HUMAN MUTATION 26(3), 255–261, 2005

RESEARCH ARTICLE

The 5′ Region of the *MSH2* Gene Involved in Hereditary Non-Polyposis Colorectal Cancer Contains a High Density of Recombinogenic Sequences

Françoise Charbonnier,[1,2] Stephanie Baert-Desurmont,[1,2] Ping Liang,[3] Frederic Di Fiore,[1] Cosette Martin,[2] Stephanie Frerot,[2] Sylviane Olschwang,[4] Qing Wang,[5] Marie-Pierre Buisine,[6] Brigitte Gilbert,[7] Mef Nilbert,[8] Annika Lindblom,[9] and Thierry Frebourg[1,2]*

[1]Inserm U614-IFRMP, Faculty of Medicine, Rouen, France; [2]Department of Genetics, Rouen University Hospital, Rouen, France; [3]Department of Genetics, Roswell Park Cancer Institute, Buffalo, New York; [4]Inserm U434 and Saint Antoine Hospital, Paris, France; [5]Molecular Oncology, Centre Léon Bérard, Lyon, France; [6]Laboratory of Biochemistry and Molecular Biology, Lille University Hospital, Lille, France; [7]Department of Genetics, Poitiers University Hospital, Poitiers, France; [8]Department of Oncology, University Hospital, Lund, Sweden; [9]Department of Clinical Genetics, Karolinska Hospital, Stockholm, Sweden

Communicated by Richard G.H. Cotton

***MSH2* rearrangements are involved in approximately 10% of hereditary non-polyposis colorectal cancer (HNPCC) families, and in most of the rearrangements, exon 1 is deleted. We scanned by quantitative multiplex polymerase chain reaction (PCR) of short fluorescent fragments (QMPSF) 200 kb of genomic sequences upstream of the *MSH2* transcription initiation site in 21 HNPCC families with exon 1 deletions. This QMPSF scan revealed 12 distinct 5′ breakpoints located up to 200 kb upstream of the *MSH2* transcription initiation site. Sequencing analysis of the rearranged allele in 17 families revealed that most of the deletions (15/17) resulted from homologous *Alu*-mediated recombination. QMPSF and sequencing analysis in these 21 families led us to detect the presence of 20 distinct 5′ breakpoints. In 14 out of 15 *Alu*-mediated recombinations, we found, either within the identical region in which the recombination had probably occurred or in its vicinity, the 26-bp *Alu* core sequence containing the recombinogenic *Chi*-like motif. Compared to the equivalent regions of other human genes, the *MSH2* upstream region was found to contain a high density of *Alu* repeats (30% within 228 kb and 43% within 50 kb), most of which belong to the old *Alu S* subfamilies. In conclusion, this study demonstrates the heterogeneity of the breakpoints within the *MSH2* upstream region and reveals the remarkable density of recombinogenic *Alu* sequences in this region. Hum Mutat 26(3), 255–261, 2005.**
© 2005 Wiley-Liss, Inc.

KEY WORDS: HNPCC; *MSH2*; QMPSF; deletion; *Alu*; repeat

INTRODUCTION

Hereditary non-polyposis colorectal cancer (HNPCC, MIM# 114500), which represents the most common form of inherited colorectal cancer, results from germline alterations of the mismatch repair genes, *MSH2* (MIM# 120435), *MLH1* (MIM# 120436), and *MSH6* (MIM# 600678) (for review see Lucci-Cordisco et al. [2003]). We have previously shown, using quantitative multiplex polymerase chain reaction (PCR) of short fluorescent fragments (QMPSF) [Charbonnier et al., 2000; 2002], that *MSH2* rearrangements are involved in at least 10% of the HNPCC families fulfilling the Amsterdam (AMS) criteria, which justifies the inclusion of their search in the routine diagnosis of HNPCC [Charbonnier et al., 2002; Di Fiore et al., 2004]. In most of the studies, the frequency of *MSH2* rearrangements has been shown to be higher than that of *MLH1*: we found a genomic rearrangement of *MSH2* in 27 out of 120 (22%) and of *MLH1* in 6/86 (7%) HNPCC AMS+ families without point mutations within *MSH2* and *MLH1* [Charbonnier et al., 2002]. This difference in the mutational spectrum of *MSH2* and *MLH1* was also documented using multiplex ligation-dependent probe amplification (MPLA), both by Nakagawa et al. [2003] who detected, among 70 HNPCC families with MSI positive tumors and without point mutations, four different *MSH2* deletions and a single case of *MLH1* rearrangement, and more recently by Bunyan et al. [2004], who detected among 122 HNPCC families without point mutations seven cases of *MSH2* deletions and no *MLH1* rearrangement. In 52 AMS+ HNPCC families without point

Received 6 December 2004; accepted revised manuscript 28 April 2005.

*Correspondence to: Pr. Thierry Frebourg, Inserm U614-IFRMP, Faculty of Medicine, 22 Boulevard Gambetta, 76183 Rouen, France. E-mail: Frebourg@chu-rouen.fr

Grant sponsor: l'Association pour la Recherche sur le Cancer; Grant sponsor: La Ligue Nationale de Lutte Contre le Cancer.
DOI 10.1002/humu.20216
Published online in Wiley InterScience (www.interscience.wiley.com).

