ARTICLE

Cell type specificity in alternative splicing of the human mismatch repair gene *hMSH2*

Luka A Clarke, Peter Jordan and Maria Guida Boavida

Centro de Genética Humana, Instituto Nacional de Saúde Dr Ricardo Jorge, Lisbon, Portugal

Human non-polyposis colorectal cancer is caused by germline mutations in the DNA mismatch repair genes *hMSH2* and *hMHL1*. Several alternatively spliced mRNA species of these genes are present in peripheral blood lymphocytes of normal individuals, which can confound RT-PCR based techniques of mutation detection. Using RT-PCR, we compared the pattern of alternative splicing in whole peripheral blood lymphocytes (PBLs), separated T and B cells, lymphoblastoid cell lines (LCLs) from the same individuals, and a variety of tissues. Alternatively spliced forms of *hMLH1* lacking exons 9/10, 10/11 and 9/10/11 were found to have similar patterns of expression in T cells, B cells, and LCLs. By contrast, a subset of *hMSH2* transcripts, some of which were produced by utilisation of novel splicing motifs, were generally expressed in T but not in B cells. LCLs derived from the same blood samples showed no expression of any *hMSH2* splicing variants. The *hMSH2* Δ ex13 transcript, while absent from LCLs, was expressed in whole PBLs and both T and B cell fractions. This transcript was furthermore largely undetectable in tissues other than mononuclear blood cells. These data provide evidence for tissue specificity in the regulation of alternative splicing in *hMSH2*. In particular we show that LCLs generally do not express alternatively spliced forms of *hMSH2* mRNA and are thus suited for RT-PCR based mutation screening in that gene. *European Journal of Human Genetics* (2000) **8**, 347–352.

Introduction

Germline mutations in the mismatch repair (MMR) genes *hMSH2* and *hMLH1* are responsible for the majority of hereditary non-polyposis colorectal cancer[1] (HNPCC) cases.[2,3] A recent report[4] described a sensitivity for detecting the responsible mutations of 80% by systematic genomic sequencing, 64% by the *in vitro* synthesised protein truncation assay (IVSP[5]) and 93% when both techniques were combined. IVSP is therefore a powerful complementary tool for the detection of shortened proteins encoded by genes containing germline nonsense mutations, splice site mutations resulting in exon skipping, or gross genomic deletions, and as it can cover large portions of the coding region at once, it is more rapid than the laborious exon by exon approach required in genomic DNA sequencing. However,

naturally occurring alternative splicing transcripts of both the *hMSH2* and *hMHL1* genes can seriously confound mutation detection by this method, since they may yield IVSP products indistinguishable from those that would be caused by germline mutations.[4,6]

A number of studies have described alternatively spliced mRNA transcripts of these genes in normal individuals as well as some tissue-specific variations in their expression. Splicing variants of *hMLH1* involving exons 9, 10 and 11 are widespread, having been detected in whole and separated lymphocyte populations, lymphoblastoid cell lines (LCLs) and normal tissues,[7,8] although with a slight reduction in expression in LCLs.[6] In *hMSH2*, the transcript lacking exon 13 (Δex13) has been reported either to be restricted to mononuclear blood cells[8,9] or to be present at a low level in all tissues tested.[10] To date no firm conclusions have been drawn about cell type specificity in MMR gene alternative splicing.

Given that some variation exists, one cell or tissue type might prove more suitable for RNA-based methods of HNPCC mutation detection due to a lower overall level of MMR gene alternative splicing. In particular, RNA from LCLs

Correspondence: Centro de Genética Humana, Instituto Nacional de Saúde 'Dr Ricardo Jorge', Avenida Padre Cruz, 1649-016 Lisboa, Portugal. Tel: + 351 21 751 9380; Fax: + 351 21 759 0441; E-mail: cgh@insarj.pt

has been reported to contain reduced amounts of confounding natural transcript variants as compared with whole peripheral blood lymphocytes (PBLs).[6] Such a difference might be related to stimulation of the cell cycle in LCLs or alternatively reflect cell type-specific differences.

