Am. J. Hum. Genet. 63:749–759, 1998

Systematic Analysis of hMSH2 and hMLH1 in Young Colon Cancer Patients and Controls

Susan M. Farrington,[1] Juili Lin-Goerke,[2] Jessica Ling,[2] Yute Wang,[2] John D. Burczak,[2] David J. Robbins,[2,*] and Malcolm G. Dunlop[1]

[1]University of Edinburgh Department of Surgery and Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh; and [2]SmithKline Beecham Pharmaceuticals, Collegeville, PA

Summary

Germ-line mutations in DNA mismatch–repair genes impart a markedly elevated cancer risk, often presenting as autosomal dominant hereditary nonpolyposis colorectal cancer (HNPCC). However, there are no pathognomonic features of HNPCC, not all gene carriers have a family history of the disease, and families fulfilling the Amsterdam criteria are relatively uncommon. Genetic testing of probands with early-onset colorectal cancer, irrespective of family history, is one approach that would allow predictive genetic testing of at-risk relatives. We cloned and sequenced hMSH2 and hMLH1 introns, to optimize genomic sequencing. We then systematically analyzed the entire hMSH2 and hMLH1 genes, by genomic sequencing and in vitro synthesized-protein–truncation assay (IVSP), in 50 colorectal cancer patients <30 years of age at diagnosis. To determine polymorphic variants, 26 anonymous donors also were sequenced. All subjects analyzed had at least 1 of 37 different polymorphic or pathogenic variants. IVSP complemented genomic sequencing, by detection of mutations not identified by genomic analysis. Fourteen cancer patients (28%) had pathogenic mutations, and a number of other variants also may have had a pathogenic significance that remains to be elucidated. Tumor replication-error status was useful in targeting sequencing efforts for this cohort of young patients: sensitivity was 86%, specificity 73%, and positive and negative predictive values 63% and 90%, respectively. These data indicate that an appreciable proportion of young colon cancer probands carry a germ-line mutation in a DNA mismatch–repair gene.

Received October 10, 1997; accepted for publication June 29, 1998; electronically published August 7, 1998.

Address for correspondence and reprints: Dr. Malcolm G. Dunlop University of Edinburgh Department of Surgery and MRC Human Genetics Unit, Western General Hospital, Edinburgh, EH4 2UX, United Kingdom. E-mail: Malcolm.Dunlop@hgu.mrc.ac.uk

* Present affiliation: SmithKline Beecham Clinical Laboratories, Van Nuys, CA.

0002-9297/98/6303-0014$02.00

Introduction

Colorectal cancer (MIM 114500) is one of the most common fatal cancers in developed countries and represents a significant public-health issue. The United States and the United Kingdom are high-incidence countries, with ~133,500 new cases and ~55,300 deaths (Parker et al. 1996*b*) in the United States and 30,941 new cases and ~17,000 deaths in the United Kingdom (Her Majesty's Stationery Office, United Kingdom Cancer Registry Data) per year. The population lifetime risk is 1 in 25 in the United States and in northern Europe (Sharp et al. 1993). In the United States, the annual incidence increased from 98,757 in 1973 to 131,200 in 1997 (see the SEER [Surveillance, Epidemiology, and End Results] database), while population-based cancer registration in Scotland identified a 7% increased incidence during the years 1981–90 (Sharp et al. 1993). Identification of people predisposed to the disease would allow targeting of effective preventative measures, with the aim of reduction of the considerable cancer-related mortality (Burke et al. 1997).

One group with a very high colorectal cancer risk consists of those individuals who carry germ-line mutations in genes that participate in DNA-mismatch repair. hMSH2 (Fishel et al. 1993; Leach et al. 1993) and hMLH1 (Bronner et al. 1994; Papadopoulos et al. 1994) are involved most frequently, but mutations in hMSH6 (Akiyama et al. 1997) and in hPMS1 and hPMS2 (Nicolaides et al. 1994) also occur in a minority of cases. Such mutations usually are associated with marked familial aggregation of colorectal, uterine, and other cancers constituting the clinically defined autosomal dominant syndrome hereditary nonpolyposis colorectal cancer (HNPCC; MIM 120435 and 120436) (Lynch et al. 1993; Mary et al. 1994; Nystrom-Lahti et al. 1995; Wijnen et al. 1995, 1997; Liu et al. 1996; International Collaborative Group on HNPCC database). However, an appreciable proportion of patients who have very early–onset colorectal cancer but who do not fulfill pragmatic criteria for HNPCC (Vasen et al. 1991) also carry mismatch-repair–gene mutations (Liu et al. 1995; Dun-

lop et al. 1997). Thus, restriction of genetic testing to individuals from families fulfilling HNPCC criteria is likely to exclude a significant fraction of gene carriers in the general population. However, screening unselected patients with sporadic cancer represents an enormous workload and may provide a very low yield of mutation carriers (Liu et al. 1995; Tomlinson et al. 1997; Wijnen et al. 1997). The majority of kindreds analyzed for hMLH1 and hMSH2 mutations have been selected specifically because of multiple affected cases; therefore, lifetime penetrance of mutations in these families is correspondingly high, at ~80% (Vasen et al. 1996). However, when a population-based case-finding approach is used, penetrance appears to be lower than that for families fulfilling HNPCC criteria (Dunlop et al. 1997). Concerns over penetrance and the potential effect of ascertainment bias mean that the population prevalence of mutations in DNA mismatch–repair genes cannot be estimated with any certainty. It is clear that the definition of indications for genetic testing and the interpretation of results are critical for hereditary cancer syndromes (Giardiello et al. 1997). We set out to determine the prevalence of hMSH2 and hMLH1 alterations in a cohort of 50 patients with extremely early onset of colon cancer, irrespective of family history. We also wished to determine the prevalence of variants in a control population, since polymorphisms could have a significant confounding effect on strategies based on testing at-risk individuals rather than probands.

