are located in the MLH1 gene (i.e., –93G>A promoter SNP, I219V coding SNP, and intronic IVS14-19A>G SNP located 19 nucleotides upstream from the exon 15 splice acceptor site) and in the MSH2 gene (i.e., G322D coding SNP and intronic IVS12-6T>C SNP located six nucleotides upstream from the exon 13 splice acceptor site). The MLH1 –93G>A polymorphism is located in the core promoter region, 93 nucleotides upstream of the transcription start site in potential transcription factor binding sites (11). MLH1 –93G>A has been associated with several cancers, including lung and breast cancers (12,13), and with increased risks of hyperplastic polyps and colonic adenomas in long-term smokers (14). The MLH1 I219V polymorphism, which is located in exon 8 at nucleotide position 655 (with A and G alleles, A>G), was shown in a multilocus analysis with other metabolic SNPs to be associated with an increased risk for childhood acute lymphoblastic leukemia (15). A more recent evaluation of SNPs (16) found a statistically significant association between the I219V homozygous variant (GG) and an increased risk of breast cancer (odds ratio [OR] = 2.90, 95% confidence interval [CI] = 1.02 to 8.24). Hutter et al. (17) showed that the MLH1 I219V polymorphism is linked with another downstream SNP, IVS14-19A>G. In a European HNPCC population, the IVS14-19 variant G allele is overrepresented on chromosomes bearing a germline MLH1 mutation, and there was a strong association of the variant G allele and MLH1 substitution and deletion mutations (18).

The G322D polymorphism is located in the coding region of the MSH2 gene, in exon 6, at nucleotide position 965, with G and A alleles (G>A). G322D is a rare polymorphism compared with the other four polymorphisms (19–21). Functional studies of its yeast homologue, G317D, indicate that this alteration results in a modest decrease in mismatch repair efficiency (22,23). However, segregation studies in confirmed HNPCC or HNPCC-like families have provided no consistent evidence of its association with colorectal cancer among affected family members (21). The MSH2 IVS12-6T>C polymorphism is located six nucleotides upstream of the exon 13 splice acceptor site. An association between this intronic SNP and non-Hodgkin lymphoma has been observed; the C allele occurred at a frequency of 11.4% in cancer patients and 5.0% in the normal control subjects (24,25). The MSH2 IVS12-6T>C polymorphism may also be associated with a predisposition to cancer in patients with ulcerative colitis; the risk of developing colorectal cancer was three times higher for patients with ulcerative colitis who carried the variant C allele than for those who carried the wild-type T allele (26).

In this case–control study of two populations, one in Ontario and the other in Newfoundland and Labrador, we investigated whether any of these five polymorphisms in mismatch repair genes constitute low-penetrance alleles and contribute to colorectal cancer susceptibility by evaluating the association between each SNP and colorectal cancer risk. Because low-penetrance alleles in candidate genes may not only be associated with cancer incidence but may also influence cancer phenotype and prognosis, we also evaluated their association with clinical and pathologic tumor characteristics among case patients and control subjects.

Materials and Methods

Single-Nucleotide Polymorphism Selection Criteria

The five polymorphisms analyzed in this study were selected on the basis of extensive database and literature searches. The databases that we surveyed included the following: Ensembl (http://www.ebi.ac.uk/ensembl/), Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php), Human Genome Variation Database (http://hgvbase.cgb.ki.se/), International Society for Gastrointestinal Hereditary Tumors (http://www.insight-group.org/), National Center for Biotechnology Information SNP Database (http://www.ncbi.nlm.nih.gov/SNP/), and Swiss-Prot (http://www.ebi.ac.uk/swissprot/). We selected validated SNPs with a minor allele frequency of greater than 1% that had multiple independent submissions to the SNP databases and/or multiple citations in the literature, that were confirmed by frequency or genotype data in which all alleles had been observed in at least two chromosomes, and that were located in putative functional domains and regulatory regions.

Study Subjects

We conducted this case–control study with subjects from two different populations—those from the province of Ontario, and those from the province of Newfoundland and Labrador (hereafter referred to as Newfoundland). Case patients accrued from the Ontario and Newfoundland populations were stratified by family risk according to Amsterdam criteria I and II (27,28). The Amsterdam I criteria include case patients with at least three family members who have been diagnosed with colorectal cancer in two successive generations, with one affected family member being a first-degree relative of the other two, and with at least one of the three being younger than age 50 years at diagnosis (27). The Amsterdam II criteria include the Amsterdam I criteria but are less stringent because they take into account HNPCC-associated tumors, such as those arising in the endometrium, small bowel, ureter, or renal pelvis (28). Additionally, case patients were also stratified by family risk as described by Cotterchio et al. (29). Briefly, case patients were classified as having a high familial risk if they met Amsterdam I criteria (27) and did not have familial adenomatous polyposis. Case patients were classified as having intermediate familial or other pathologic risk if they met any of the following criteria: case patient had two relatives with any of the HNPCC-associated tumors, with one affected family member being a first-degree relative of at least one of the other two; case patient had a family member with HNPCC diagnosed at age 35 years or younger; case patient was younger than age 50 years and had one first or second-degree relative diagnosed with colorectal cancer under the age of 50; case patient was age 35 years or younger irrespective of family history; case patient had multiple primary colorectal tumors; case patient had other primary HNPCC-associated tumors; or case patient had one of the following—multiple polyps, Peutz–Jeghers disease, hamartomatous polyp, juvenile polyp, inflammatory bowel disease, or any unusual colorectal cancer histologies (such as carcinosarcoma, adenosquamous, spindle cell, metaplastic, choriocarcinoma, signet ring, undifferentiated, trophoblastic differentiation, small-cell neuroendocrine carcinoma); or case patient was of Ashkenazi Jewish ancestry. All other case patients were classified as having a low or sporadic risk.

Case patients and control subjects from the Ontario population were obtained from the Ontario Familial Colorectal Cancer Registry, which is part of a U.S. National Cancer Institute (NCI)–supported consortium, the Colon Cancer Family Registry (http://epi.grants.cancer.gov/CFR/). Living residents of Ontario with pathology-confirmed colorectal cancer, aged 20–74 years, and diagnosed between July 1, 1997, and June 30, 2000, were identified and recruited for the Ontario Familial Colorectal Cancer Registry from the population-based Ontario Cancer Registry, as described by Cotterchio et al. (29). Family history information was collected by a mailed questionnaire and was used to construct pedigrees that would be used to classify the patient by his or her family risk. A total of 3776 patients with colorectal cancer were identified in Ontario; after we obtained their physicians' approval, the patients were asked to complete and return the family history questionnaire. The familial risk for each patient was determined by use of the Amsterdam criteria (27,28) and other risk criteria described above (i.e., high, intermediate or other family risk, and low or sporadic risk) (29). Among the 3776 case patients contacted, 1593 were willing and able to participate in the registry. Of the 1593 case patients, 1103 had a blood sample available, from whom we identified 1004 case patients for this study by restricting inclusion to those with an adequate blood sample available and with colorectal cancer indicated as the primary tumor site. Because the majority of case patients and control subjects with specified ethnicity in Ontario (92.5% and 86.9%, respectively) and in Newfoundland (98.6% and 96.8%, respectively) are of white European or Caucasian ethnicity, we decided to exclude from the analyses those who were non-white and those who did not report ethnicity to minimize the potential for population stratification. Of 1004 case patients in Ontario, 929 were white and were successfully genotyped for all five SNPs, and thus, they constitute the case patients from Ontario. All information was collected from three mailed questionnaires (family, personal, and diet questionnaires) and phone and in-person interviews, as well as blood and tissue specimens that were obtained after informed written consent was provided to participate in the Ontario Familial Colorectal Cancer Registry, as described in protocols approved by the research ethics boards of Mount Sinai Hospital and the University of Toronto. No case patients with familial adenomatous polyposis were included in the registry.