© 2005 WILEY-LISS, INC.

mutations, Wang et al. [2003] reported similar detection rates of rearrangements for *MSH2* and *MLH1* (6/52, 12%), but four patients harbored the same *MLH1* exon 1–10 deletion, potentially associated to a founder effect, which may have led to an overestimation of the relative contribution of *MLH1* rearrangements in HNPCC.

MSH2 deletions present a remarkable heterogeneity: we have reported, in 41 HNPCC families, 19 distinct *MSH2* exonic deletions, removing exon(s) 1, 1–2, 1–4, 1–6, 1–7, 1–8, 1–11, 1–15, 2, 3, 4–6, 5, 5–6, 7, 7–10, 8, 9–10, 12–13, and 13–15 [Charbonnier et al., 2002; Di Fiore et al., 2004] and we have subsequently detected, in 10 other families, five additional *MSH2* deletions removing exons 4, 3–11, 8–16, 9–16, and 11–14. The majority of the *MSH2* deletions reported so far encompass exon 1 [Wang et al., 2003; Taylor et al., 2003; Nakagawa et al., 2003; Di Fiore et al., 2004; Bunyan et al., 2004]. QMPSF analysis of the promoter region revealed that, in these cases, the promoter was either partially or completely deleted and an initial QMPSF scanning of the *MSH2* 5′ region revealed the presence of at least seven distinct 5′ breakpoints [Di Fiore et al., 2004]. We have now performed a fine QMPSF analysis of 200 kb of genomic sequences upstream of the *MSH2* transcription initiation site in families with a *MSH2* rearrangement removing exon 1, in order to localize and characterize accurately the 5′ breakpoints.

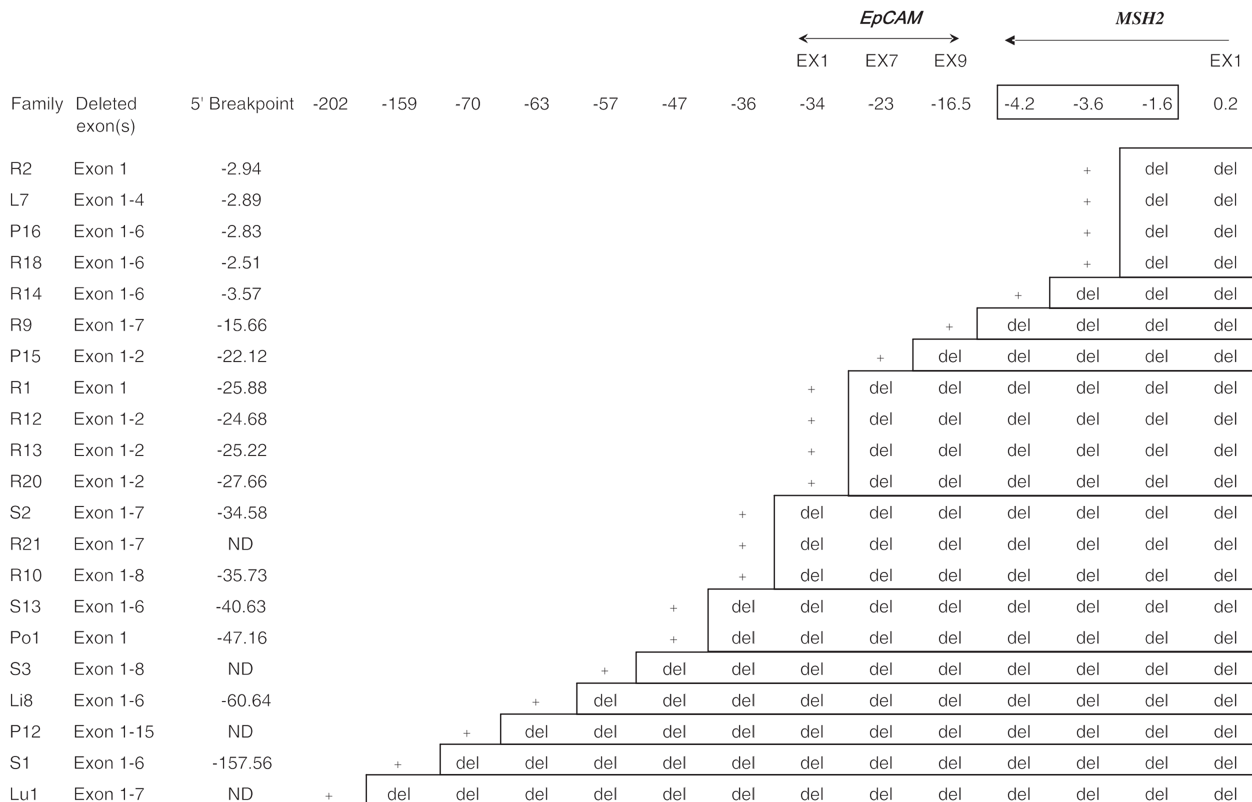
MATERIALS AND METHODS

Families

This study included 21 families (Fig. 1), 12 of which strictly fulfilled the AMS criteria (at least three relatives with colorectal cancer, cancer of the endometrium, small bowel, ureter, or renal pelvis, one of whom is a first-degree relative of the other two; at least two successive generations affected; and at least one cancer diagnosed before age 50 [Vasen et al., 1999]). In each family, QMPSF analysis of the 16 exons of *MSH2* [Charbonnier et al., 2000] had previously revealed an heterozygous genomic deletion encompassing exon 1.