Patients and Methods

Patients and Samples

A total of 76 subjects were studied: 50 unrelated Scottish patients diagnosed with colorectal cancer at <30 years of age and 26 anonymous donors from the United States. There were 15 male and 11 female anonymous donors, who were cancer free at the time of sampling and whose mean age was 41 years, which corresponded well with the current age of the cancer patients. The cancer cohort was identified retrospectively from cancer registrations since 1970; thus, there is a potential survivorship bias, but our unpublished data do not suggest that this is a major effect (see Discussion). None of the study subjects were referred specifically because of a family history of colon cancer. All cancer patients had histologically confirmed colorectal cancer. Peripheral blood was drawn, and DNA was purified from peripheral-blood leukocytes. Attempts were made to establish lymphoblastoid cell lines from each cancer patient.

A detailed family history was obtained for all 50 patients with cancer diagnosed at <30 years of age, and paraffin-embedded archival tumor material, along with matched normal tissue, was obtained for 42 of these patients. Pedigree ascertainment was rigorous and included a patient interview or questionnaire, review of hospital and pathology records, and verification through birth, death, and marriage records for central Scotland. At least three, and usually four, generations of ancestors were traced for every proband, and in no cases were the pedigrees connected. For every relative in each kindred, current health/cancer status was ascertained, or the relative was determined to have died. Family history was categorized as follows: 0 = no family history of cancer; 1 = relative with noncolorectal cancer; 2 = relative with colorectal cancer who does not fulfill the Amsterdam criteria; and 3 = nuclear family fulfills the Amsterdam criteria for HNPCC (Vasen et al. 1991).

Analysis of Tumor Microsatellite Instability

Archival paraffin-embedded tumor tissue and matched normal tissue (usually adjacent normal colon tissue) were sectioned at 10 μm and were laid on glass slides. A representative section was stained with hematoxylin/eosin and was examined by microscopy. Only areas with >90% tumor were used in the analysis of microsatellite instability. Adjacent normal tissue was scraped off and discarded. Normal tissue was obtained, in the same fashion, from a separate block when possible, ensuring that no tumor tissue was included in the section. Depending on the size and cellularity of the sections, 3–15 sections were needed for tumor/normal tissue. DNA was purified by use of the Qiagen Tissue Kit, according to the manufacturer's protocol.

As described elsewhere (Liu et al. 1995), 2–5 μl of tumor and normal template solution was used in PCR reactions, except that, in this study, all primers were fluorescently labeled. The marker loci analyzed were four $(CA)_n$ repeats (D2S123, D5S122, D5S346, and D13S160) and four poly-A repeats (BAT25, BAT26, BAT40, and Pax6-I253). The Pax6-I253 repeat is a poly-A tract at the 3′ end of Pax6 in exon 13, and the primer sequences are listed at http://www.hgu.mrc.ac.uk/Users/Malcolm.Dunlop/MMRprim.htm. Analysis was performed by use of an ABI 310 Automated Genetic Analyzer, with GeneScan software. Each tumor DNA sample was compared with matched normal DNA. At the start of the analysis, tumor microsatellite instability was defined as marker band shifts at a minimum of two loci.

Genomic Sequencing

DNA was extracted from peripheral blood by use of the Nucleon DNA Extraction kit (Scotlab) or the Puregene DNA Isolation kit (Gentra Systems), according to the manufacturer's instructions. We cloned and sequenced a number of introns from both hMLH1 and hMSH2, to optimize genomic sequencing. Specific primers derived from these sequences and previously published sequences (Kolodner et al. 1994) that worked well in this study are listed at http://www.hgu.mrc.ac.uk/

Users/Malcolm.Dunlop/MMRprim.htm. Each exon of hMSH2 and hMLH1 was amplified by PCR using 40 ng of genomic DNA in a volume of 50 μl. Final reaction concentrations were 1 × PCR Buffer II (Perkin Elmer), 3.0 mM $MgCl_2$ (or 1.5 mM, for hMSH2 exon 1), 0.2 mM dNTPs, 10 pmol of each specific oligonucleotide primer, and 1.25 U *Taq* polymerase. Amplification was hot started at 94°C for 3 min, followed by 35 cycles of 94°C for 20 s, 55°C for 20 s, and 72°C for 40 s. The final reaction was extended at 72°C for 10 min, followed by storage at 4°C. Cycle sequencing was performed by use of the PRISM Ready Dye Terminator Cycle Sequencing kit with AmpliTaq DNA polymerase, FS (Taq-FS; Perkin Elmer/Applied Biosystems), and Applied Biosystems DNA sequencer model 373A or 377 (Parker et al. 1996*a*), according to the manufacturer's instructions. DNA sequence analysis was performed by use of Sequencher 3.0 software (Gene Codes), by comparison of published genomic sequences of hMLH1 (Han et al. 1995; Kolodner et al. 1995; Genome Database accession number 249617) and hMSH2 (Kolodner et al. 1994; Genome Database accession number 203983) with those from cancer cases or population donors.

In Vitro Synthesized-Protein–Truncation Assay (IVSP)

cDNA was generated by reverse transcription of RNA purified from lymphoblastoid cell lines from the affected index case. PCR amplification of the cDNA was used to introduce a 17-bp consensus T7 promoter sequence and a mammalian translation-initiation sequence in frame with a unique hMLH1 or hMSH2 sequence. PCR primer sequences and conditions were similar to those previously described in this article, with some modification (Liu et al. 1995). Each gene was amplified in two or three overlapping segments. Resultant PCR products were used in a coupled transcription-translation reaction (Promega) incorporating 2–5 μCi of ^{35}S-methionine. Labeled in vitro–transcribed protein products from the reaction were heat denatured and were analyzed by use of 8%, 10%, and 12% SDS-PAGE gels. Gels were washed in fixative and were autoradiographed overnight at room temperature. All samples showing truncated protein products were reamplified independently, and an additional IVSP analysis was performed, for confirmation. For each analysis, normal control samples were run in parallel, and the wild-type full-length protein was noted. In most of the analyses, artifactual bands were visible—presumably owing to internal initiation—since these were visible in samples from normal controls.