In Ontario, population control subjects who had not been diagnosed with colorectal cancer were accrued by randomly selected residential telephone numbers during the years 1999 and 2000, and for the years 2002 and 2003 by use of population-based Tax Assessment Rolls of the provincial government (29,30). The Ministry of Finance Property Assessment Database (year 2000) was used to identify age- (5-year groups) and sex-matched controls, and thus, control subjects were frequency matched to case patients by sex and 5-year age group. Permission to use this file for recruiting control subjects was granted to an investigator in this study (J. R. McLaughlin, a coinvestigator with the Ontario Familial Colorectal Cancer Registry). The subjects were sent a letter of invitation with a reply form to assess eligibility (no colorectal cancer and age- and sex-matched to case patients). Nonresponders received a follow-up telephone call. Another letter was sent if they did not respond within 7 weeks. A total of 2736 control subjects from Ontario agreed to participate in the study, with 1957 completing all three questionnaires (family, personal, and diet questionnaires). Of the 1957, 1314 control subjects provided blood samples, and 1098 of them were white. These 1098 control subjects were successfully genotyped for all five SNPs and thus constituted the control subjects from Ontario in this study. The remaining case patients and control subjects did not complete one or more questionnaires, did not have an adequate blood sample, or were not white.

The accrual pattern followed by the Newfoundland Familial Colorectal Cancer Registry was similar to that followed by the Ontario Familial Colorectal Cancer Registry. Case patients with colorectal cancer who were younger than age 75 years and diagnosed between January 1, 1999, and December 31, 2003, were identified through the Newfoundland tumor registry. No additional sampling was done in Newfoundland. We identified 1175 case patients with colorectal cancer and obtained consent from physicians to contact 1144 of them. Of the 1144 patients, 747 responded to the family history questionnaire, and of those, 555 provided blood samples. We were able to obtain 504 case

patients from the Newfoundland Familial Colorectal Cancer Registry, of whom 430 provided ethnicity information and were classified as white. The recruitment of population control subjects who had not been diagnosed with colorectal cancer in Newfoundland was accomplished through random digit dialing, and the control subjects were again matched to case patients by sex and 5-year age group. We identified 1602 control subjects from Newfoundland who agreed to participate in the study, of whom 703 completed all three questionnaires. Of the 703, 530 control subjects provided blood samples. The remaining case patients and control subjects did not complete one or more questionnaires, did not have an adequate blood sample, or were not white. Because the accrual of the Newfoundland population control subjects is being performed on an ongoing basis, at the time of this analysis, the 275 white control subjects who had provided an adequate blood sample and completed all three questionnaires were included in this study.

The mean age for control subjects from both provinces was calculated from the date of completion of the family history questionnaire, and that for case patients was calculated from the age at diagnosis. Case patients accrued from the Ontario and Newfoundland populations were stratified according to Amsterdam criteria I and II (27,28) as described above. We collected data on tumor microsatellite instability (MSI) status, tumor location, tumor stage, and tumor grade, when available, through review of pathologic and/or surgical reports. Tumors were staged and graded according to the methodology of the American Joint Committee on Cancer (31).

Molecular Genetic Analysis

Single-nucleotide polymorphism genotyping with the fluorogenic 5′ nuclease assay. Peripheral blood lymphocytes were isolated from whole blood by use of Ficoll–Paque gradient centrifugation according to the manufacturer's protocol (Amersham Biosciences, Baie d'Urfé, Quebec, Canada). Phenol–chloroform or the Qiagen DNA extraction kit (Qiagen Inc, Montgomery County, Maryland) was used to extract genomic DNA from lymphocytes. The fluorogenic 5′ nuclease polymerase chain reaction (PCR) assay or the TaqMan assay (32) was used to genotype each of the following five SNPs: MLH1 –93G>A, I219V, IVS14-19A>G, G322D, and IVS12-6T>C. Primers and probes were designed with Primer Express version 2.0 software (Applied Biosystems, Foster City, CA). Primers were tested for successful PCR amplification of the 119 (G322D), 80 (I219V), 70 (MLH1 –93G>A), 85 (IVS14-19A>G), and 105 (IVS12-6T>C) base-pair products on the basis of the visualization of a robust amplicon product on an agarose gel (33,34). Sequences of primers and probes are listed in Supplementary Table 1 (available online). The master reaction mixtures for the polymorphisms G322D, I219V, and MLH1 –93G>A were prepared with reagents in the TaqMan 1000 Rxn PCR Core Reagents kit (Applied Biosystems). For G322D, the master reaction mixture contained (final concentrations): 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 2.5 mM $MgCl_2$, 300 nM forward primer, 300 nM reverse primer, 100 nM wild-type FAM-labeled probe, 350 nM variant VIC-labeled probe, AmpErase uracil-*N*-glycosylase (0.01 U/μL; Applied Biosystems), and AmpliTaq Gold DNA polymerase (0.025 U/μL; Applied Biosystems). The PCR conditions for G322D were as follows: 2 minutes at 50 °C, 10 minutes at 95 °C (AmpliTaq Gold activation), 15 seconds at 92 °C (DNA denaturation), and 1 minute at 58.5 °C (primer–probe annealing and primer extension). The denaturation–annealing–extension sequence (15 seconds at 92 °C and 1 minute at 58.5 °C) was repeated for 39 cycles. For the I219V polymorphism, all components of the master reaction mixture were used in the final concentrations that were used for G322D, except for the following modifications: 3.0 mM $MgCl_2$, 75 nM FAM-labeled probe, and 150 nM variant TET-labeled probe. The PCR conditions for I219V were also similar to those used for G322D, except that the annealing temperature for I219V was 57 °C for 1 minute. The G322D and I219V SNP assays used 49 μL of the master reaction mixture and 1 μL of DNA template (at 2–7.5 ng/μL).

For the MLH1 93G>A polymorphism, the master reaction mixture contained (final concentrations): 96 μM dATP, 96 μM dCTP, 96 μM cGTP, 96 μM dUTP, 3.5 mM $MgCl_2$, 300 nM forward primer, 300 nM reverse primer, 15 nM wild-type FAM-labeled probe, 100 nM variant TET-labeled probe, and AmpliTaq Gold DNA polymerase (0.025 U/μL). The conditions used were identical to those used for G322D, except the annealing temperature was 62 °C. The MLH1 93G>A SNP assay used 29 μL of the master reaction mixture and 1 μL of DNA template (at 2–7.5 ng/μL).

The MLH1 IVS14-19A>G and MSH2 IVS12-6T>C polymorphisms were genotyped by use of the Eurogentec qtPCR kit (Eurogentec, San Diego, CA). The master reaction mixture contained (final concentrations): 0.72 mM primers and 0.16 mM probes (both VIC and FAM labeled). Assays for both of these polymorphisms used the conditions described for G322D. The IVS14-19A>G and IVS12-6T>C SNP assays used 12.5 μL of the master reaction mixture and 1 μL of DNA template (at 2–7.5 ng/μL).

All SNP genotyping assays were conducted in 96-well polypropylene plates (Axygen Scientific, Union City, CA), and the results were analyzed with the Applied Biosystems 7900HT Sequence Detection system and the accompanying software—SDS versions 2.0 and/or 2.1 (Applied Biosystems). Independent quality control for genotyping was done on 5%–10% of samples with restriction fragment length polymorphism (35) and/or sequencing (36,37).