QMPSF Analysis of the *MSH2* 5′ Region

Short genomic fragments (between 116 and 294 bp) of the *MSH2* promoter and upstream region were PCR amplified, using 31 6-Fam labeled primer pairs (primer sequences are available upon request), in several multiplex PCRs. Exon 1 and an additional fragment, corresponding to another gene used as a control, were systematically included in each multiplex PCR. Primers were designed in order to have a T_m comprised between 50°C and 60°C. To clean up DNA before PCR, 1 μl of genomic DNA (100 ng/μl) was preincubated before PCR with 5 μl of Generelease^{TM} (Bioventure, Inc., Murfreesboro, TN), 15 seconds at 65°C, 15 seconds at 8°C, and 45 seconds at 65°C, 1 minute 30



											EpCAM				MSH2	
										EX1	EX7	EX9				EX1
Family	Deleted exon(s)	5' Breakpoint	-202	-159	-70	-63	-57	-47	-36	-34	-23	-16.5	-4.2	-3.6	-1.6	0.2
R2	Exon 1	-2.94												+	del	del
L7	Exon 1-4	-2.89												+	del	del
P16	Exon 1-6	-2.83												+	del	del
R18	Exon 1-6	-2.51												+	del	del
R14	Exon 1-6	-3.57											+	del	del	del
R9	Exon 1-7	-15.66										+	del	del	del	del
P15	Exon 1-2	-22.12									+	del	del	del	del	del
R1	Exon 1	-25.88								+	del	del	del	del	del	del
R12	Exon 1-2	-24.68								+	del	del	del	del	del	del
R13	Exon 1-2	-25.22								+	del	del	del	del	del	del
R20	Exon 1-2	-27.66								+	del	del	del	del	del	del
S2	Exon 1-7	-34.58							+	del	del	del	del	del	del	del
R21	Exon 1-7	ND							+	del	del	del	del	del	del	del
R10	Exon 1-8	-35.73							+	del	del	del	del	del	del	del
S13	Exon 1-6	-40.63						+	del	del	del	del	del	del	del	del
Po1	Exon 1	-47.16						+	del	del	del	del	del	del	del	del
S3	Exon 1-8	ND					+	del	del	del	del	del	del	del	del	del
Li8	Exon 1-6	-60.64				+	del	del	del	del	del	del	del	del	del	del
P12	Exon 1-15	ND			+	del	del	del	del	del	del	del	del	del	del	del
S1	Exon 1-6	-157.56		+	del	del	del	del	del	del	del	del	del	del	del	del
Lu1	Exon 1-7	ND	+	del	del	del	del	del	del	del	del	del	del	del	del	del

FIGURE 1. **Localization of the 5′ breakpoints of *MSH2* rearrangements removing exon 1. Numbers in the upper line indicate the approximate locations (in kb) of QMPSF amplicons numbered according to the transcription initiation site of *MSH2*. We have not included all the QMPSF amplicons used in this study. The box in the upper line delimits the *MSH2* promoter region, as defined by Iwahashi et al. [1998]. Exon 1 of *MSH2* and exons 1, 7, and 9 of *TACSTD1* are indicated above this line. Families showing the same QMPSF profile are boxed: del indicates a deletion detected by QMPSF, whereas + indicates an amplicon that is not deleted. The third column shows the accurate location of the 5′ breakpoint (numbered in kb according to the first nucleotide of the translation initiation codon of *MSH2*), as determined by sequencing; ND: not determined. Several breakpoints have been previously shown in Di Fiore et al. [2004].**

seconds at 97°C, 30 seconds at 8°C, 1 minute 30 seconds at 65°C, 30 seconds at 97°C, 30 seconds at 65°C, then maintained at 4°C. PCRs were performed in a final volume of 25 μl containing 100 ng of genomic DNA, 0.2–2 μM of each pair of primers, and 1 unit of Thermoprime plus DNA polymerase (Abgene, Epson, UK). After electrophoresis for 3 hr on an automated sequencer (Applied Biosystems, Foster City, CA), data were analyzed using the Gene scanner Model 672 Fluorescent Fragment Analyzer (Applied Biosystems). Electropherograms were superimposed to those generated from control DNAs and the areas of the corresponding peaks between the different samples were visually compared. In QMPSF, a 50% reduction of the peak(s) area indicates a heterozygote deletion of the corresponding genomic fragment [Charbonnier et al., 2000; 2002; Di Fiore et al., 2004]. Each positive result was controlled in a second independent fluorescent multiplex PCR.

Long-Range PCR

In each family, the rearranged allele was PCR-amplified using the Expand™ Long Template PCR system from Roche Diagnostics (Mannheim, Germany), according to the manufacturer's protocol. The PCR consisted of 10 cycles of 10 seconds at 92°C, 30 seconds at 65°C (with a 1°C decrease/cycle), and 10 minutes at 68°C followed by 30 cycles of 10 seconds at 92°C, 30 seconds at 55°C, and 10 minutes and 20 seconds at 68°C. Primer sequences are presented in Table 1.