Long-Range PCR

For long-range PCR of sample 817, the GeneAmp XL PCR kit (Perkin Elmer) was used with forward primer 5′-GGC CAT TGT CAC AGA GGA TAA GA-3′ and reverse primer 5′-ACA CAG CCC ACG AAG GAG TG-3′. The reaction mixture contained ~400 ng of genomic DNA in a volume of 50 μl. Final reaction concentrations were 1 × PCR Buffer II (Perkin Elmer), 1.5 mM $Mg(OAc)_2$, 0.8 mM dNTPs, 40 pmol of each specific oligonucleotide primer, and 4 U rTth DNA polymerase. Amplification was hot started at 94°C for 1 min, followed by 26 cycles of 94°C for 15 s and 68°C for 10 min. The final reaction was extended at 72°C for 10 min, followed by storage at 4°C.

Results

Tumor Microsatellite Instability

We obtained paraffin-embedded archival matched tumor and normal material from the respective pathology departments for 42 of the cancer patients. Forty of these samples could be amplified by PCR, to allow reliable assessment of microsatellite-instability status, and 19 (48%) exhibited the instability phenotype. This proportion is similar to that of our previous series of patients <35 years of age (Liu et al. 1995) but is very much higher than the 15% noted for a non–age-selected Scottish cohort of colorectal cancer patients (Bubb et al. 1996) analyzed in our laboratories. Tumors exhibiting the instability phenotype almost exclusively showed band shifts at six or more markers. The lowest proportion of markers showing band shifts was 4 (50%) of 8. Instability status related to the detection of germ-line mutations, by genomic sequencing and by IVSP, for all 50 patients studied is shown in table 1. We could not detect a mutation, by genomic sequencing, in one patient with a microsatellite-unstable tumor; unfortunately, we also were unable to obtain IVSP data, because the lymphoblastoid cell line did not transform and the patient succumbed to malignant disease. Even when this case was included in the analysis, it was clear that microsatellite instability has considerable value in the prediction of germ-line mismatch-repair–gene mutations (table 2). For this very early–onset cohort, sensitivity of microsatellite-instability status for mutation was 86% (12/14), and specificity was 73% (19/26). Positive and negative predictive values were 63% (12/19) and 90% (19/21), respectively. Microsatellite-instability status and family history are summarized in table 3, with all five analyzable tumors from families fulfilling the Amsterdam criteria exhibiting microsatellite instability. The proportion of microsatellite-unstable tumors correlated with the degree of family history, as has been noted by others (Wijnen et al. 1997).

Genomic Sequencing

Using the combination of genomic sequencing and IVSP analysis, we identified a total of 15 germ-line mutations (table 4), and 5 of these mutations are novel (patients 579, 815, 817, 830, and 1157). The relation-

Table 1

Tumor Microsatellite Instability, Family History, and Germ-Line hMSH2/hMLH1 Mutations, for Young Cancer Probands

Patient	Family History[a]	Tumor Microsatellite Instability	IVSP		Genomic Sequencing	
			hMSH2	hMLH1	hMSH2	hMLH1
825	1	POS	POS	NEG	POS	NEG
818	2	POS	POS	NEG	POS	NEG
528	0	POS	ND	ND	POS	NEG
830	2	POS	ND	ND	POS	NEG
814	2	POS	POS	NEG	POS	NEG
579	2	POS	POS	NEG	POS/NEG	NEG
815	2	POS	NEG	POS	NEG	POS
804	3	POS	NEG	POS	NEG	POS
533	3	POS	NEG	POS	NEG	POS
329	3	POS	NEG	NEG	NEG	POS
817	2	POS	NEG	POS	NEG	NEG
1052	3	POS	NEG	NEG	NEG	NEG
982	3	POS	NEG	NEG	NEG	NEG
889	1	POS	NEG	POS	NEG	NEG
871	0	POS	NEG	NEG	NEG	NEG
869	0	POS	ND	ND	NEG	NEG
862	1	POS	NEG	NEG	NEG	NEG
813	2	POS	NEG	NEG	NEG	NEG
615	1	POS	NEG	NEG	NEG	NEG
1157	0	NEG	NEG	NEG	POS	NEG
696	1	NEG	NEG	NEG	NEG	POS
1323	2	NEG	NEG	NEG	NEG	NEG
1279	2	NEG	NEG	NEG	NEG	NEG
1161	1	NEG	NEG	NEG	NEG	NEG
1012	1	NEG	NEG	NEG	NEG	NEG
964	0	NEG	NEG	NEG	NEG	NEG
887	0	NEG	NEG	NEG	NEG	NEG
876	0	NEG	NEG	NEG	NEG	NEG
860	1	NEG	NEG	NEG	NEG	NEG
822	0	NEG	NEG	NEG	NEG	NEG
819	1	NEG	NEG	NEG	NEG	NEG
812	2	NEG	NEG	NEG	NEG	NEG
809	1	NEG	NEG	NEG	NEG	NEG
757	0	NEG	NEG	NEG	NEG	NEG
737	1	NEG	NEG	NEG	NEG	NEG
649	0	NEG	ND	ND	NEG	NEG
622	1	NEG	NEG	NEG	NEG	NEG
551	0	NEG	NEG	NEG	NEG	NEG
510	1	NEG	NEG	NEG	NEG	NEG
106	1	NEG	NEG	NEG	NEG	NEG
1005	0	ND	NEG	NEG	NEG	NEG
983	0	ND	NEG	Incomplete	NEG	NEG
960	3	ND	ND	ND	NEG	NEG
824	2	ND	ND	ND	NEG	NEG
823	1	ND	NEG	NEG	NEG	NEG
816	1	ND	ND	ND	NEG	NEG
523	2	ND	NEG	NEG	NEG	NEG
324	2	ND	ND	ND	NEG	NEG
841	1	…	ND	ND	NEG	NEG
559	0	…	NEG	NEG	NEG	NEG

Note.—Designations and data for patients with pathogenic mutations are underlined. ND = not determined.