Tumor microsatellite instability analysis. Tumor microsatellite instability (MSI) analysis was performed as described previously (38). Briefly, paraffin-embedded colorectal tumor tissue from case patients with incident cases of colorectal cancer, and also paraffin-embedded normal colorectal tissue from the same patients were microdissected in areas with more than 70% cellularity in tumor and normal cell populations. MSI analysis was carried out with five or more microsatellite markers from the panel of 10 microsatellite markers, as recommended by the NCI; these markers consist of the mononucleotides BAT25, BAT26, BAT40, and BAT34C4; the dinucleotides D2S123, D5S346, ACTC, D18S55, and D10S197; and one penta-mono-tetra compound marker, MYC-L (39). MSI was indicated by the presence of altered or additional bands of the PCR-amplified product from the tumor tissue, compared with the bands from matched normal colon tissue. MSI status was assigned as MSI high (MSI-H, ≥30% unstable markers among all markers tested), MSI low (MSI-L, <30% markers unstable), or microsatellite stable (MSS, no unstable markers) as described by the

NCI-recommended guidelines for MSI testing (39). For the analysis, MSI-L and MSS groups were combined into one group (hereafter referred to as "MSS/L") to distinguish them from the MSI-H group, who were mismatch repair–deficient. Primers were obtained from Applied Biosystems, and primer sequences are listed in Supplemental Table 2 (available online).

PCRs for MSI analysis were prepared with reagents in the Platinum Taq Kit (Invitrogen, Burlington, ON, Canada) and in the Gold Taq Kit (Applied Biosystems). For BAT25 and D5S346, the master mixture contained (final concentrations) 2.0 mM $MgCl_2$, all four deoxyribonucleotide triphosphates (each at 0.4 mM), forward primer (2 ng/μL), reverse primer (2 ng/μL), 2 U of Platinum Taq Polymerase, and 1× PCR buffer. The master reaction mixtures for BAT26, ACTC, and D2S123 were identical to that for BAT25 and D5S346, except that 1 U Gold Taq polymerase was used. The master reaction mixtures for D18S55, BAT40, and MYC-L were also identical to that for BAT25 and D5S346, except that 2.25 mM $MgCl_2$ was used. The master reaction mixture for BAT34C4 also used 2.25 mM $MgCl_2$ but with 1 U of Gold Taq polymerase, and the master reaction mixture for D10S197 used 1.0 mM $MgCl_2$ and 1 U of Gold Taq polymerase. All assays used 20 μL of the master reaction mixture and 2 μL of DNA template (at 25 ng/μL). The PCR conditions for all microsatellite sequences were as follows: 5 minutes at 94 °C (Taq activation), 30 seconds at 94 °C (DNA denaturation), 30 seconds at 55 °C (primer annealing), and 30 seconds at 72 °C (primer extension). The denaturation–annealing–extension sequence was repeated for 35 cycles, followed by 10 minutes at 72 °C. Upon completion of the PCRs, the products of individual DNA samples were pooled into two groups as follows: group 1 contained 15% of the ACTC, 11.3% D5S346, 23.7% BAT26, 30% D2S123, and 20% BAT25, group 2 contained 10% D18S55, 15% BAT34C4, 20% D10S197, 35% BAT40, and 20% MYC-L. Once the pooled PCR products were mixed by brief vortexing, 1.5 μL of pooled PCR products was mixed with 8 μL of Hi-Di formamide (Applied Biosystems) and 0.5 μL of ROX standard (i.e., a passive reference dye) (Applied Biosystems) and loaded onto 96-well polypropylene plates. All MSI assays were conducted in 96-well polypropylene plates (Axygen Scientific), and the results were analyzed with the Applied Biosystems 3130xl DNA Analyzer system and the accompanying software—GeneMapper version 3.7 for Microsatellite Instability (Applied Biosystems).

Statistical Analysis

The associations of the variant alleles with colorectal cancer incidence, age at onset, MSI status, tumor location, tumor grade, tumor stage, and family risk status (Amsterdam criteria I and II) were evaluated with a two-sided Pearson's chi-square or Fisher's exact test, in which a P value of less than .05 was considered to be statistically significant. Unconditional logistic regression was also used to evaluate the association between each SNP and colorectal cancer, after adjusting for age and sex. The Cochran–Armitage Trend Test was performed to examine the trend between MLH1 –93 variant allele carriers (of zero, one, or two copies) and case patients with MSI-H tumors. Chi-square and Fisher's exact tests, the trend test, and logistic regression were performed with SAS version 9.0 (SAS Institute, Cary, NC). All statistical tests were two-sided, and the results were adjusted by use of the Bonferroni correction method for multiple comparisons.

Before the initiation of this study, we had performed power calculations. It was estimated that 600 case patients and 600 control subjects would be made available from the Ontario and Newfoundland registries and that the rarest allele frequency of selected polymorphisms was 5% (rarest carrier frequency of 10%), if we assumed Hardy–Weinberg equilibrium. From 600 case patients and 600 control subjects, we had 80% power to detect an odds ratio of 1.9 [if we assumed a correlation coefficient of 0.2 between case patients and control subjects as described by Dupont and Plummer (40)] with a two-sided alpha of 0.01. This more conservative value of alpha was used to reduce the likelihood of a chance positive result among the approximately 20 comparisons that were made in each population.

Results

Populations of Case Patients With Colorectal Cancer and Control Subjects

We genotyped a total of 929 case patients and 1098 control subjects for all five SNPs in Ontario, and a total of 430 case patients and 275 control subjects in Newfoundland. The ages (mean ± standard deviation) at diagnosis for case patients in Ontario and Newfoundland were 59.8 ± 9 years and 60.4 ± 9 years, respectively. The age at diagnosis, family history, and histopathologic feature distributions for all case patients enrolled in this study from both provinces are shown in Table 1. Overall, no differences were observed in the distribution of characteristics between the case patients from the two provinces. The ages (mean ± standard deviation) of the control subjects from Ontario and Newfoundland were 63.7 ± 9 years and 60.5 ± 9 years, respectively. No differences in age or sex distributions between case patients and control subjects were observed in either province.

Distribution of Genotypes and Alleles

The variant allele frequencies of the selected SNPs in the general populations and in the case patient populations of Ontario and Newfoundland were not known. Whether there were any differences in variant allele frequencies for selected polymorphisms between the general populations of Ontario and Newfoundland was also unknown. The variant allele frequencies for the SNPs examined are shown in Table 2. All SNPs examined were in Hardy–Weinberg equilibrium among control populations of both provinces. Overall, the variant allele frequencies were similar between case patients and control subjects in both Ontario and Newfoundland. We observed no statistically significant differences in allele frequencies between the general populations (represented by the control subjects) of Ontario and Newfoundland (data not shown).