TABLE 1. **PCR Amplification of the Rearranged Allele in 18 HNPCC Families Harboring a *MSH2* Deletion Encompassing Exon 1**

Family	Deleted exon(s)	Sense primer		Antisense primer		Size of the amplified allele
		Sequence	Location[a]	Sequence	Location[a]	
R2	Exon 1	5′ aacacatcttctttctcagtatacagtctt 3′	–3265	5′ atgccaaataccaatcattc 3′	+5358	4.0 kb
L7	Exons 1–4	5′ tatcagaaatggtagtagcttctctaaagg 3′	–3658	5′ aaaggttaagggctctgact 3′	+11221	1.5 kb
P16	Exons 1–6	5′ tatcagaaatggtagtagcttctctaaagg 3′	–3658	5′ accaccaccaactttatgag 3′	+26836	7 kb
R18	Exons 1–6	5′ tatcagaaatggtagtagcttctctaaagg 3′	–3658	5′ accaccaccaactttatgag 3′	+26836	2 kb
R14	Exons 1–6	5′ ccttaaacacctggacttaaggaatttttc 3′	–4181	5′ accaccaccaactttatgag 3′	+26836	3 kb
R9	Exons 1–7	5′ cagataaaggagatgggtgag 3′	–16623	5′ agaggagtcacaaaaactgc 3′	+42398	1.5 kb
P15	Exons 1–2	5′ atggacctgacagtaaatgg 3′	–23390	5′ ggtaaaacacattcctttgg 3′	+7140	1.5 kb
R1	Exon 1	5′ acagtgcacttcagaaggag 3′	–26182	5′ atgccaaataccaatcattc 3′	+5358	2.5 kb
R12	Exons 1–2	5′ agcctttgaaaagcctagtt 3′	–25622	5′ ggtaaaacacattcctttgg 3′	+7140	2 kb
R13	Exons 1–2	5′ agcctttgaaaagcctagtt 3′	–25622	5′ ggtaaaacacattcctttgg 3′	+7140	1 kb
R20	Exons 1–2	5′ gagaagagcaaaacctgaag 3′	–29332	5′ ggtaaaacacattcctttgg 3′	+7140	2.3 kb
S2	Exons 1–7	5′ ttctcccgattaggtaaacaga 3′	–36153	5′ tttctgcattcacctagtgttg 3′	+35810	5.5 kb
R10	Exons 1–8	5′ ttctcccgattaggtaaacaga 3′	–36153	5′ tttgtgaactcccttatctgct 3′	+49729	1 kb
S13	Exons 1–6	5′ gcttctgaattctgagtgatctgta 3′	–48605	5′ accaccaccaactttatgag 3′	+26836	9 kb
PO1	Exon 1	5′ gcttctgaattctgagtgatctgta 3′	–48605	5′ atgccaaataccaatcattc 3′	+5358	5.5 kb
Li8	Exons 1–6	5′ attgatttagggaggaacacct 3′	–63116	5′ cctttggtccagatatcattaaaag 3′	+19835	3 kb
S1	Exons 1–6	5′ ttctaaaagctcaggaccacat 3′	–158866	5′ ttttgactcataaagccatcct 3′	+15653	1.7 kb
LU1	Exons 1–7	5′ cccaaacagaagtgaagatagc 3′	–203624	5′ cttaaggccaatactccctctt 3′	+31593	11 kb

[a]Numbered according to the first nucleotide of the translation initiation codon of *MSH2* (build 35.1, NT_022184.14 (26218482–26526582)).

TABLE 2. **Description of *MSH2* Genomic Rearrangements Encompassing Exon 1**

Family	Deleted exon(s)	Estimated size of the deletion	5′ repeat element[a]	3′ repeat element[a]	Tract of sequence identity	Location of the 26-bp *Alu* core sequence
R2[c]	Exon 1	4.6 kb	*Alu Sg*	*Alu Sx*	24 bp	66 bp dowstream the breakpoint[b]
L7	Exons 1–4	12.6 kb	*Alu Sg*	–	–	–
P16	Exons 1–6	23.3 kb	*Alu Sg*	*Alu Sq*	11 bp	1 bp upstream the breakpoint
R18	Exons 1–6	27.6 kb	*Alu Sp*	*Alu Sc*	17 pb	31 bp upstream the breakpoint
R14	Exons 1–6	28 kb	*Alu Sx*	*AluY*	31 bp	at the breakpoint
R9[c]	Exons 1–7	57.4 kb	*Alu Sg/x*	*Alu Sx*	21 bp	72 bp dowstream the breakpoint
P15	Exons 1–2	28.8 kb	*Alu Sx*	*Alu Sg*	22 bp	208 bp upstream the breakpoint
R1	Exon 1	28.8 kb	*Alu Sx*	*Alu Sc*	4 bp	43 bp dowstream the breakpoint
R12	Exons 1–2	30.7 kb	*Alu Sp*	*Alu Sq/x*	19 bp	25 bp dowstream the breakpoint
R13	Exons 1–2	31.7 kb	*AluY*	*Alu Sg*	35 bp	at the breakpoint
R20	Exons 1–2	34 kb	*L1*	*Alu Sc*	–	–
S2	Exons 1–7	66.5 kb	*Alu Sp*	*Alu Sp*	None[d]	–
R10	Exons 1–8	85 kb	*Alu Sq*	*AluY*	7 bp	31 bp dowstream the breakpoint
S13	Exons 1–6	65 kb	*AluY*	*AluY*	15 bp	16 bp upstream the breakpoint
Po1	Exon 1	49 kb	*Alu Sp*	*Alu Sx*	16 bp	at the breakpoint
Li8	Exons 1–6	80 kb	*Alu Sg*	*Alu Sg*	29 pb	62 pb upstream the breakpoint
S1	Exons 1–6	172 kb	*Alu Sx*	*AluJo*	25 bp	16 bp upstream the breakpoint

[a]In all cases, the *Alu* sequences were in the same orientation.
[b]The breakpoint was predicted to occur within the sequence identity tract.
[c]The rearrangement in this family has partially been described in Charbonnier et al. [2002].
[d]In this family, the rearranged allele furthermore contains an insertion; nucleotide sequences alignment suggests that the breakpoint had occurred within this insertion.