[a] Categories are described in Patients and Methods.

ship of family history and tumor microsatellite instability to mutation status is discussed in the relevant sections. Genomic sequencing initially only detected 12 (80%) of the 15 possible mutations. However, when informed by IVSP data, genomic analysis characterized a total of 14 (93%) mutations. However, it is important to note that this improved detection rate only resulted when complementary IVSP analysis was used.

Table 2

Relationship of Tumor Microsatellite-Instability Status to Germ-Line Mismatch-Repair–Gene Mutations

Tumor Microsatellite-Instability Status	No. of Tumors, by Germ-Line–Mutation Status		
	Detected	Not Detected	Total
Positive	12	7	19
Negative	2	19	21
Total	14	26	40

Two mutations are worthy of specific comment. The mutation in patient 1157 disrupts the initiation codon so that an in-frame methionine at codon 26 appears to act as a surrogate transcription-start site, although studies indicate that the transcript is unstable. Characterization of the mutation identified in patient 579 proved complex. Replicate hMSH2 IVSPs for patient 579 detected a very short protein fragment, which could not be explained on the basis of the His→Tyr mutation at codon 639, identified by genomic sequencing. In view of the IVSP results, we performed additional genomic sequencing and eventually identified the second mutation, at the splice acceptor site of exon 14 (table 4), which initially had been missed. Using restriction-site changes induced by each mutation, we traced both variants through the family and showed that they reside on the same allele. Extensive sequencing of reverse transcription–PCR products revealed that this complex double mutation results in an in-frame deletion of exons 12–14, thus accounting for the very short IVSP fragment. A His→Tyr mutation at codon 639 has been reported elsewhere (Leach et al. 1993; Liu et al. 1994) and results in a surrogate splice donor site and a 92-bp frameshift deletion of nt 1914–2006, generating a premature-termination codon 17-bp downstream of the exon 13 splice acceptor site. To determine whether any mRNA containing the 92-bp splice mutation reported by Liu et al. (1994) was expressed in patient 579, we serially amplified and sequenced overlapping fragments from cDNA that included exons 11–15. We only detected a wild-type sequence when the 3′ primer was 5′ of the splice acceptor site of exon 15. Hence, we have established that the double mutation identified in patient 579 is distinct from that reported by Liu et al. (1994) and that both mutations are required for deletion of exons 12–14.

Sequencing in cancer patients and controls identified a total of 37 variants, including presumed polymorphisms and pathogenic mutations (tables 4 and 5). The polymorphisms are also of considerable interest, since some of these polymorphisms possibly could contribute to cancer predisposition. However, we did not consider them pathogenic unless either there was a nonconservative amino acid change or the variant arose at a conserved sequence around a splice site. Among the 26 donor DNAs, no previously published mutations were found, but 6 donor DNAs had a total of four novel variants (table 5). The significance of these changes is uncertain and cannot be investigated, owing to the anonymity of the donors. The His→Tyr change at codon 718 of hMLH1 was present in three U.S. donors, but we later discovered that two of these donors were siblings.

IVSP

Lymphoblastoid cell lines were available from 41 of the patients <30 years of age. Failure to transform usually was due to concurrent chemotherapy and/or terminal cancer. In one case, we were not able to obtain results from one segment of hMLH1, despite repeated attempts (table 1). Thus, near-complete or complete IVSP analysis of both genes was possible for 41 cases, and 9 (22%) were positive (4 for hMSH2 and 5 for hMLH1). Six of these cases were detected by screening with genomic sequencing. For the two cases that were found to be completely negative, by genomic sequencing, we were prompted to perform additional detailed genomic analysis, as a result of reproducible IVSP and cDNA findings. These studies showed that patient 817 has a large genomic deletion (see below). The other patient (889) showed a reproducible novel IVSP band but no genomic DNA change and an apparent deletion of the 3′ end of the gene, by cDNA sequencing. We have not been able to delineate the genomic change in this patient. In a third patient (579), there also was a dis-

Table 3

Relationship of Family History and Tumor Microsatellite-Instability Status to Germ-Line MLH1/hMSH2 Mutations, for Young Colorectal Cancer Probands

Tumor Microsatellite-Instability Status	No. of Cases, by Family-History Category[a]				
	0	1	2	3	Total
All cases:					
Positive	3	4	7	5	19
Negative	8	10	3	0	21
Not determined	3	3	3	1	10
Total	14	17	13	6	50
hMSH2:					
Positive	2	1	4	0	7
Negative	12	16	9	6	43
hMLH1:					
Positive	0	2	2	3	7
Negative	14	15	11	3	43

Note.—Mutation status is from combined genomic sequencing/IVSP data. IVSP data were not available for 9 cases for hMSH2 and 10 cases for hMLH1 (table 1).

[a] Family history categories are described in Patients and Methods.

Table 4

Pathogenic hMLH1 and hMSH2 Mutations Identified in Young Colorectal Cancer Probands

Gene and Patient	Mutation	Nucleotide Change	Effect on Coding Sequence	Location
hMLH1:				
329	616delAAG	Deletion of AAG at 1846–1848	Deletion of Lys616	Exon 16
533	IVS8–3delTA	Deletion of TA at 677−3	Splice mutation	IVS 8
696	K618A	AA→GC at 1852–1853	Lys→Ala at 618	Exon 16
804	R659X	C→T at 1975	Arg→Stop at 659	Exon 17
815	IVS1+1G→A	G→A at 116+1	Splice mutation	IVS 1
817*	del exon 13	Deletion of ~3 kb involving IVS 12, through exon 13 to IVS 13	Deletion of codons 470–520 (exon 13)	IVS 12–13, exon 13
889	...	Not identified	Truncation on IVSP	Exons 12–19
hMSH2:				
528*	R406X	C→T at 1216	Arg→Stop at 406	Exon 7
579	H639Y	C→T at 1915	Double mutation results in deletion of codons 588–820 (exons 12–14)	Exon 12, IVS 13
	IVS13−1G→T	G→T at 2211		
814*	Q601X	C→T at 1801	Gln→Stop at 601	Exon 12
818*	Q252X	C→T at 754	Gln→Stop at 252	Exon 4
825*	delCTGT	Deletion of CTGT at 808–811	Deletion of codons 265–314 (exon 5)	Exon 5
830	R680X	C→T at 2038	Arg→Stop at 680	Exon 13
1157	M1L	A→T at 1	New initiation at codon 26	Exon 1