Here we provide evidence that a subset of *hMSH2* splicing variants is preferentially expressed in T rather than B lymphocytes, and in consequence, that LCLs are essentially free of *hMSH2* variants. In contrast, common *hMLH1* alternative transcripts are detectable in LCLs although with reduced expression levels in some cases. We studied expression of both genes by RT-PCR in normal individuals, comparing PBLs. fractions enriched in T and B lymphocytes, and LCLs.

Materials and methods

Peripheral blood was drawn from six healthy volunteers and PBLs isolated by centrifugation through lymphocyte separation medium (Gibco BRL, Paisley, UK). PBLs equivalent to 5 ml of blood were lysed for isolation of RNA, and a volume equivalent to 15 ml was passed through a human CD3$^+$ T cell enrichment column (R&D Systems, Minneapolis, USA; www.rndsystems.com) to obtain a T lymphocyte-enriched fraction. The fraction bound within the column was presumed to contain a high proportion of B cells, and was recovered by elution with trypsin at 37°C. LCLs were established from further blood samples taken from four of the six individuals by infection of PBL with the Epstein-Barr virus according to the standard protocol. The virus infects and immortalises B lymphocytes, and cell lines thus obtained were grown for 2–3 weeks prior to analysis. Further LCLs derived from the blood of 20 unrelated individuals were also obtained. Cellular mRNA was isolated from all four cell fractions (PBL, LCL, T cell fraction and B cell fraction) using the 'Quickprep' mRNA purification kit (Pharmacia Biotech, St. Albans, Herts, UK). Total RNA isolated from human tissues of foetal (colon mucosa, small intestine, liver, spleen, adrenals, pancreas, heart, brain, skin and placenta) or adult origin (colon mucosa, colon crypt epithelial cells, liver and macrophages) were a gift from Dr E Chastre, Paris. All cDNAs were prepared by reverse transcription with an oligo-dT primer and 50 ng of RNA, using the 'Ready To Go You-Prime' kit (Pharmacia Biotech, St. Albans, Herts, UK).

Amplification of *hMSH2* and *hMLH1* cDNAs was carried out using nested PCR with the sets of primers described below. Unless otherwise indicated, reaction conditions were initial denaturation at 95°C for 5 min followed by 35 cycles of 95°C for 30 s, 60°C for 1 min, 72°C for 2 min, and a final extension period of 5 min at 72°C. The first round PCR reactions contained 1.5 mM MgCl$_2$, 50–100 ng cDNA template, 300 ng each of forward and reverse oligonucleotide primers and 2.5 units of Amplitaq DNA polymerase (Perkin Elmer, Branchburg, NJ, USA). In the second round 1 μl of the first round PCR product served as template.

Primer set 1 was used to amplify the *hMSH2* cDNA (Accession number: HS04045) from nucleotide 25 in exon 1 to nucleotide 1165 in exon 7. First round: SA (5'-CAG CCG AAG GAG ACG GTG C-3') and SB (5'-CTT CTT GGC AAG TCG GTT AAG-3'). Second round: SA$_N$ (5'-GCT GCA GTT GGA GAG CGC GGC-3') and SB$_N$ (5'-GGT TAA GAT CTG GGA ATC GAC-3').

Primer set 2 was used to amplify the *hMSH2* cDNA from nucleotide 25 in exon 1 to nucleotide 2190 in exon 13. First round: SA and SD (5'-GAG GAT AGA AGC AGT TTC AAC-3'). Second round: SA$_N$ and SD$_N$ (5'-AA CAT TTC AGC CAT GAA CGT-3').