The distribution of the genotypes for all five SNPs—MLH1 –93G>A, MLH1 I219V, MLH1 IVS14-19G>A, MSH2 G322D, and MSH2 IVS12-6T>C—among control subjects and case patients in Ontario and Newfoundland are shown in Table 3. We found no differences in genotype frequency distribution between all case patients and all control subjects within each province. When case patients were stratified by MSI status (i.e., MSI-H and

Table 1. Distribution of age, family history, and clinicopathologic features among case patients with colorectal cancer from Ontario and Newfoundland*

Feature	Case patients, No. (%)	
	Ontario	Newfoundland
No. of white case patients genotyped	929 (100)	430 (100)
Age at diagnosis, y		
<50	126 (13.6)	46 (10.7)
≥50	802 (86.3)	384 (89.3)
Unavailable	1 (0.1)	0 (0)
Family history†		
Amsterdam I only	42 (4.5)	14 (3.3)
Amsterdam II only	15 (1.6)	0 (0)
Non-Amsterdam	872 (93.9)	413 (96.0)
Unavailable	0 (0)	3 (0.7)
Histologic grade		
1	80 (8.6)	55 (12.8)
2	568 (61.1)	309 (71.9)
3	92 (9.9)	32 (7.4)
Unavailable	189 (20.3)	34 (7.9)
TNM stage‡		
1	178 (19.2)	NA
2	302 (32.5)	NA
3	247 (26.6)	NA
4	47 (5.0)	NA
Unknown	155 (16.7)	NA
MSI		
MSI-H	117 (12.6)	33 (7.7)
MSS/MSI-L	649 (69.9)	261 (60.7)
Unavailable	163 (17.5)	136 (31.6)
Tumor location		
Proximal	364 (39.2)	154 (35.8)
Distal	554 (59.6)	259 (60.2)
Other	10 (1.1)	16 (3.7)
Unavailable	1 (0.1)	1 (0.2)

* TNM = tumor–node–metastasis; NA = not available; MSI = microsatellite instability; MSI-H = high-frequency microsatellite instability; MSS/L = microsatellite stable or low-frequency microsatellite instability.

† Family history criteria are as described previously (27,28).

‡ Available only for Ontario.

MSS/L) and compared with control subjects (Table 4), we observed a statistically significant association between the MLH1 –93G>A polymorphism and MSI-H tumors in both the Ontario and Newfoundland populations (for Ontario heterozygotes, OR = 1.84, 95% CI = 1.20 to 2.83; for Ontario homozygotes, OR = 3.23, 95% CI = 1.65 to 6.30; for Newfoundland heterozygotes, OR = 2.56, 95% CI = 1.14 to 5.75; and for Newfoundland homozygotes, OR = 8.88, 95% CI = 2.33 to 33.9). However, genotype frequencies were similar among case patients with MSS/L tumors and control subjects and thus not statistically significant (Table 4). We also observed a statistically significantly increasing trend for the association between the number of MLH1 –93 variant A alleles carried (zero, one, or two alleles) and MSI-H status in both the Ontario and Newfoundland populations (both P_{trend}<.001) (Table 4).

Because low-penetrance alleles may not only be associated with cancer incidence but also may influence cancer phenotype and prognosis, we examined associations between available clinicopathologic tumor features among case patients and the variant alleles for each of the five SNPs. For the MLH1 polymorphisms I219V and IVS14-19A>G and the MSH2 polymorphism G322D, we found no association between any clinicopathologic characteristic and the variant alleles of these three SNPs in case patients from either Ontario or Newfoundland (Supplementary Tables 3–10, available online). In addition, no association was found between clinicopathologic features and the MSH2 IVS12-6T>C polymorphism, except for tumors with MSI status, in which the IVS12-6T>C polymorphism was associated with MSI-H tumors among case patients from Ontario (P = .04; Supplementary Table 5, available online) but not with MSI-H tumors among case patients from Newfoundland (Supplementary Table 6, available online). This association between MSH2 IVS12-6T>C and MSI-H tumors among case patients from Ontario did not remain statistically significant after Bonferroni correction for multiple comparisons was applied.

Analysis of MSI status and the genotype distribution of the MLH1 –93G>A promoter SNP among case patients from Ontario and Newfoundland found strong associations (P = .001 and P = .003, respectively) between the variant allele and MSI status (Tables 5 and 6), with a higher proportion of the MLH1 –93G>A variant allele carriers having MSI-H tumors than MSS/L tumors. Both of these results remained statistically significant after Bonferroni correction for multiple comparisons. Analysis of Ontario case patients with a strong family history,

Table 2. Allele frequencies of each polymorphism in participants in Ontario and Newfoundland*

Gene	SNP	Population	Variant allele frequencies, %		*P* value†
			Case patients	Control subjects	
MLH1	–93G>A	Ontario	22.6	21.4	.22
		Newfoundland	22.4	19.3	.17
	I219V	Ontario	30.4	31.3	.73
		Newfoundland	29.0	32.0	.39
	IVS14-19A>G	Ontario	43.2	43.2	.95
		Newfoundland	42.2	45.3	.76
MSH2	G322D	Ontario	1.7	1.3	.21
		Newfoundland	1.3	1.1	.52
	IVS12-6T>C	Ontario	9.3	10.7	.22
		Newfoundland	12.3	12.9	.93

* Newfoundland = Newfoundland and Labrador; SNP = single-nucleotide polymorphism.

† Pearson's chi-square test used. All statistical tests were two-sided. A *P* value of less than .01 was considered to be statistically significant after adjusting for multiple comparisons by use of the Bonferroni method of correction.

Table 3. Association of MLH1 and MSH2 polymorphisms with risk of colorectal cancer in Ontario and Newfoundland*

Population, SNP, and genotype	Case patients, No. (%)	Control subjects, No. (%)	OR (95% CI)
Ontario			
MLH1			
–93G>A			
GG	554 (59.6)	687 (62.6)	1.00 (referent)
GA	331 (35.6)	352 (32.1)	1.19 (0.98 to 1.45)
AA	44 (4.7)	59 (5.4)	0.89 (0.59 to 1.35)
I219V			
AA	451 (48.5)	514 (46.8)	1.00 (referent)
AG	391 (42.1)	485 (44.2)	0.93 (0.77 to 1.13)
GG	87 (9.4)	99 (9.0)	1.05 (0.75 to 1.45)
IVS14-19A>G			
AA	297 (32.0)	355 (32.3)	1.00 (referent)
AG	462 (49.7)	538 (49.0)	1.02 (0.83 to 1.25)
GG	170 (18.3)	205 (18.7)	1.02 (0.78 to 1.33)
MSH2			
G322D			
GG	898 (96.7)	1069 (97.4)	1.00 (referent)
GA and AA†	31 (3.3)	29 (2.6)	1.30 (0.77 to 2.20)
IVS12-6T>C			
TT	770 (82.9)	878 (80.0)	1.00 (referent)
TC	146 (15.7)	205 (18.7)	0.81 (0.64 to 1.03)
CC	13 (1.4)	15 (1.3)	0.99 (0.46 to 2.13)
Newfoundland			
MLH1			
–93G>A			
GG	260 (60.5)	176 (64.0)	1.00 (referent)
GA	147 (34.2)	92 (33.4)	1.08 (0.78 to 1.50)
AA	23 (5.3)	7 (2.5)	2.20 (0.92 to 5.26)
I219V			
AA	213 (49.5)	128 (46.5)	1.00 (referent)
AG	184 (42.8)	118 (42.9)	0.94 (0.68 to 1.30)
GG	33 (7.7)	29 (10.6)	0.68 (0.39 to 1.17)
IVS14-19A>G			
AA	143 (33.3)	85 (30.9)	1.00 (referent)
AG	211 (49.1)	131 (47.6)	0.96 (0.67 to 1.35)
GG	76 (17.7)	59 (21.5)	0.78 (0.50 to 1.21)
MSH2			
G322D			
GG	419 (97.4)	269 (97.8)	1.00 (referent)
GA and AA†	11 (2.6)	6 (2.2)	1.17 (0.42 to 3.21)
IVS12-6T>C			
TT	332 (77.2)	211 (76.7)	1.00 (referent)
TC	90 (20.9)	57 (20.7)	1.03 (0.70 to 1.50)
CC	8 (1.9)	7 (2.6)	0.70 (0.25 to 1.97)

* OR = odds ratios adjusted for age and sex; CI = confidence interval; SNP = single-nucleotide polymorphism.