Note.—Asterisks (*) designate mutations previously reported from our laboratories (Liu et al. 1995; Dunlop et al. 1997), but genomic mutation of patient 817 has not been previously reported. IVS = intervening sequence.

crepancy between IVSP and genomic-sequencing data. As discussed in the previous section, this patient has two mutations, and the codon 639 His→Tyr change was detected by sequencing. The splicing mutation initially was missed by genomic sequencing but was identified by IVSP analysis, which informed the subsequent genomic and cDNA sequencing that fully characterized this complex mutation.

In this study, when RNA purified from lymphoblastoid cell lines was used, IVSP was a robust method of identifying detectable mutations. There were 11 mutations that could have been identified by IVSP (including the mutation in patient 889, which was detectable only by IVSP). Although we did not have RNA from two patients, sequencing and IVSP were complementary, since IVSP identified 9 of 13 mutations (sensitivity 69%) and genomic sequencing detected 12 of 15 mutations (sensitivity 80%).

Long-Range PCR

Replicate IVSPs and cDNA sequencing of samples from patient 817 reproducibly demonstrated a truncation in hMLH1 due to deletion of exon 13. However, extensive genomic sequencing failed to identify the mutation at the DNA level. Hence, we analyzed the intronic region around exon 13 by long-range PCR, to determine whether any large genomic deletion had removed that exon completely. The forward primer was in exon 12 and the reverse in exon 14, resulting in an ~15.5-kb wild-type product. By use of this approach, patient 817 was shown to carry a genomic deletion of ~3 kb, which resulted in removal of exon 13 (fig. 1).

Family-History and Germ-Line–Mutation Status of the Patients <30 Years of Age

Three (50%) of the six young probands from families that fulfilled the Amsterdam criteria had germ-line mutations. The nuclear family of patient 814 did not fulfill these criteria and, so, was categorized as being in family-history group 2. However, when several affected distant relatives unknown to the nuclear family were identified, as a result of research interests, it became clear that this was a striking cancer family. There was a loose correlation between family history and mutation prevalence, but the association was not of practical value (table 3). The prevalence of detectable mutations in each of the family-history groups was 50% (3/6) for group 3 (HNPCC), 46% (6/13) for group 2, 18% (3/17) for group 1, and 14% (2/14) for group 0. Thus, if analysis had been restricted to the Amsterdam-criteria families, only 21% of all mutations would have been identified. Even if any first- or second-degree relative with colo-

Table 5

hMSH2/hMLH1 Variants Considered to be Nonpathogenic Polymorphisms in Patients or Controls

Gene and Variant	Nucleotide Change	Effect on Coding Sequence	Location	Frequency			Reference
				Allele 1	Allele 2	Percentage	
hMLH1:							
E/E13[a]	39G→A	No change at Glu13	Exon 1	140	2	1.4	
A/A125[a]	372A→G	No change at Ala125	Exon 4	141	1	.7	
IVS7+11G→C	G→C at 588+11	None apparent	IVS 7	141	1	.7	
I/V219[a]	654A→G	Ile or Val at 219	Exon 8	97	45	32.7	Tomlinson et al. (1997)
IVS11−7insA	1039−7insA	None apparent	IVS 11	141	1	.7	
IVS13+14G→A	G→A at 1558+14	None apparent	IVS 13	140	2	1.4	Tannergard et al. (1995)
IVS14−19A→G[a]	A→G at 1668−19	None apparent	IVS 14	133	9	6.3	Buerstedde et al. (1995)
L/L636	1908G→A	No change at Leu636	Exon 17	141	1	.7	
L/L653	1959G→T	No change at Leu653	Exon 17	141	1	.7	Buerstedde et al. (1995)
G/G706[a]	2118C→T	No change at Gly706	Exon 19	137	5	3.5	
H/Y718	2152C→T	His or Tyr at 718	Exon 19	139	3	2.1	
L/V729[a]	2184C→G	Leu or Val at 729	Exon 19	141	1	.7	
hMSH2:							
IVS1+9C→G[a]	C→G at 211+9	None apparent	IVS 1	121	21	14.8	Bubb et al. (1996)
K/K74[a]	219G→A	No change at Lys74	Exon 2	141	1	.7	
L/L191	574C→T	No change at Leu191	Exon 3	139	3	2.1	Moslein et al. (1996)
IVS9−9T→A	T→A at 1511−9	None apparent	IVS 9	21	121	85.2	Borresen et al. (1995)
IVS10+12A→G[a]	A→G at 1661+12	None apparent	IVS 10	54	88	62.0	Wijnen et al. (1994)
IVS10+6T→C[a]	T→C at 1661+6	None apparent	IVS 10	0	142	100.0	
IVS10−9G→A[a]	G→A at 1662−9	None apparent	IVS 10	141	1	.7	
C/G641	121T→G	Cys or Gly at 641	Exon 12	141	1	.7	
IVS12−6T→C[a]	T→C at 2006−6	None apparent	IVS 12	133	9	6.3	Leach et al. (1993)
I/V770	2308A→G	Ile or Val at 770	Exon 14	141	1	.7	

Note.—Five patients with mutations that we had reported previously were not sequenced for all exons; thus, the total number of alleles sequenced is 142 (45 patients and 26 controls). IVS = intervening sequence.