Sequence Analysis

After electrophoresis on low-melt agarose gel, PCR products were purified using the Qiaquick Gel Extraction kit (Qiagen SA, Courtabeuf, France) and directly sequenced using the Big Dye V3.0 Terminator Kit (Applied Biosystems) and an automated sequencer (Applied Biosystems).

Bioinformatics

Alignments of the nucleotidic sequences of the rearranged and wild-type alleles were achieved using the Clustal algorithm from BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Search for repeat motifs within the *MSH2* locus (working draft sequence, Build 35.1, NT_022184.14 [26218482-26526582]) was performed using the Repeat Masker Program (http://repeatmasker.genome.washington.edu). Survey of *Alu* sequence density for the 50 kb and 5 kb upstream regions of *MSH2* gene and of all the 24,336 human genes annotated in the NCBI GenBank Build 35 (http://www.ncbi.nlm.nih.gov) was performed using in house PERL-based computer scripts. The statistic P-values for the differences in *Alu* sequence densities between *MSH2* and all the human genes were calculated using Z scores of the *Alu* density distributions.

RESULTS

We first analyzed by QMPSF the *MSH2* promoter region and a stretch of 200 kb upstream of the transcription initiation site in 21 HNPCC families harboring different types of exonic deletions that remove exon 1 (Fig. 1). This analysis of the 5′ region was performed using a total of 31 QMPSF amplicons. We also used QMPSF to delineate the 3′ boundary of the breakpoint, using the specific amplicon for each exon [Charbonnier et al., 2000] and then appropriate amplicons for the downstream intron. The QMPSF profiles revealed, among these 21 HNPCC families, the presence of at least 12 different 5′ breakpoints scattered between –1.6 kb and –202 kb upstream of the transcription initiation site of *MSH2* (Fig. 1). In most of the families (16/21), the 5′ breakpoint was located upstream of the *MSH2* promoter region. In 15 families, the coding region of the *TACSTD1* gene (MIM# 185535), which is located –16 kb upstream of *MSH2* and