† Heterozygous (GA) and homozygous (AA) variants were collapsed because of low variant A allele frequency.

as defined by Amsterdam I (27) and/or Amsterdam II criteria (28), found a strong, statistically significant association between the MLH1 –93G>A variant A allele and family history of colorectal cancer (for Amsterdam I, $P = .004$; and for Amsterdam I and II, $P = .016$) (Table 5). The association of the MLH1 –93G>A polymorphism with disease meeting Amsterdam I criteria remained statistically significant after Bonferroni correction for multiple comparisons; however, the association of MLH1 –93G>A and disease meeting Amsterdam I and II criteria did not remain statistically significant. No association between the MLH1 –93G>A variant and family history was observed among Newfoundland case patients. However, tumors in Newfoundland case patients who carried the variant allele were more likely to be located in the proximal region of the colon than in the distal region ($P = .04$; Table 6). This association between MLH1 −93G>A and tumor location did not remain statistically significant after Bonferroni correction for multiple comparisons. Tumor location, histologic grade, and tumor–node–metastasis (TNM) stage of the tumors were not associated with the MLH1 –93G>A variant A allele among the Ontario case patients.

Discussion

This is the first large-scale case–control study, to our knowledge, to examine the allele frequencies of these five polymorphisms

Table 4. Risk of colorectal cancer by microsatellite instability status for the MLH1 –93G>A polymorphism only in Ontario and Newfoundland*

Population, tumor MSI status, and genotype	Case patients, No. (%)	OR (95% CI)
Ontario		
MSI-H		
GG	55 (48.2)	1.00 (referent)
GA	45 (39.5)	1.84 (1.20 to 2.83)
AA	14 (12.3)	3.23 (1.65 to 6.30)
P_{trend}†		<.001
MSS/L		
GG	379 (60.7)	1.00 (referent)
GA	221 (35.4)	1.17 (0.94 to 1.46)
AA	24 (3.9)	0.69 (0.42 to 1.15)
Newfoundland		
MSI-H		
GG	12 (36.4)	1.00 (referent)
GA	16 (48.5)	2.56 (1.14 to 5.75)
AA	5 (15.1)	8.88 (2.33 to 33.9)
P_{trend}†		<.001
MSS/L		
GG	166 (63.6)	1.00 (referent)
GA	83 (31.8)	0.96 (0.66 to 1.39)
AA	12 (4.6)	1.70 (0.65 to 4.51)

* The number of control subjects (percentage) from Ontario by genotype was as follows: GG = 687 (62.6); GA = 352 (32.1); and AA = 59 (5.4). The number of control subjects from Newfoundland by genotype is as follows: GG = 176 (64.0); GA = 92 (33.4); and AA = 7 (2.5). MSI = microsatellite instability; OR = odds ratio adjusted for age and sex; CI = confidence interval; MSI-H = high-frequency microsatellite instability; MSS/L = microsatellite stable or low-frequency microsatellite instability.

† Cochran–Armitage trend test was used. All statistical tests were two-sided.

in mismatch repair genes and their association with the incidence of colorectal cancer in two populations. The MLH1 –93G>A promoter polymorphism appeared to be strongly associated with MSI-H tumors in both populations. The MLH1 –93G>A variant allele, either in the homozygous or heterozygous state, was associated with a higher risk of developing MSI-H tumors than the wild-type allele. The absolute risks of colorectal cancer for heterozygous and homozygous carriers of the variant A allele in Ontario were 7.4% and 6.9%, respectively. For Newfoundland, the absolute risks of colorectal cancer were 13.1% for heterozygous carriers of the variant A allele and 12.6% for homozygous carriers. Thus, the MLH1 –93G>A polymorphism may modify colorectal tumorigenesis. Carriers of the wild-type (G) allele predominantly had MSS/L tumors, and the genotype frequencies among case patients with MSS/L tumors were very similar to those among the control subjects (Table 4). Among case patients in Ontario, a statistically significant association was found between the MLH1 −93G>A promoter variant allele and a strong family history of colorectal cancer, as defined by the Amsterdam criteria (27,28). This association was not observed in the Newfoundland case patients, perhaps because of the small number of patients meeting the Amsterdam criteria and also being carriers of the variant A allele in this study. The MLH1 –93G>A variant allele was not associated with tumor location, histologic grade, or TNM stage among Ontario case patients, but it was associated with proximally located tumors among Newfoundland case patients. This association was not statistically significant after Bonferroni correction for multiple comparisons. At the time of analysis, TNM stage was not available for Newfoundland case patients and so could not be analyzed in this population.

Because we performed many statistical tests, some statistically significant results could have occurred by chance, and so the Bonferroni correction method for multiple comparisons was applied. The major finding of this study—a strong association between the MLH1 –93G>A variant and MSI-H colorectal cancer tumors—was highly statistically significant in two independent populations even after Bonferroni adjustment and was, therefore, unlikely to have occurred by chance.

Polymorphism frequencies often vary by ethnic background (41). The frequency of the MLH1 –93G>A variant A allele in the Ontario general population was 21.4% and that in the Newfoundland general population was 19.3%; both values are considerably lower than those published in Asian populations—46% in the Japanese population and 50% in the Korean population (11–13). The differences between our white populations and those Asian populations are not surprising. Indeed, D132H, a SNP in the MLH1 gene, has been associated with colorectal cancer in an Ashkenazi Jewish population in Israel (42). However, the D132H SNP occurs at a very low frequency in North America and was not detected in a group of patients with colorectal cancer or endometrial cancer in the United States (43).

The MLH1 –93G>A promoter SNP has been associated with other cancers, including lung and breast cancers (12–14). A recent study found a statistically significant association between the variant AA genotype and an increased risk for squamous cell lung carcinoma in a Korean population (13). Another study found that the GG genotype was statistically significantly associated with an increased risk of breast cancer in Korean women (12). These differing results may reflect differences in the tissue specificity and diverse cellular functions of MLH1 (44). Recently, Yu et al. (14) reported that the MLH1 –93G>A promoter SNP was associated with an increased risk of colorectal polyps and adenomas among long-term smokers in the United States (14). The results of Yu et al. are consistent with our finding that this polymorphism has modifying effects in colorectal cancers.

Functional studies have shown that the MLH1 promoter region, from nucleotide position –184 to the transcription start site, in which the G to A alteration occurs, is essential for transcription of the MLH1 gene (11). Within this region, there are two potential binding sites for transcription factors GT-IIB (GT-motif 2B) and NF-IL6 (interleukin-6–regulated nuclear factor) (11,45,46). The G>A alteration may, therefore, affect the transcriptional activation of MLH1. This polymorphism may alternatively be in linkage disequilibrium with another coding or intronic polymorphism and/or an MLH1 mutation that reduces MLH1 function. Because the A allele of the MLH1 –93G>A polymorphism has also been found in patients with MSS tumors, we can rule out the possibility of this polymorphism being linked to a founder MLH1 mutation in our populations.