[a] Observed in cancer patients but not considered to be pathogenic, on the basis of the available evidence.

rectal or uterine cancer had been considered as an indicator for gene analysis, only 64% of the mutations would have been identified (tables 1 and 3).

Discussion

We report a systematic investigation of hMSH2 and hMLH1, the two genes most commonly involved in hereditary predisposition to colorectal cancer. Detailed analysis of the entire coding sequence of both genes was completed for a total of 76 individuals, and we characterized an appreciable number of variants. In all, there were 37 variants, and 15 (41%) of these appear to be pathogenic, on the basis of dramatic gene alterations such as truncation, splice errors, short deletions, and nonconservative amino acid changes. Fourteen (28%) of the cancer patients had mutations, indicating that patients with very early–onset colorectal cancer merit analysis of DNA mismatch–repair genes. The large deletions noted in patient 817 and possibly in patient 889 are of some interest, since other, similar deletions that appear to be mediated by *Alu* recombination have been reported (Nystrom-Lahti et al. 1995; Mauillon et al. 1996). Hence, when mutation screening fails to identify sequence changes in HNPCC families or in early-onset cases, consideration should be given to methods for detection of large deletions, such as the long-range PCR used here, Southern blot analysis, or FISH using gene-specific cosmids. Our genomic DNA–sequencing approach could not have detected promoter mutations that might affect hMLH1 or hMSH2 expression, but, to date, promoter alterations have been demonstrated only as somatic events in tumor cells (Scherer et al. 1996; Kane et al. 1997).

The sensitivity of genomic sequencing was 80%, and that of IVSP was 64%. In this study, IVSP and genomic

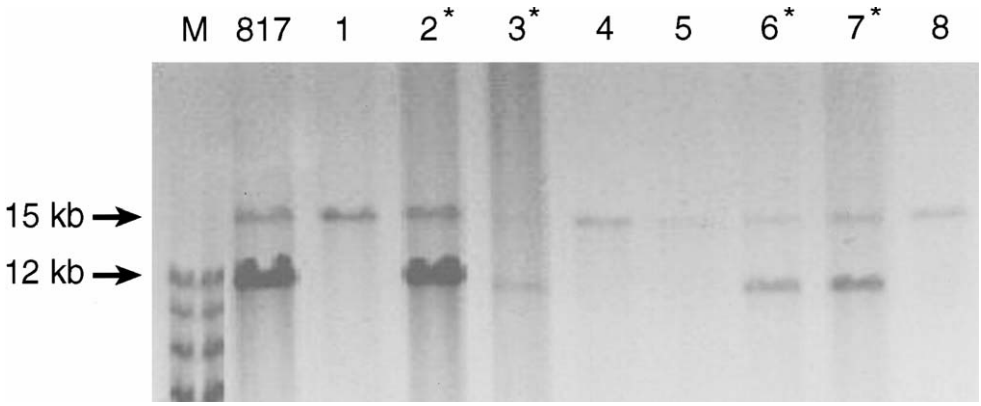

Figure 1 Long-range PCR of hMLH1 exons 12–14 (~15 kb). Patient 817 shows a genomic deletion of ~3 kb that includes exon 13 in the hMLH1 gene. The asterisks (*) indicate four additional affected or at-risk family members who carry the same deletion allele detectable by long-range PCR and the exon 13 deletion detected by reverse transcription–PCR analysis. Lane M, Genomic size marker.

sequencing were complementary, although RNA-based mutation analysis raises practical problems for large-scale mutation detection. Nonetheless, routine genomic analysis of DNA fragments spanning intron/exon boundaries would have missed an appreciable proportion of all mutations in this patient cohort. For the purposes of this study, our analysis was restricted to RNA purified from lymphoblastoid cell lines. We and others have experienced problems with RNA from peripheral blood leukocytes (Kohonen-Corish et al. 1996) when analyzing hMLH1 in particular. Since a relatively small number of patients have very early–onset colorectal cancer, it may be practical to obtain RNA from other sources, such as from skin biopsy or fresh surgical specimens, although this also can be problematic.

The cohort of patients with onset of colorectal cancer at <30 years of age who were analyzed for this study constitute a highly select group. Colorectal cancer in this age group accounts for <0.5% of all cancer registrations in Scotland (Sharp et al. 1993). The patients were identified retrospectively—from a period of 23 years, to 1993—and, therefore, sampling could occur only if the patients had survived. If survival is influenced by mutation status, then our assessment of mutation prevalence may ultimately be biased. However, we found no evidence of a survival effect for HNPCC gene carriers in Scotland, for the time period in question (M.G.D., unpublished data). In addition, the group of patients in this study developing cancer at <30 years of age had been identified prospectively over the last 4 years, and a number of gene carriers identified during this period have since died from their disease. Hence, we believe that the proportion (28%) of mutation carriers in the cohort studied is representative. Although we have yet to demonstrate clinical utility, patients who develop colorectal cancer at <30 years of age seem to represent a population in which analysis of DNA mismatch–repair genes is worthwhile, regardless of family history.

In addition to contributing to the growing data on the mutational spectra of hMSH2 and hMSH1, the polymorphic variants described here will have value as intragenic markers for linkage studies in which mutations cannot be identified in large families. It is important to consider the prevalence of these variants when the possible pathogenic nature of sequence alterations identified during predictive testing is assessed. All the samples analyzed had one or more polymorphisms, when compared with sequences stored in GenBank (accession nos. U03911 and U04045 [for hMSH2 cDNA] and U07343 and U07418 [for hMLH1 cDNA]). Such a high prevalence may confound mutation-screening techniques such as SSCP and may result in a high level of secondary sequencing of polymorphisms. The benefit of the affected-proband approach used in this study is that there is de facto evidence of an association between the putative mutation and early-onset cancer. In cases for which there is a dramatic mutation, such as a deletion and/or a truncating mutation, there is no problem in the assignment of cancer risk to that change. However, amino acid substitutions do cause a dilemma in predictive testing. Although not definitive, association of a missense mutation with early-onset cancer does contribute considerable clinical significance if similar changes never have been identified in healthy donors.