Primer set 3 was used to detect the *hMSH2* Δex13 variant and amplified from nucleotide 1846 in exon 12 to nucleotide 2292 in exon 14. First round, annealing temperature (Tm) 54°C: SC$_N$ (5'-CTC ATG GAT AAG AAC AGA ATA-3') and SF$_N$ (5'-CTA TGT CAA TTG CAA ACA GTC-3'). Second round was at Tm 58°C with primers MC11 and MC12.[10]

Primer set 4 was used to amplify the *hMLH1* cDNA (Accession number: HS07343) from codon 184 in exon 7 to codon 395 in exon 12. First round, Tm 58°C: LA (5'-ATG TCG TTC GTG GCA GGG GT-3') and LD$_N$ (5'-CCA AAG AGA AAT AGT CTG CAA-3'). Second round primers were MA2F and MAR1.[7]

PCR products were directly sequenced using the T7 Sequenase version 2.0 kit (Amersham Life Science, Little Chalfont, Bucks, UK) and ^{35}S-labelled dATP. Sequences were then run on 6% polyacrylamide gels and visualised by x-ray film audioradiography.

Results

Expression of *hMSH2* was analysed with primer sets 1 (exons 1–7), 2 (exons 1–13) and 3 (exons 12–14), and full length products were amplified in all PBL samples tested. In addition, a number of minor products were amplified and subsequently identified by direct sequencing. Some corresponded to previously prescribed splicing variants, including the common Δex5 variant,[8] the Δex2–6 product reported by this laboratory,[11] and the common Δex13 variant.[9,10] Furthermore, several previously undescribed forms were detected that used cryptic splice sites, some of which consisted of short (5–9 bp) sequence motifs that were identical at both the novel splice donor and acceptor sites. The individual *hMSH2* splicing variants identified are depicted in Figure 1.

Primer set 4 amplified five previously described *hMLH1* products between exons 7 and 12 of 663, 550, 456, 415 and 302 bp.[7,8] Their identities were confirmed by direct sequencing as *hMLH1* full length, *hMLH1* Δex9, Δex9–10, Δex10–11 and Δex9–11, respectively. In addition, one individual showed a strong 442 bp product resulting from cryptic splicing around a 5 bp repeated sequence motif. The individual *hMLH1* splicing variants identified are depicted in Figure 2.
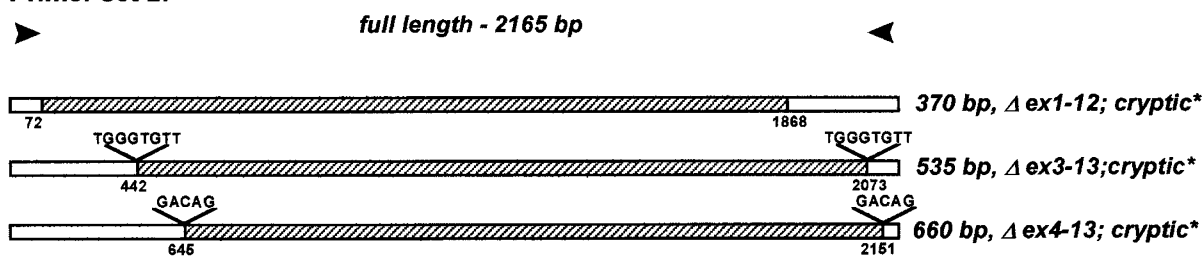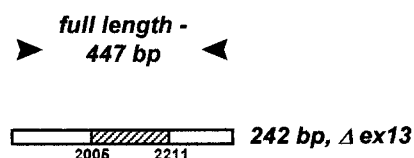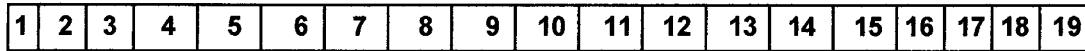
hMSH2
coding exons
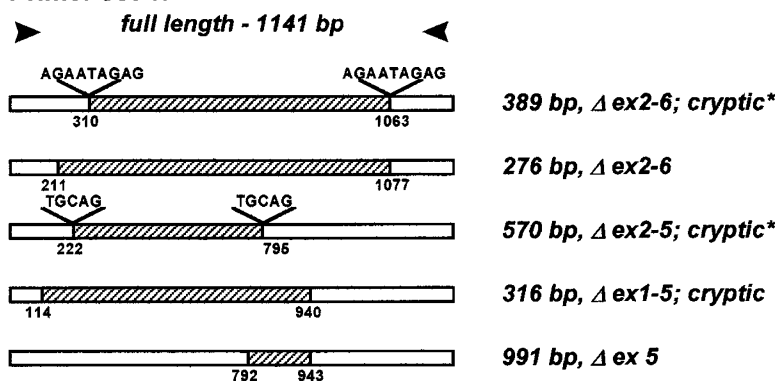
Figure 1 Schematic illustration of alternatively spliced *hMSH2* transcripts analysed in this study. Bold arrows indicate the positions of the three primer pairs used in the second round of nested RT-PCR with respect to the coding exons. The expected sizes of the full length amplification products are shown between them. The shaded region on each PCR product corresponds to the portion spliced out; numbers indicate nucleotides joined by alternative splicing events. Nucleotide positions are taken from the *hMSH2* sequence bearing accession number HS04045. The size of each spliced PCR product, and identification of exons wholly or partly spliced out is indicated on the right of each splicing variant. Where one or more non-constitutive splice sites have been used, the word *cryptic* appears. Sequences of short repeated sequence motifs present at both cryptic splice donor and acceptor sites are given in their respective positions. Asterisks denote previously undescribed splicing variants.