Traditionally, several different empirically developed clinical criteria, such as the Amsterdam criteria, were used to assess an individual's risk of developing colorectal tumors in HNPCC-like families. These criteria were based primarily on family history

Table 5. Genotype frequencies of MLH1 –93G>A polymorphism and clinicopathologic features of patients with colorectal cancer in Ontario*

Feature	No.	Genotype frequency, No. (%)			Chi-square†	DF	*P* value‡
		GG	GA	AA			
Total	929	554 (59.6)	331 (35.6)	44 (4.7)	NC	NC	NC
Sex							
Male	496	309 (55.8)	167 (50.5)	20 (45.5)	3.528	2	.17
Female	433	245 (44.2)	164 (49.5)	24 (54.5)			
Age at diagnosis, y§							
<50	126	82 (14.8)	38 (11.5)	6 (13.6)	1.904	2	.39
≥50	802	472 (85.2)	292 (88.5)	38 (86.4)			
Family history‖							
Amsterdam I only	42	27 (4.9)	9 (2.7)	6 (13.6)	11.119	2	.004
Non-Amsterdam I	887	527 (95.1)	322 (97.3)	38 (86.4)			
Amsterdam I & II	57	34 (6.1)	16 (4.8)	7 (15.9)	8.272	2	.016
Non-Amsterdam I or II	872	520 (93.9)	315 (95.2)	37 (84.1)			
Tumor location§							
Proximal	364	219 (39.5)	129 (39.1)	16 (36.4)	2.197	4	.935
Distal	554	328 (59.2)	198 (60.0)	28 (63.6)			
Other	10	7 (1.3)	3 (0.9)	0 (0.0)			
MSI§							
MSS/L	649	391 (87.3)	231 (83.4)	27 (65.9)	16.315	2	.001
MSI-H	117	57 (12.7)	46 (16.6)	14 (34.1)			
Histologic grade§							
1	80	49 (11.1)	30 (11.2)	1 (3.0)	Fisher¶	NA	.13¶
2	568	346 (78.6)	196 (73.4)	26 (78.8)			
3	92	45 (10.2)	41 (15.4)	6 (18.2)			

* DF = degrees of freedom; NC = not calculated; MSI = microsatellite instability; MSI-H = high-frequency microsatellite instability; MSS/L = microsatellite stable or low-frequency microsatellite instability; NA = not applicable.

† Value of the two-sided Pearson's chi-square test unless otherwise indicated.

‡ Two-sided Pearson's chi-square test, unless otherwise indicated. A *P* value of less than .007 was considered to be statistically significant after adjusting for multiple comparisons with the Bonferroni method of correction.

§ All samples with unavailable data have been omitted from the analysis.

‖ Family history criteria are as described previously (27, 28).

¶ Two-sided Fisher's exact test.

and early age at onset of cancer and gave little consideration to particular tumor characteristics or phenotype. This topic was recently addressed in a study by Lindor et al. (47), which found that relatives of Amsterdam I patients with MSS/L tumors have a statistically significantly lower risk of developing colorectal tumors than relatives of Amsterdam I patients with MSI-H tumors (47). This type of familial segregation of MSS/L tumors was termed "familial colorectal cancer type X" syndrome (47,48). Our study identified 42 Ontario case patients meeting Amsterdam I criteria—17 (40%) carried MSI-H tumors, 15 (36%) carried MSS/L tumors, and 10 (24%) lacked MSI data. Of the 17 case patients with MSI-H tumors, eight (47%) carried the variant allele, and of the 15 case patients with MSS/L tumors, only three (20%) carried the variant allele. Similarly, of the 14 Newfoundland case patients who met Amsterdam I criteria, three (21%) carried MSI-H tumors, seven (50%) carried MSS/L tumors, and four (29%) lacked MSI data. Of three case patients with MSI-H tumors, two (67%) were carriers of the variant allele, and of the seven case patients with MSS/L tumors, three (42%) were carriers of the variant allele. Because the MLH1 –93G>A polymorphism is found in a subset of Amsterdam I case patients with MSS/L tumors, further investigation appears to be warranted to examine the association between this SNP MSS/L tumors, and familial colorectal cancer type X syndrome.

We found that the remaining SNPs examined were not associated with colorectal cancer in either population (Tables 3 and 4), and we did not observe an association between the variant alleles and any clinicopathologic characteristics examined. The MSH2 IVS12-6T>C polymorphism was associated with MSI-H tumor status in the Ontario population of case patients but not in the Newfoundland population.

Our study has several limitations. One is the potential for selection bias. The major reason for nonparticipation in Ontario was unwillingness and/or inability to participate (only 42% were able and willing to participate). It is unlikely that this self-selection would be related to subject's genotype, unless genotype is related to advanced disease stages (49). However, because the general clinical and pathologic characteristics of colorectal cancer of our case patient populations were similar to previously published reports (5,8,50–52), our study was not particularly limited by this potential source of bias.

Another potential source of bias was self-report of ethnicity. In Newfoundland, 15% of study subjects (case patients and study controls) did not report ethnicity, and in Ontario, 9% of study subjects reported mixed ethnicity. To minimize the potential of population stratification and to make the two populations comparable, non-whites and subjects not reporting ethnicity were excluded from subsequent analyses.

Table 6. Genotype frequencies of MLH1 –93G>A polymorphism and clinicopathologic features of case patients with colorectal cancer in Newfoundland*

Feature	No.	Genotype frequency, No. (%)			Chi-square†	DF	*P* value‡
		GG	GA	AA			
Total	430	260 (60.5)	147 (34.2)	23 (5.3)	NC	NC	NC
Sex							
Male	262	158 (60.8)	90 (61.2)	14 (60.9)	0.008	2	.99
Female	168	102 (39.2)	57 (38.8)	9 (39.1)			
Age at diagnosis, y							
<50	46	28 (10.8)	17 (11.6)	1 (4.3)	1.088	2	.58
≥50	384	232 (89.2)	130 (88.4)	22 (95.7)			
Family history§,‖							
Amsterdam I	14	8 (3.1)	6 (4.1)	0 (0.0)	Fisher	NA	.81
Non-Amsterdam I	413	249 (96.9)	141 (95.9)	23 (100.0)			
Tumor location‖							
Proximal	154	88 (33.8)	52 (35.6)	14 (60.9)	9.908	4	.04
Distal	259	164 (63.1)	88 (60.3)	7 (30.4)			
Other	16	8 (3.1)	6 (4.1)	2 (8.7)			
MSI‖							
MSS/L	261	166 (93.3)	83 (83.8)	12 (70.6)	11.655	2	.003
MSI-H	33	12 (6.7)	16 (16.2)	5 (29.4)			
Histologic grade‖							
1	55	38 (15.4)	15 (11.6)	2 (10.0)	Fisher¶	NA	.42¶
2	309	193 (78.1)	99 (76.7)	17 (85.0)			
3	32	16 (6.5)	15 (11.6)	1 (5.0)			

* DF = degrees of freedom; NC = not calculated; NA = not applicable; MSI = microsatellite instability; MSI-H = high-frequency microsatellite instability; MSS/L = microsatellite stable or low-frequency microsatellite instability.

† Value of the two-sided Pearson's chi-square test unless otherwise indicated.

‡ Two-sided Pearson's chi-square test was used, unless indicated otherwise. A *P* value of less than .008 was considered to be statistically significant after adjusting for multiple comparisons with the Bonferroni method of correction.

§ No probands meeting Amsterdam II criteria have been identified in the Newfoundland population. Amsterdam I criteria were as described previously (27).

‖ All samples with unavailable data were omitted from the analysis.

¶ Two-sided Fisher's exact test was used.

Another limitation was the unavailability of some clinical data from our study subjects. Clinical and pathologic characteristics were not available for a variety of reasons (e.g., tumor material not available for MSI analysis, technical difficulties with the MSI analyses, or death of case patients before determination of the tumor's stage or grade). However, the proportion of case patients with missing data was small and was unlikely to affect our results. Accrual of study subjects was not complete in Newfoundland at the time of the analyses because control subject accrual in Newfoundland was still ongoing, which limited our sample size and statistical power.