The microsatellite-instability phenotype in the tumors from the young patients investigated in this study consisted of band shifts at almost every locus examined. This differs from the microsatellite instability observed in non–age-selected cohorts in which one or two markers may have been altered (Lothe et al. 1993; Bubb et al. 1996). Microsatellite instability was a useful predictor of a germ-line mutation in this study group, with a pos-

itive predictive value of 63%. However, reliance on the instability phenotype as the sole indicator of the need for mismatch-repair–gene analysis would have excluded some patients who carry mutations (table 2). Of some interest are the seven cases that are microsatellite unstable but for which we were unable to detect a mutation in hMLH1 or hMSH2. These cases may be due to mutations in other DNA mismatch–repair genes, such as PMS1, PMS2, hMSH3, hMSH6, or other MLH/MSH homologues, especially since three cases were associated with a family history of cancer (tables 1 and 3).

Although a strong family history correlated with mutation prevalence in this study, family history was not a useful determinant of germ-line mutations, in practice. In this study, the prevalence of mutations in families fulfilling the Amsterdam criteria was 50%, which is identical to that in a Dutch study (Wijnen et al. 1997). Another study that included some HNPCC families showing linkage to HMSH2 or hMLH1 identified 80% of mutations (Liu et al. 1996); therefore, it is clear that analysis of samples from HNPCC families is highly likely to identify mutations. However, in practice, recognition of HNPCC families is confounded by lack of pathognomonic features, deficiencies in family information, adoption, early death of relatives that is due to unrelated causes, and incomplete gene penetrance. Although many mutation carriers in this study did have one first- or second-degree relative with colorectal or uterine cancer, mutation screening of people with only one affected relative would be impractical, owing to the enormous workload and the detection of variants of unknown significance. Approximately 1% of the entire population have either two first-degree relatives affected by colorectal cancer or one relative affected at $\leqslant 45$ years of age (Dunlop and Campbell 1997), whereas a substantially higher proportion have only one affected relative. Thus, screening for mutations in DNA mismatch–repair genes should concentrate not only on the relatively few HNPCC families but also on early-onset cases, as they arise, irrespective of family history and perhaps targeted by tumor microsatellite-instability status. It is interesting to note that 11 (25%) of 44 probands who came from families that did not fulfil the Amsterdam criteria had mutations. This differs substantially from the results of two previous studies (Tomlinson et al. 1997; Wijnen et al. 1997), in which mutations were detected in <6% and <8% of cases, respectively. However, two reasons can account for this. First, we used a combined approach of genomic sequencing and IVSP, instead of exon screening by SSCP and denaturing gradient-gel electrophoresis, used in these previous studies, which would not have detected mutations in patients 817 and 889. Second, the extremely early age at onset in our cohort may explain this discrepancy, since the average age of the probands was less than that in the previous studies.

The feasibility of offering commercially available testing for genetic diseases such as HNPCC, as well as the possible outcomes from the results (Burke et al. 1997), has been studied elsewhere (Plummer and Casey 1996). The current use of commercial testing and the relationship to genetic counseling for familial adenomatous polyposis also has been described elsewhere (Giardiello et al. 1997). Genetic testing for HNPCC could be offered to anyone who chooses to be tested, or testing could be restricted to individuals who can supply sufficient proof that they fulfill stringent familial inheritance criteria. The optimum testing criteria likely lie between these extremes, and the findings presented here suggest that consideration should be given to the offering of genetic testing to all index colon cancer patients <30 years of age.

Acknowledgments

We thank all patient and control subjects who participated in this study and acknowledge the considerable contribution made by the many clinicians who contributed to the assimilation of these cohorts. We are grateful to Rhona De Mey and Alison Fordyce, for genealogy work, and to Tracey Shipman and Sheila McBeath, for excellent technical support. This work was supported by grants from the Scottish Health Department (K/MRS/50/C2417), Cancer Research Campaign (SP2326/0101), Scottish Hospitals Endowment Research Trust (SHERT 1331), Edinburgh University Cancer Research Fund, and Tenovus Scotland. M.G.D. is a Medical Research Council Clinician Scientist Fellow.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank (for hMSH2 cDNA [U03911 and U04045] and hMLH1 cDNA [U07343 and U07418])

Genome Database, http://www.gdb.org/ (for hMSH2 [203983] and for hMLH1, including links to genomic intron/exon-boundary sequences [249617])

International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer, http://www.nfdht.nl/database/mdbchoice.htm (for the mismatch-repair–gene mutation database)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for colorectal cancer [MIM 114500] and HNPCC [MIM 120435 and 120436])

SEER Cancer Statistics Review, 1973–1994, http://www-seer.ims.nci.nih.gov/Publications/CSR7394/

References

Akiyama Y, Sato H, Yamada T, Nagasaki H, Tsuchiya A, Abe R, Yuasa Y (1997) Germ-line mutation of the hMSH6/GTPB

gene in an atypical hereditary nonpolyposis colorectal cancer kindred. Cancer Res 57:3920–3923

Borresen A-L, Lothe RA, Meling GI, Lystad S, Morrison P, Lipford J, Kane MF, et al (1995) Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinomas. Hum Mol Genet 4:2065–2072

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, et al (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary nonpolyposis colorectal cancer. Nature 368: 258–261

Bubb VJ, Curtis LJ, Cunningham C, Dunlop MG, Carothers AD, Morris RG, White S, et al (1996) Microsatellite instability and the role of hMSH2 in sporadic colorectal cancer. Oncogene 12:2641–2649

Buerstedde JM, Alday P, Torhorst J, Weber W, Muller H, Scott R (1995) Detection of new mutations in six out of 10 Swiss HNPCC families by genomic sequencing of the hMSH2 and hMLH1 genes. J Med Genet 32:909–912

Burke W, Petersen G, Lynch P, Botkin J, Daly M, Garber J, Ellis Khan MJ, et al (1997) Recommendations for follow-up care of individuals with an inherited predisposition to cancer. I. Hereditary nonpolyposis colon cancer. JAMA 227: 915–919