By contrast to the heterogeneity of products amplified from whole PBLs, only full length *hMSH2* products were amplified with primer sets 1 and 2 from LCLs of the same individuals (see Figure 3). This result was supported by the screening of a further 20 LCLs from unrelated individuals, which were also found to be essentially free of *hMSH2* splicing transcripts. Separation of PBLs into T and B cell fractions revealed that a subset of alternative splicing variants of *hMSH2* were T cell specific, suggesting that absence of these variant forms in LCLs resulted from the lack of

alternative splicing in B cells. The Δex13 splicing variant amplified by primer set 3, however, although absent from all 24 LCLs, was found to be expressed in both the T cell and B cell fractions of all subjects (see Figure 3). This indicated that expression of *hMSH2* Δex13 might be specific to a further cell type present in the B cell fraction. Following screening of a panel of human foetal and adult tissue samples, besides being found in whole PBLs, the Δex13 transcript was detected in macrophages and very faintly in foetal skin, but in no other tissue samples (see Figure 4), a result broadly consistent with

hMLH1
coding exons

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |

Primer set 4:

full length - 663 bp

550 bp; *Δ ex9*
667 791

456 bp; *Δ ex9-10*
667 885

415 bp; *Δ ex10-11*
790 1039

302 bp; *Δ ex9-11*
667 1039

ACCTC   ACCTC
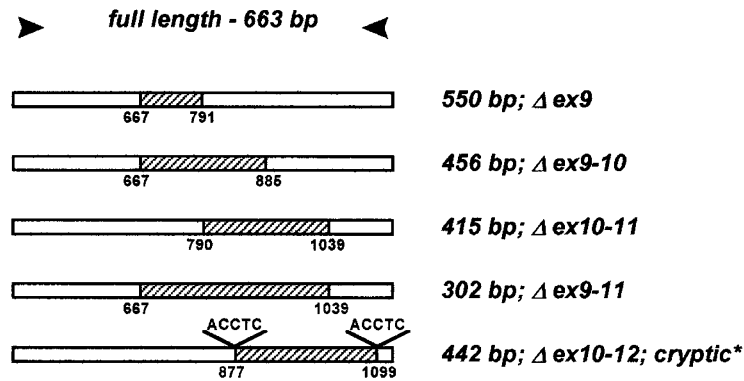442 bp; *Δ ex10-12; cryptic**
877 1099