Our study also has numerous strengths. The large sample size gave us high precision and was less susceptible to fluctuating results. All statistical analyses were adjusted for the main confounding variables of age and sex. In addition, we were able to address the issue of multiple comparisons by using the Bonferroni correction method. A major strength of our study is the use of two separate population-based registries (Ontario and Newfoundland) that had been accrued with similar strategies, thus providing us with confirmation that our observations reflect true associations and are less likely to be due to chance.

The important findings in this study—the associations between the MLH1 –93G>A polymorphism and MSI status in both populations and between the MLH1 –93G>A polymorphism and a strong family history of colorectal cancer (Amsterdam criteria) among Ontario case patients—indicate that low-penetrance alleles of mismatch repair genes may be associated with the risk of colorectal cancer. Because as many as 25–30% of all patients with colorectal cancer exhibit familial aggregation without Mendelian patterns of inheritance (53,54), it would be informative to determine whether MLH1 –93G>A is associated with family risk defined by use of less stringent criteria than Amsterdam I or II.

Most MSI-H colorectal cancer tumors are sporadic, in which methylation of the MLH1 promoter results in mismatch repair–deficient tumors (55). Results from this study indicate that the MLH1 –93G>A promoter polymorphism may act as a modifier allele of colorectal cancer that contributes to susceptibility of MSI-H tumors; however, the mechanism by which subtle alterations of mismatch repair genes contribute to the MSI-H phenotype remains unknown. A possible mechanism through which the MLH1 –93G>A polymorphism could affect gene transcription is by altering the promoter's sensitivity to methylation. The MLH1 promoter is hypermethylated in 15%–20% of sporadic colorectal cancers (56,57). The MLH1 –93G>A promoter polymorphism is located in a CpG island, adjacent to CpG sites that are able to undergo methylation (58). The MLH1 promoter is bidirectional and is used not only by MLH1 but also by another gene, EPM2AIP1, which is located approximately 321 base pairs upstream and has been implicated in myoclonus epilepsy (59,60). Because the MLH1 –93G>A polymorphism occurs in the core promoter region of both genes, it may have a preferential effect on the directionality

of gene transcription. The effect of promoter polymorphisms on gene transcription has been illustrated with the progesterone receptor gene in endometrial cancers; in this cancer, the promoter of the progesterone receptor has a SNP that selectively transcribes one gene isoform over another (61).

In addition to mismatch repair, MLH1 plays a role in the control of the G_2–M cell cycle checkpoint (62). Reductions in the cellular levels of MLH1 tend to have a greater impact on its role in cell cycle control than on DNA mismatch repair (62). Decreased levels of MLH1 expression associated with the MLH1 –93G>A polymorphism may lead to impaired cell cycle control, allowing cells to proceed with cell division before proper DNA repair can be accomplished. This impaired control would overwhelm the mismatch repair mechanism, leading to the accumulation of mutations and subsequent MSI.

Different alterations in mismatch repair genes may have various effects on phenotype, depending on their location within the coding region or regulatory domain of the gene, and on the amino acid or nucleotide change that results. Individual missense alterations that commonly occur in mismatch repair genes and other cancer susceptibility genes may not be pathogenic and severe enough to cause colorectal cancer, but they may affect the level of protein (through mRNA levels or rates of translation) required by the specific cell type to perform normal function (63). Such may be the case with the MLH1 –93G>A promoter polymorphism, which we found to be associated with an increased susceptibility to MSI-H colorectal cancers. Additional characterization of these changes and the cumulative effects that these alterations have on disease risk may lead to new insights into the contribution of low-penetrance alleles to cancer incidence and disease progression.

References

(1) Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, et al. Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. Gastroenterology 2003;124:544–60.

(2) Jemal A, Thomas A, Murray T, Thun M. Cancer statistics, 2002. CA Cancer J Clin 2002;52:23–47.

(3) Wang WS, Chen PM, Su Y. Colorectal carcinoma: from tumorigenesis to treatment. Cell Mol Life Sci 2006;63:663–71.

(4) Ilyas M, Straub J, Tomlinson IP, Bodmer WF. Genetic pathways in colorectal and other cancers. Eur J Cancer 1999;35:1986–2002.

(5) Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919–32.

(6) Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 2005;352:1851–60.

(7) Giardiello FM, Brensinger JD, Petersen GM. AGA technical review on hereditary colorectal cancer and genetic testing. Gastroenterology 2001; 121:198–213.

(8) de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004;4:769–80.

(9) de Jong MM, Nolte IM, te Meerman GJ, van der Graaf WTA, de Vries EGE, Sijmons RH, et al. Low-penetrance genes and their involvement in colorectal cancer susceptibility. Cancer Epidemiol Biomarkers Prev 2002; 11:1332–1352.

(10) Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. Oncogene 2004;23:6471–6.

(11) Ito E, Yanagisawa Y, Iwahashi Y, Suzuki Y, Nagasaki H, Akiyama Y, et al. A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene hMLH1. Biochem Biophys Res Commun 1999; 256:488–94.

(12) Lee KM, Choi JY, Kang C, Kang CP, Park SK, Cho H, et al. Genetic polymorphisms of selected DNA repair genes, estrogen and progesterone receptor status, and breast cancer risk. Clin Cancer Res 2005; 11:4620–6.

(13) Park SH, Lee GY, Jeon HS, Lee SJ, Kim KM, Jang SS, et al. –93G→A polymorphism of hMLH1 and risk of primary lung cancer. Int J Cancer 2004;112:678–82.

(14) Yu JH, Bigler J, Whitton J, Potter JD, Ulrich CM. Mismatch repair polymorphisms and colorectal polyps: hMLH1-93G>A variant modifies risk associated with smoking. Am J Gastroenterol 2006;101:1313–9.

(15) Mathonnet G, Krajinovic M, Labuda D, Sinnett D. Role of DNA mismatch repair genetic polymorphisms in the risk of childhood acute lymphoblastic leukaemia. Br J Haematol 2003;123:45–8.

(16) Listgarten J, Damaraju S, Poulin B, Cook L, Dufour J, Driga A, et al. Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. Clin Cancer Res 2004;10:2725–37.

(17) Hutter P, Couturier A, Rey-Berthod C. Two common forms of the human MLH1 gene may be associated with functional differences. J Med Genet 2000;37:776–81.

(18) Hutter P, Wijnen J, Rey-Berthod C, Thiffault I, Verkuijlen P, Farber D, et al. An MLH1 haplotype is over-represented on chromosomes carrying an HNPCC predisposing mutation in MLH1. J Med Genet 2002; 39:323–7.

(19) Froggatt NJ, Joyce JA, Davies R, Gareth D, Evans R, Ponder BA, et al. A frequent hMSH2 mutation in hereditary non-polyposis colon cancer syndrome. Lancet 1995;345:727.

(20) Salovaara R, Loukola A, Kristo P, Kaariainen H, Ahtola H, Eskelinen M, et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. J Clin Oncol 2000;18:2193–200.

(21) Liu T, Stathopoulos P, Lindblom P, Rubio C, Wasteson Arver B, Iselius L, et al. MSH2 codon 322 Gly to Asp seems not to confer an increased risk for colorectal cancer susceptibility. Eur J Cancer 1998;34:1981.

(22) Ellison AR, Lofing J, Bitter GA. Functional analysis of human MLH1 and MSH2 missense variants and hybrid human-yeast MLH1 proteins in Saccharomyces cerevisiae. Hum Mol Genet 2001;10:1889–900.

(23) Drotschmann K, Clark AB, Kunkel TA. Mutator phenotypes of common polymorphisms and missense mutations in MSH2. Curr Biol 1999; 9:907–10.