Dunlop MG, Campbell H (1997) Screening for people with a family history of colorectal cancer. Br Med J 314:1779–1780

Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, Liu B, et al (1997) Cancer risk associated with germline DNA mismatch repair gene mutations. Hum Mol Genet 6:105–110

Fishel RA, Lescoe MK, Rao MRS, Copland N, Jenkins N, Garber J, Kane M, et al (1993) The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colorectal cancer. Cell 75:1027–1038

Giardiello FM, Brensinger JD, Petersen GM, Luce MC, Hylind LM, Bacon JA, Booker SV, et al (1997) The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. N Engl J Med 336:823–827

Han H-J, Maruyama M, Baba S, Park J-G, Nakamura Y (1995) Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary nonpolyposis colorectal cancer (HNPCC). Hum Mol Genet 4:237–242

Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, et al (1997) Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair–defective human tumor cell lines. Cancer Res 57:808–811

Kohonen-Corish M, Ross VL, Doe WF, Kool DA, Edkins E, Faragher I, Wijnen J, et al (1996) RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. Am J Hum Genet 59:818–824

Kolodner RD, Hall NR, Lipford J, Kane MF, Morrison PT, Finan PJ, Burn J, et al (1995) Structure of the human MLH1 locus and analysis of a large hereditary nonpolyposis colorectal carcinoma kindred for MLH1 mutations. Cancer Res 55:242–248

Kolodner RD, Hall NR, Lipford J, Kane MF, Rao MRS, Morrison P, Wirth L, et al (1994) Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for MSH2 mutations. Genomics 24:516–526

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, et al (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 75: 1215–1225

Liu B, Farrington SM, Petersen GM, Hamilton SR, Parsons R, Papadopoulos N, Fujiwara T, et al (1995) Genetic instability occurs in the majority of young patients with colorectal cancer. Nat Med 1:348–352

Liu B, Parsons RE, Hamilton SR, Petersen GM, Lynch HT, Watson P, Markowitz S, et al (1994) hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds. Cancer Res 54:4590–4594

Liu B, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, Jass JR, et al (1996) Analysis of mismatch repair genes in hereditary nonpolyposis colorectal cancer patients. Nat Med 2:169–174

Lothe RA, Peltomaki P, Meling GI, Aaltonen LA, Nystrom-Lahti M, Pylkkanen L, Heimdal K, et al (1993) Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. Cancer Res 53: 5849–5852

Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, Cavalieri RJ, et al (1993) Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology 104: 1535–1549

Mary J-L, Bishop T, Kolodner R, Lipford JR, Kane M, Weber W, Torhorst J, et al (1994) Mutational analysis of the hMSH2 gene reveals a three base pair deletion in a family predisposed to colorectal cancer development. Hum Mol Genet 3:2067–2069

Mauillon JL, Michel P, Limacher J-M, Latouche J-B, Dechelotte P, Charbonnier F, Martin C, et al (1996) Identification of novel germline hMLH1 mutations including a 22 kb Alu-mediated deletion in patients with familial colorectal cancer. Cancer Res 56:5728–5733

Moslein G, Tester DJ, Lindor NM, Honchel R, Cunningham JM, French AJ, Hailing KC, et al (1996) Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. Hum Mol Genet 5:1245–1252

Nicolaides NC, Papadopoulos BL, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, et al (1994) Mutations of two PMS homologues in hereditary nonpolyposis colorectal cancer. Nature 371:75–80

Nystrom-Lahti M, Kristo P, Nicolaides NC, Chang S-Y, Aaltonen LA, Moisio A-L, Jarvinen HJ, et al (1995) Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat Med 1:1203–1206

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, et al (1994) Mutation of a mutL homolog in hereditary colon cancer. Science 263: 1625–1629

Parker LT, Zakeri H, Deng Q, Spurgeon S, Kwok P-Y, Nickerson DA (1996*a*) AmpliTaq DNA polymerase, FS dye-terminator sequencing: analysis of peak height patterns. BioTechniques 21:694–699

Parker SL, Tong T, Bolden S, Wingo PA (1996*b*) Cancer statistics, 1996. CA Cancer J Clin 46:5–27

Plummer SJ, Casey G (1996) Are we any closer to genetic testing for common malignancies? Nat Med 2:156–158

Scherer SJ, Welter C, Zang K-D, Dooley S (1996) Specific *in vitro* binding of p53 to the promoter region of the human mismatch repair gene hMSH2. Biochem Biophys Res Commun 221:722–728

Sharp L, Black RJ, Harkness EF, Finlayson AR, Muir CS (1993) Cancer registration statistics Scotland 1981–1990. National Health Service in Scotland, Information and Statistics Division, Edinburgh

Tannergard P, Lipford JR, Kolodner R, Frodin JE, Nordenskjold M, Lindblom A (1995) Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. Cancer Res 55:6092–6096

Tomlinson IPM, Beck NE, Homfray T, Harocopos CJ, Bodmer WF (1997) Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. J Med Genet 34:39–42

Vasen HFA, Mecklin J-P, Meera Khan P, Lynch HT (1991) The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 34:424–425

Vasen HF, Wijnen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, Nagengast FM, et al (1996) Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 110: 1020–1027

Wijnen J, Fodde R, Meera Khan P (1994) DGGE polymorphism in intron 10 of MSH2, the HNPCC gene. Hum Mol Genet 3:2268

Wijnen J, Meera Khan P, Vasen H, van der Klift H, Mulder A, van Leeuwen–Cornelisse I, Bakker B, et al (1997) Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. Am J Hum Genet 61:329–335

Wijnen J, Vasen H, Meera Khan P, Menko FH, van der Klift H, van Leeuwen C, van den Broek M, et al (1995) Seven new mutations in *hMSH2,* an HNPCC gene, identified by denaturing gradient-gel electrophoresis. Am J Hum Genet 56:1060–1066