Figure 2   Schematic illustration of alternatively spliced *hMLH1* transcripts analysed in this study. Bold arrows indicate the position of the primer pair used in the second round of nested RT-PCR with respect to the coding exons. The expected size of the full length amplification product is shown between them. The shaded region on each PCR product corresponds to the portion spliced out; numbers indicate nucleotides joined by alternative splicing events. Nucleotide positions are taken from the *hMLH1* sequence bearing accession number HS07343. The size of each spliced PCR product, and identification of exons wholly or partially spliced out is indicated on the right of each splicing variant. Where non-constitutive splice sites have been used, the word *cryptic* appears. Sequences of short repeated sequence motifs present at both cryptic splice donor and acceptor sites are given in their respective positions. The asterisk denotes a previously undescribed splicing variant.
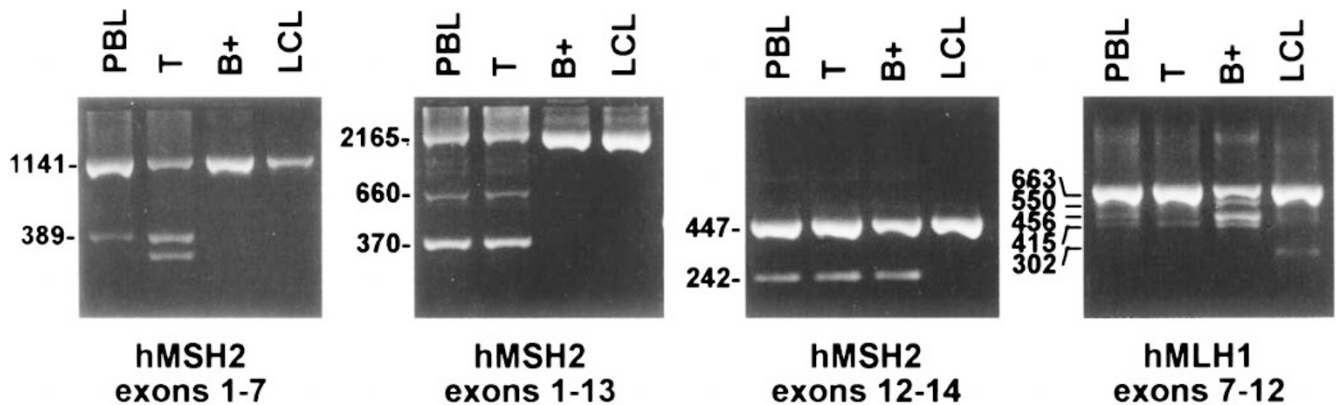

Figure 3   Selected results of nested RT-PCR of the *hMSH2* and *hMLH1* cDNAs using four primer sets. In each case, the cDNA templates were prepared from peripheral blood lymphocytes (PBL), a T cell-enriched fraction (T), the remainder B cell-rich fraction (B$^+$) and the corresponding lymphoblastoid cell line (LCL) from the same individual. Full length products of 1141 bp, 2165 bp, 447 bp and 663 bp were amplified by the four primer sets respectively, as labelled. A 389 bp product (formed by cryptic splicing between exons 2 and 6) was amplified by primer set 1 (*hMSH2* exons 1–7) in whole PBL and T cells but not in B cells or LCLs. Similarly, minor products of 660 bp and 370 bp (formed by cryptic splicing between exons 4–13 and 1–12 respectively) were amplified by primer set 2 (*hMSH2* exons 1–13) in whole and T lymphocytes but not in B cells or LCLs. Primer set 3 (*hMSH2* exons 12–14) amplified a 242 bp product lacking exon 13 in whole PBL, T and B cells but not in LCLs. Minor bands of 550, 456, 415 and 302 bp formed by splicing of *hMLH1* exons 9, 9–10, 10–11 and 9–11 respectively, were amplified by primer set 4 (*hMLH1* exons 7–12) in all lymphocyte fractions and LCLs.
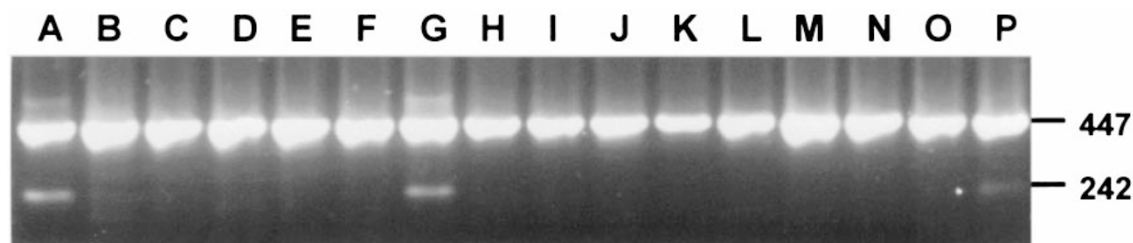
Figure 4 Nested RT-PCR amplification of *hMSH2* cDNA with primer set 3 from a panel of human adult and foetal tissues. A: PBLs, B: Mucosa 1, C: Mucosa 2, D: Colon*, E: Small Intestine*, F: Crypt, G: Macrophage, H: Liver, I: Liver*, J: Spleen*, K: Adrenal*, L: Pancreas*, M: Placenta*, N: Heart*, O: Brain*, P: Skin* (* tissue of foetal origin). The full length product of 447 bp was detected in all tissues tested, but the 242 bp band corresponding to the Δex13 transcript was only present in whole PBLs, macrophages and foetal skin.