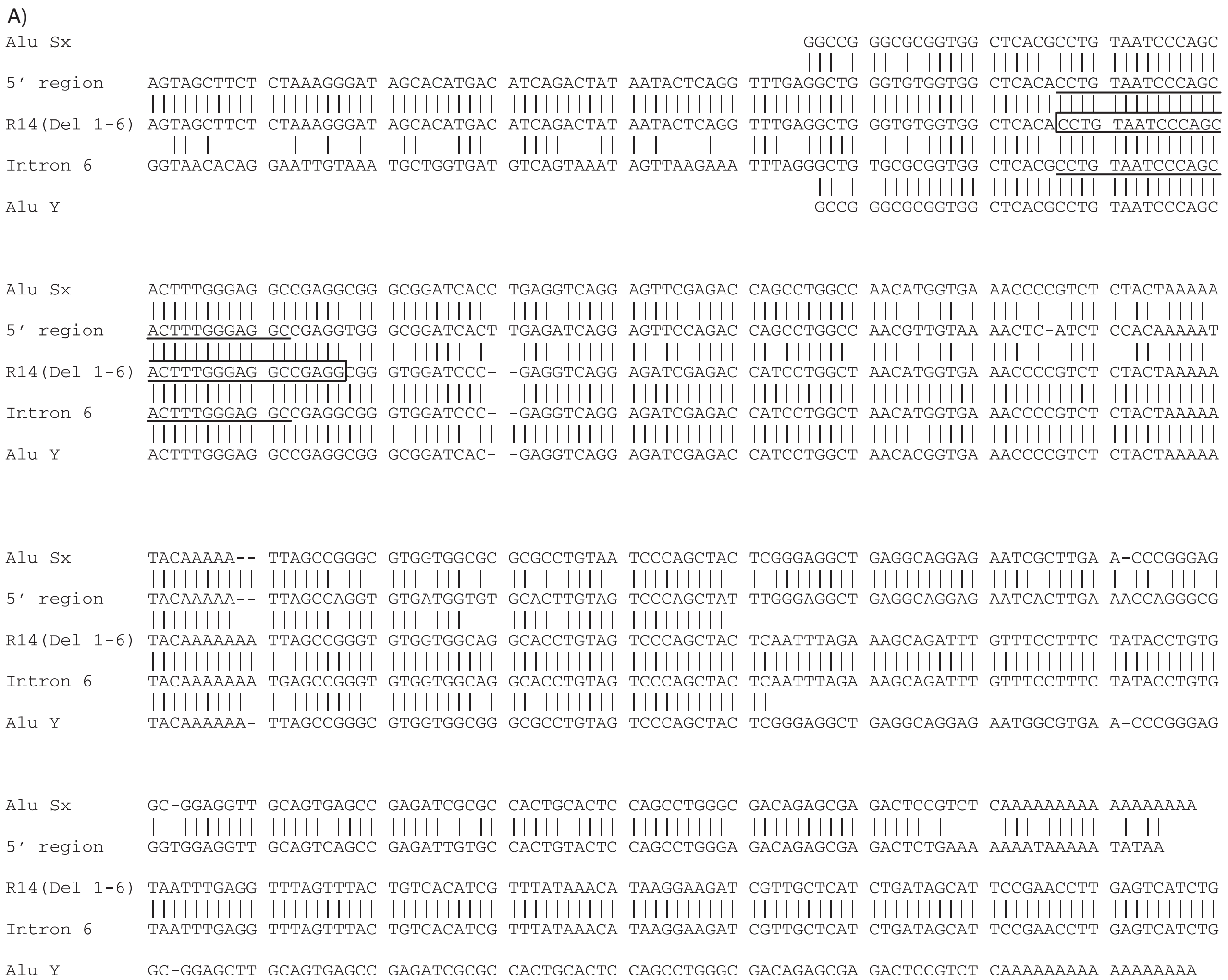


FIGURE 2. **Schematic representation of *MSH2 Alu* mediated rearrangements involving exon 1. Alignment between the nucleotide sequences of the normal *MSH2* 5′ region, the deleted allele, the normal intron, and the *Alu* consensus sequences. The open box indicates the homologous region in which the recombination has occurred. The *Alu* core sequences containing the recombinogenic *Chi*-like motif are underlined. A: Representative recombination occurring within the *Alu* core sequence detected in Family R14 harboring a *MSH2* exon 1–6 deletion; 5′ region: −3645 to −3330 (numbered from the first nucleotide of the translation initiation codon of *MSH2*); intron 6: c1076+11028 to c1076+11385 (NM_000251). B: Representative recombination occurring at the vicinity of the *Alu* core sequence detected in Family R12 harboring a *MSH2* exons 1–2 deletion; 5′ region : –24845 to –24579 (corresponding to intron 5 of the *TACSTD1/EpCAM* gene); intron 2: c.366+571 to c.366+794.**

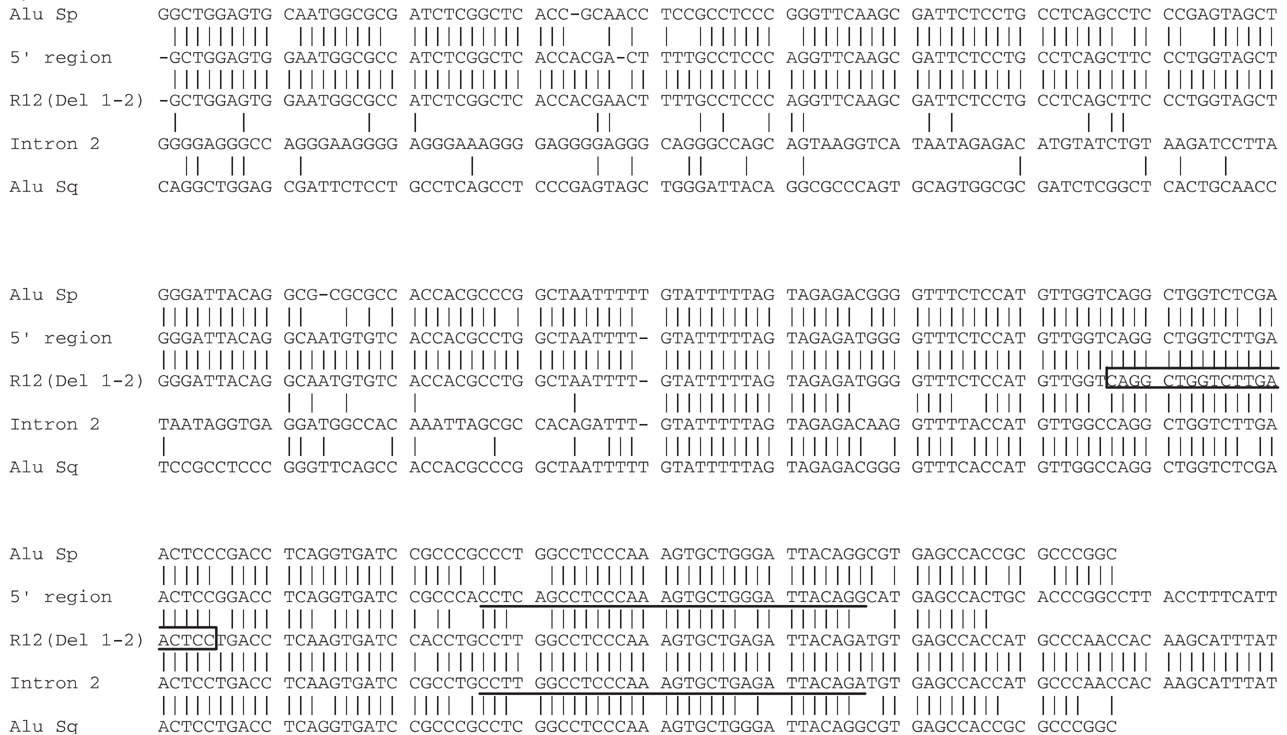


FIGURE 2. **(Continued)**

encodes the membrane glycoprotein EpCAM involved in homophilic adhesion, was either partially or completely deleted (Fig. 1). On the basis of the QMPSF results, we then designed PCR primers to amplify the deleted allele and we achieved successful PCR-amplification of the rearranged allele in 18/21 families (Table 1). In 17 of these families, we were able to sequence the rearranged allele and to identify the 5′ and 3′ breakpoints (Table 2). Taken together, QMPSF mapping and sequencing analysis revealed, among the 21 families, the presence of 20 distinct 5′ breakpoints (Fig. 1).