(24) Paz-y-Mino C, Fiallo BF, Morillo SA, Acosta A, Gimenez P, Ocampo L, et al. Analysis of the polymorphism [gIVS12-6T > C] in the hMSH2 gene in lymphoma and leukemia. Leuk Lymphoma 2003;44:505–8.

(25) Paz-y-Mino C, Perez JC, Fiallo BF, Leone PE. A polymorphism in the hMSH2 gene (gIVS12-6T>C) associated with non-Hodgkin lymphomas. Cancer Genet Cytogenet 2002;133:29–33.

(26) Brentnall TA, Rubin CE, Crispin DA, Stevens A, Batchelor RH, Haggitt RC, et al. A germline substitution in the human MSH2 gene is associated with high-grade dysplasia and cancer in ulcerative colitis. Gastroenterology 1995;109:151–5.

(27) Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424–5.

(28) Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. Gastroenterology 1999;116:1453–6.

(29) Cotterchio M, McKeown-Eyssen G, Sutherland H, Buchan G, Aronson M, Easson AM, et al. Ontario familial colon cancer registry: methods and first-year response rates. Chronic Dis Can 2000;21:81–6.

(30) Croitoru ME, Cleary SP, Di Nicola N, Manno M, Selander T, Aronson M, et al. Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. J Natl Cancer Inst 2004;96: 1631–4.

(31) Green FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG, et al. AJCC cancer staging manual. 6th ed. New York (NY): Springer; 2002.

(32) Livak KJ. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet Anal 1999;14:143–9.

(33) Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 1987;155: 335–50.
(34) Smith NH, Selander RK. Sequence invariance of the antigen-coding central region of the phase 1 flagellar filament gene (fliC) among strains of Salmonella typhimurium. J Bacteriol 1990;172:603–9.
(35) Wyman AR, White R. A highly polymorphic locus in human DNA. Proc Natl Acad Sci U S A 1980;77:6754–8.
(36) Watts D, MacBeath JR. Automated fluorescent DNA sequencing on the ABI PRISM 310 Genetic Analyzer. Methods Mol Biol 2001;167:153–70.
(37) MacBeath JR, Harvey SS, Oldroyd NJ. Automated fluorescent DNA sequencing on the ABI PRISM 377. Methods Mol Biol 2001;167: 119–52.
(38) Lindor NM, Burgart LJ, Leontovich O, Goldberg RM, Cunningham JM, Sargent DJ, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. J Clin Oncol 2002;20:1043–8.
(39) Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248–57.
(40) Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. Control Clin Trials 1998;19:589–601.
(41) Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. Science 2002;298:2381–2385.
(42) Lipkin SM, Rozek LS, Rennert G, Yang W, Chen PC, Hacia J, et al. The MLH1 D132H variant is associated with susceptibility to sporadic colorectal cancer. Nat Genet 2004;36:694–9.
(43) Shin BY, Chen H, Rozek LS, Paxton L, Peel DJ, Anton-Culver H, et al. Low allele frequency of MLH1 D132H in American colorectal and endometrial cancer patients. Dis Colon Rectum 2005;48:1723–7.
(44) Schofield MJ, Hsieh P. DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol 2003;57:579–608.
(45) Zenke M, Grundstrom T, Matthes H, Wintzerith M, Schatz C, Wildeman A, et al. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J 1986;5:387–97.
(46) Kinoshita S, Akira S, Kishimoto T. A member of the C/EBP family, NF-IL6 beta, forms a heterodimer and transcriptionally synergizes with NF-IL6. Proc Natl Acad Sci U S A 1992;89:1473–6.
(47) Lindor NM, Rabe K, Petersen GM, Haile R, Casey G, Baron J, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA 2005; 293:1979–85.
(48) Kaz AM, Brentnall TA. Genetic testing for colon cancer. Nat Clin Pract Gastroenterol Hepatol 2006;3:670–9.
(49) Hung RJ, Brennan P, Canzian F, Szeszenia-Dabrowska N, Zaridze D, Lissowska J, et al. Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. J Natl Cancer Inst 2005;97:567–76.
(50) Pawlik TM, Raut CP, Rodriguez-Bigas MA. Colorectal carcinogenesis: MSI-H versus MSI-L. Dis Markers 2004;20:199–206.
(51) Diaz LA Jr. The current clinical value of genomic instability. Semin Cancer Biol 2005;15:67–71.
(52) Benson AB 3rd. New approaches to the adjuvant therapy of colon cancer. Oncologist 2006;11:973–80.
(53) Burt RW, Bishop DT, Lynch HT, Rozen P, Winawer SJ. Risk and surveillance of individuals with heritable factors for colorectal cancer. WHO Collaborating Centre for the Prevention of Colorectal Cancer. Bull World Health Organ 1990;68:655–65.
(54) Bodmer WF. Cancer genetics: colorectal cancer as a model. J Hum Genet 2006;51:391–6.
(55) Wheeler JM. Epigenetics, mismatch repair genes and colorectal cancer. Ann R Coll Surg Engl 2005;87:15–20.
(56) Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A 1998; 95:6870–5.
(57) Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S, et al. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci U S A 1998;95:8698–702.
(58) Deng G, Chen A, Pong E, Kim YS. Methylation in hMLH1 promoter interferes with its binding to transcription factor CBF and inhibits gene expression. Oncogene 2001;20:7120–7.
(59) Hitchins M, Williams R, Cheong K, Halani N, Lin VA, Packham D, et al. MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. Gastroenterology 2005;129:1392–9.
(60) Ianzano L, Zhao XC, Minassian BA, Scherer SW. Identification of a novel protein interacting with laforin, the EPM2a progressive myoclonus epilepsy gene product. Genomics 2003;81:579–87.
(61) De Vivo I, Hankinson SE, Colditz GA, Hunter DJ. A functional polymorphism in the progesterone receptor gene is associated with an increase in breast cancer risk. Cancer Res 2003;63:5236–8.
(62) Cejka P, Stojic L, Mojas N, Russell AM, Heinimann K, Cannavo E, et al. Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. EMBO J 2003;22:2245–54.
(63) Fodde R, Smits R. Cancer biology. A matter of dosage. Science 2002; 298:761–3.

Notes

S. Raptis and M. Mrkonjic contributed equally to this work.

This work was supported by a Team Grant from the Canadian Institutes of Health Research (CIHR grant CRT-43821 to J. R. McLaughlin, B. Bapat, J. A. Knight, S. S. Gallinger, R. C. Green, P. S. Parfrey, and B. H. Younghusband). In addition, this work was supported by the NCI, National Institutes of Health, under Request For Applications CA-95-011 and through cooperative agreements with members of the colon family registry and Principal Investigators (U01 CA074783 awarded to Ontario Registry for Studies of Familial Colorectal Cancer). The content of this article does not necessarily reflect the views or policies of the NCI or any of the collaborating centers in the Cancer Family Registry (CFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or CFR. In addition, this work was supported in part by the American Institute for Cancer Research grant 99B055 (B. Bapat and J. A. Knight). S. Raptis was supported by graduate studentships from the University of Toronto (Frank Fletcher Memorial Fund and University of Toronto Fellowship) and the Samuel Lunenfeld Research Institute. M. Mrkonjic was supported by a graduate studentship from the Team in Interdisciplinary Research on Colorectal Cancer with funding from the CIHR and by graduate studentships from the University of Toronto (Frank Fletcher Memorial Fund, Laboratory Medicine and Pathobiology Graduate Award, and University of Toronto Fellowship) and from the Samuel Lunenfeld Research Institute. The authors had full responsibility for the design of the study, the collection of data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

Manuscript received October 30, 2006; revised January 18, 2007; accepted February 2, 2007.