data from two previous studies in which the Δex13 transcript was confined to mononuclear blood cells.[8,9] Macrophage contamination may therefore have accounted for the anomalous presence of this transcript in the B cell fraction. When the same bank of tissue samples was screened with primer set 1, other splicing variants, including *hMSH2* Δex5 and Δex2–6, were found to lack such a specific pattern, and were not detected in macrophages (data not shown). In this study, the Δex5 transcript was often expressed at levels no higher than non-specific background bands, which precluded a thorough confirmation of a cell-type specific pattern of occurrence. Expression of this transcript in LCLs can therefore not be excluded.

The *hMLH1* products amplified by primer set 4 provided less evidence of tissue or cell type specificity, being present in all human tissue samples (data not shown) and in all lymphocyte cell fractions from the individuals tested. In LCLs, however, some variation in the expression level of the spliced forms was observed (see Figure 3).

Discussion

Alternative splicing of pre-messenger RNA is a common phenomenon mediated by the interaction of RNA-binding proteins with specific splice site recognition elements.[12,13] The extent of alternative splicing of a given transcript is thought to reflect the tissue-specific expression profile of such RNA-binding proteins. In this report we have demonstrated preferential expression in T lymphocytes of a subset of alternatively spliced *hMSH2* transcripts that were absent from the B lymphocyte-rich fraction. Interestingly, many of these transcripts were produced by splicing around short sequence motifs present at both the 5' and 3' ends of the spliced region. Since one copy of each of these repeated motifs remains in the spliced end product, the exact site of splicing is unknown, but only one of these variants fully conforms to either the GT–AG or the AT–AC rules of the major or minor spliceosome consensus sequences[14] (see Table 1). However, recent evidence indicates much splice site diversity, including TA–TT intron splicing of an *hMSH2* variant and a constitutive AT–AA intron in *hMSH3*.[15]

Consistent with the lack of alternative splicing in B cells, there was a striking absence of the majority of *hMSH2* splicing variants from LCLs, which are established by immortalisation of B-lymphocytes. By contrast, differences in *hMLH1* expression were less well defined, with several common splicing variants being widespread. However, in agreement with a previous study,[6] LCLs revealed reduced expression of *hMLH1* variants involving exons 9–11. The physiological significance of the tissue-specific variations described above remains obscure, as does the general significance of alternative splicing in MMR genes. The lack of alternative splicing of *hMSH2* in a mitotically active cell type such as LCLs would, however, be consistent with the hypothesis that variant transcripts represent a redundant or down-regulatory form of mRNA. Such a role of aberrant transcripts might be expected in mature or resting cell populations such as T lymphocytes and macrophages with less mismatch repair activity. Possible effects on cancer predisposition resulting from translation of variant MMR proteins from individual splicing forms can also not be excluded.