Among the 17 families in which we sequenced the deleted allele, the alignment between the nucleotide sequences of the normal *MSH2* 5′ region, of the rearranged allele and of the normal intronic sequences revealed that in 15 cases the rearrangement had clearly resulted from an homologous *Alu* mediated recombination (Table 2). Most of the *Alu* elements spanning the breakpoints derived from the *Alu S* subfamilies (Table 2). In 14 *Alu*-mediated recombinations, we found, either within the identical region in which the recombination had probably occurred (Fig. 2A) or at its vicinity (Fig. 2B) the 26 bp *Alu* core sequence (Table 2), containing a part of the prokaryotic *Chi* sequence [Rudiger et al., 1995].

Search for *Alu* repeats within the 228 kb upstream region of *MSH2* revealed that this region contains a remarkably high density of *Alu* repeats (252 *Alu* repeats corresponding to 30.15% of genomic sequence), the majority belonging to the *Alu* S subfamilies (170/252). The *Alu* density gradually increased towards the *MSH2* gene, the 50-kb and 5-kb upstream regions containing 43% and 56% of *Alu* sequences, respectively (Table 3).

TABLE 3. **Comparison of the Density of *Alu* Repeats Between the 5′ Region of *MSH2* and Other Human Genes**

5′ region	*MSH2*	All human genes[a]	
228 kb[b]	30%	–	–
50 kb	43%	15%	P < 0.005
Alu S subfamily	34%	10%	P < 0.0005
5 kb	56%	14%	P < 0.001
Alu S subfamily	33%	13%	P < 0.01

[a]Mean *Alu* sequence density for the equivalent regions of the 24336 human genes annotated in the NCBI GenBank Build 35.
[b]Upstream of the transcription initiation site.

DISCUSSION

The recently developed PCR-based methods, QMPSF and MPLA, which allow easy and reliable detection on a routine basis of heterozygous genomic deletions and duplications are highlighting the contribution of these alterations in numerous genetic diseases. QMPSF analysis of the *MMR* genes has previously allowed us to document the prevalence of *MSH2* genomic deletions in HNPCC families [Charbonnier et al., 2002; Di Fiore et al., 2004]. In this study, we exploited the power of QMPSF for scanning previously unexplored regions and mapped the 5′ breakpoints of *MSH2* deletions encompassing exon 1. We detected 20 different 5′ breakpoints located up to 200 kb upstream of the *MSH2* transcription initiation site. In most of the cases (15/21) the deletion also removed partially or totally the *TACSTD1/EpCAM* gene. In agreement with a previous study reporting the detection of a *MSH2* deletion encompassing 47 kb of the 5′ region and including *TACSTD1* [Nakagawa et al., 2003], we did not observe a specific clinical presentation in the HNPCC families harboring a *TACSTD1* deletion. EpCAM is an antigen that is highly expressed in epithelial carcinomas, suggesting that its activation but not its inactivation could be associated with oncogenesis [Munz et al., 2004].

As previously reported [Charbonnier et al., 2002; Wagner et al., 2003; Wang et al., 2003; Nakagawa et al., 2003], the great majority of the *MSH2* rearrangements (15/17) was shown to result from homologous unequal *Alu*-mediated recombinations, the main mechanism of genomic deletions observed in Mendelian diseases [Batzer and Deininger, 2002; Abeysinghe et al., 2003]. The completion of the human genome sequencing has revealed that the *Alu* elements represent 10% of the genome and include 1.2 million copies and that these *Alu* repeats preferentially accumulate in gene-rich sequences [Lander et al., 2001; for review see Batzer and Deininger, 2002]. The present report reveals the remarkable density of *Alu* repeats within the 5′ region of *MSH2* (Table 3). This density is significantly higher than the mean density of *Alu* repeats for the equivalent 5′ regions of human genes (Table 3). We had already underlined the high density of *Alu* repeats within the transcribed region *MSH2* [Charbonnier et al., 2002]; indeed, the *MSH2* transcribed region (80 kb) contains 101 *Alu* repeats corresponding to 33% of genomic sequence, the majority of which are located within the first eight introns. The density of *Alu* repeats within the 5′ region explains why, in the majority of the families harboring a *MSH2* rearrangement, exon 1 is deleted.

We observed that most of the 5′ region *Alu* repeats involved in *MSH2* deletions belong to the *Alu S* subfamilies (Table 2). The density of *Alu S* repeats in the *MSH2* 5′ region is significantly higher than that observed in other human genes (Table 3). The *Alu J* and *Alu S* subfamilies have been defined as the oldest and intermediate *Alu* groups [Jurka and Smith, 1988], respectively, whereas the younger *Alu Y* group includes the most actively retrotransposed *Alu* subfamilies (for review see Jurka [2004]). Interestingly, in silico analysis of the human genome reveals that old *Alu* elements accumulate in the gene-rich regions, particularly the promoter regions, whereas the location of the new *Alu* elements is biased towards intergenic regions (Wang et al., submitted). This could explain the predominant role of old *Alu* sequences in intragenic recombinations, illustrated by our study, whereas *Alu* insertions mostly involve more recent *Alu* members [Deininger and Batzer, 1999]. In the *Alu*-mediated recombinations that we characterized, the tract's length of sequence identity, in which the recombination had probably occurred, varies from 4 to 35 bp. Interestingly, in 14 out of 15 *Alu*-mediated recombinations studied, we found the 26 bp core sequence previously associated with homologous and non-homologous recombination breakpoints [Rudiger et al., 1995]. This sequence contains the pentanucleotide CCAGC, corresponding to a part of the prokaryotic *Chi* sequence involved in homologous recombination in *E. coli* [Rudiger et al., 1995], and may by itself promote DNA exchange and homologous recombination [Kolomietz et al., 2002].

In conclusion, this study demonstrates the remarkable heterogeneity of 5′ breakpoints in the *MSH2* distributed over 200 kb upstream of the transcription initiation site, which can be explained by the high density of *MSH2* 5′ region in recombinogenic *Alu* sequences.

ACKNOWLEDGMENTS

We are grateful to Mario Tosi for helpful discussion and critical review of the manuscript, and to the clinicians who referred HNPCC families.

REFERENCES

Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN. 2003. Translocation and gross deletion breakpoints in human inherited disease and cancer. I: nucleotide composition and recombination-associated motifs. Hum Mutat 22:229–244.

Batzer MA, Deininger PL. 2002. *Alu* repeats and human genomic diversity. Nat Rev Genet 3:370–379.

Bunyan DJ, Eccles DM, Sillibourne J, Wilkins E, Thomas NS, Shea-Simonds J, Duncan PJ, Curtis CE, Robinson DO, Harvey JF, Cross NC. 2004. Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. Br J Cancer 13;91:1155–1159.

Charbonnier F, Raux G, Wang Q, Drouot N, Cordier F, Limacher JM, Saurin JC, Puisieux A, Olschwang S, Frebourg T. 2000. Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. Cancer Res 60:2760–2763.

Charbonnier F, Olschwang S, Wang Q, Boisson C, Martin C, Buisine MP, Puisieux A, Frebourg T. 2002. MSH2 in contrast to MLH1 and MSH6 is frequently inactivated by exonic and promoter rearrangements in hereditary nonpolyposis colorectal cancer. Cancer Res 62:848–853.

Deininger PL, Batzer MA. 1999. *Alu* repeats and human disease. Mol Genet Metab 67:183–193.

Di Fiore F, Charbonnier F, Martin C, Frerot S, Olschwang S, Wang Q, Boisson C, Buisine MP, Nilbert M, Lindblom A, Frebourg T. 2004. Screening for genomic rearrangements of the MMR genes must be included in the routine diagnosis of HNPCC. J Med Genet 41:18–20.

Iwahashi Y, Ito E, Yanagisawa Y, Akiyama Y, Yuasa Y, Onodera T, Maruyama K. 1998. Promoter analysis of the human mismatch repair gene hMSH2. Gene 213:141–147.

Jurka J. 2004. Evolutionary impact of human *Alu* repetitive elements. Curr Opin Genet Dev 14:603–608.

Jurka J, Smith T. 1988. A fundamental division in the *Alu* family of repeated sequences. Proc Natl Acad Sci USA 85:4775–4778.

Kolomietz E, Meyn MS, Pandita A, Squire JA. 2002. The role of *Alu* repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes Chromosomes Cancer 35:97–112.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR,

Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Lucci-Cordisco E, Zito I, Gensini F, Genuardi M. 2003. Hereditary nonpolyposis colorectal cancer and related conditions. Am J Med Genet 122A:325–334.

Munz M, Kieu C, Mack B, Schmitt B, Zeidler R, Gires O. 2004. The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. Oncogene 23:5748–5758.

Nakagawa H, Hampel H, de la Chapelle A. 2003. Identification and characterization of genomic rearrangements of MSH2 and MLH1 in Lynch syndrome (HNPCC) by novel techniques. Hum Mutat 22:258.

Rudiger NS, Gregersen N, Kielland-Brandt MC. 1995. One short well conserved region of *Alu*-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. Nucleic Acids Res 23:256–260.

Taylor CF, Charlton RS, Burn J, Sheridan E, Taylor GR. 2003. Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. Hum Mutat 22: 428–433.

Vasen HF, Watson P, Mecklin JP, Lynch HT. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 116:1453–1456.

Wagner A, Barrows A, Wijnen JT, van der Klift H, Franken PF, Verkuijlen P, Nakagawa H, Geugien M, Jaghmohan-Changur S, Breukel C, Meijers-Heijboer H, Morreau H, van Puijenbroek M, Burn J, Coronel S, Kinarski Y, Okimoto R, Watson P, Lynch JF, de la Chapelle A, Lynch HT, Fodde R. 2003. Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the MSH2 gene. Am J Hum Genet 72:1088–1100.

Wang Y, Friedl W, Lamberti C, Jungck M, Mathiak M, Pagenstecher C, Propping P, Mangold E. 2003. Hereditary nonpolyposis colorectal cancer: frequent occurrence of large genomic deletions in MSH2 and MLK1 genes. Int J Cancer 103:636–641.