Table 1 Comparison of cryptic splice site repeats detected in this study with known consensus sequences used by the major and minor spliceosomes[14]

'GT–AG' intron:	AG / GTRAGT...intron...CAG / G
'AT–AC' intron:	/ ATATCCTT...intron...YCCAC /
S389:	AGAATAGAG / CTGGAA...intron....AGAATAGAG / G
S570:	TGCAG / AGTGTT...intron...TGCAG / T
S535:	TGG / GTGTTA...intron...TGG / GTGTT
S660:	GACAG / GTACAT...intron...GACAG / T
L442:	AC / CTCAGTTT...intron...CTGAC / CTC

Each splicing variant is named with a letter (S: *hMSH2*; L: *hMLH1*) and a number (length in bp of the PCR product as shown in Figures 1 and 2). Repeated sequence motifs are underlined and suggested exon–intron boundaries shown as diagonal slashes. The only one of these splicing variants to exactly follow either the *GT–AG* or the *AT–AC* rule is S660, in which the 5′ splice junction is the natural exon 4/intron 4 boundary (S389 = *CT–AG*; S570 = *(A)GT–AG*; S535 = *GT–GG*; S660 = *GT–AG*; L442 = *CT–AC*).

In summary, the near absence of alternatively spliced *hMSH2* transcripts and the reduced expression of some *hMLH1* variants in LCLs provide an explanation for the observation shared by several laboratories that the use of LCLs in RNA-based mutation detection and IVSP resulted in less false positive products as compared with the use of PBLs. This leads to the conclusion that using LCLs is the preferred choice for RNA-based techniques when screening *hMSH2* mutations, but that naturally occurring alternative *hMLH1* transcripts might still confound mutation detection in that gene due to their ubiquitous expression.

References

1 Lynch HT, Smyrk TC, Watson P *et al*: Genetics, natural history, tumor spectrum, and pathology of hereditary non-polyposis colorectal cancer: an updated review. *Gastroenterology* 1993; **104**: 1535–1549.

2 Nyström-Lahti M, Parsons R, Sistonen P *et al*: Mismatch repair genes on chromosomes 2p and 3p account for a major share of hereditary nonpolyposis colorectal cancer families evaluable by linkage. *Am J Hum Genet* 1994; **55**: 659–665.

3 Liu B, Nicolaides NC, Markowitz S *et al*: Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat Genet* 1995; **9**: 48–55.

4 Farrington SM, Lin-Goerke J, Ling J *et al*: Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. *Am J Hum Genet* 1998; **63**: 749–759.

5 Luce MC, Marra G, Chauhan DP *et al*: In vitro transcription/ translation assay for the screening of hMLH1 and hMSH2 mutations in familial colon cancer. *Gastroenterology* 1995; **109**: 1368–1374.

6 Kohonen-Korish M, Ross VL, Doe WF *et al*: RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. *Am J Hum Genet* 1996; **59**: 818–824.

7 Charbonnier F, Martin C, Scotte M, Sibert L, Moreau V, Frebourg T: Alternative splicing of MLH1 messenger RNA in human normal cells. *Cancer Res* 1995; **55**: 1839–1841.

8 Genuardi M, Viel A, Bonora D *et al*: Characterization of MLH1 and MSH2 alternative splicing and its relevance to molecular testing of colorectal cancer susceptibility. *Hum Genet* 1998; **102**: 15–20.

9 Xia L, Shen W, Ritacca F *et al*: A truncated hMSH2 transcript occurs as a common variant in the population: implications for genetic diagnosis. *Cancer Res* 1996; **56**: 2289–2292.

10 Mori Y, Shiwaku H, Fukushige S *et al*: Alternative splicing of hMSH2 in normal human tissues. *Hum Genet* 1997; **99**: 590–595.

11 Marshall B, Isidro G, Boavida MG: Naturally occurring splicing variants of the hMSH2 gene containing nonsense codons identify possible mRNA instability motifs within the gene coding region. *Biochim Biophys Acta* 1996; **1308**: 88–92.

12 Chabot B: Directing alternative splicing: cast and scenarios. *Trends Genet* 1996; **12**: 472–478.

13 Cooper TA, Mattox W: The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet* 1997; **61**: 259–266.

14 Kreivi JP, Lamond AI: RNA splicing: unexpected spliceosome diversity. *Curr Biol* 1996; **6**: 802–805.

15 Wu Q, Krainer AR: AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol Cell Biol* 1999; **19**: 3225–3236.