LETTER TO JMG

Disruption of an exon splicing enhancer in exon 3 of *MLH1* is the cause of HNPCC in a Quebec family

S McVety, L Li, P H Gordon, G Chong, W D Foulkes

J Med Genet 2006;43:153–156. doi: 10.1136/jmg.2005.031997

Background: A 3 bp deletion located at the 5' end of exon 3 of *MLH1*, resulting in deletion of exon 3 from RNA, was recently identified.

Hypothesis: That this mutation disrupts an exon splicing enhancer (ESE) because it occurs in a purine-rich sequence previously identified as an ESE in other genes, and ESEs are often found in exons with splice signals that deviate from the consensus signals, as does the 3' splice signal in exon 3 of *MLH1*.

Design: The 3 bp deletion and several other mutations were created by polymerase chain reaction mutagenesis and tested using an in vitro splicing assay. Both mutant and wild type exon 3 sequences were cloned into an exon trapping vector and transiently expressed in Cos-1 cells.

Results: Analysis of the RNA indicates that the 3 bp deletion c.213_215delAGA (gi:28559089, NM_000249.2), a silent mutation c.216T→C, a missense mutation c.214G→C, and a nonsense mutation c.214G→T all cause varying degrees of exon skipping, suggesting the presence of an ESE at the 5' end of exon 3. These mutations are situated in a GAAGAT sequence 3 bp downstream from the start of exon 3.

Conclusions: The results of the splicing assay suggest that inclusion of exon 3 in the mRNA is ESE dependent. The exon 3 ESE is not recognised by all available motif scoring matrices, highlighting the importance of RNA analysis in the detection of ESE disrupting mutations.

Hereditary non-polyposis colorectal cancer (HNPCC) is an inherited predisposition to colorectal cancer (CRC) caused by a germline mutation in a mismatch repair gene. It has a dominant pattern of inheritance and high penetrance, and it accounts for approximately 3% of all CRC cases. Most mutations associated with HNPCC are detected in *MLH1* (50%), *MSH2* (40%), and *MSH6* (5%).¹ The mutation detection rate in HNPCC is often lower for various reasons. One is the limitation of common diagnostic techniques. For instance, large deletions, insertions, and rearrangements that are known to occur frequently in *MLH1* and *MSH2* are not easily detected by traditional methods of mutation analysis, such as exon-by-exon sequencing, single strand conformation polymorphism, and denaturing high performance liquid chromatography. Screening programmes that take a multi-modal approach—combining conventional mutation detection methods with Southern blotting, multiplex ligation dependent probe amplification, or the protein truncation test (PTT)—are more effective.² Another significant factor influencing the rate of mutation detection is the difficulty in establishing the pathogenicity of small mutations such as small in-frame deletions/insertions and single base substitutions.

Recently, it has been shown that small sequence changes may affect the activity of splicing regulatory sequences resulting in altered mRNA splicing. In particular, changes

in the coding sequence, which may or may not affect the encoded protein sequence, may disrupt exon splicing enhancers (ESEs), leading to exon skipping. ESEs are short, degenerate, frequently purine-rich sequences that are important in both constitutive and alternative splicing.^{3–7} ESEs have been identified in a large number of genes, and their disruption has been linked to several genetic disorders, including HNPCC,^{8–10} cystic fibrosis,^{11 12} Marfan's syndrome,¹³ and Becker muscular dystrophy.¹⁴ A mutation generating a new splicing regulatory element has also been reported.¹²

Most ESEs are thought to be binding sites for SR proteins,^{3–5} a family of splicing factors with RNA recognition motifs and serine/arginine-rich regions called RS domains.⁵ The RNA binding domain binds to consensus sequences in the exon, and the RS domain promote and stabilise the interactions of splice site recognition complexes, enhancing exon inclusion and regulating splice site selection.^{15–18}

We have identified a purine-rich region in exon 3 of *MLH1* which may represent a novel functional ESE. A 3 base pair in-frame deletion, c.213_215delAGA (gi:28559089, NM_000249.2), in this region of exon 3 was detected by exon by exon sequencing in a Quebec family meeting Amsterdam criteria I for HNPCC. PTT analysis revealed a truncated protein product in the patient, and cDNA sequencing indicated that the del c.213_215delAGA mutation causes exon 3 to be skipped during mRNA splicing. Analysis of normal and tumour tissue from one patient showed loss of heterozygosity in the tumour, confirming that the in-frame deletion of 3 base pairs (bp) in exon 3 is the cause of HNPCC in this family. Using site directed mutagenesis we have mutated bases in and near the ESE in exon 3, and we have tested the effects of these mutations on splicing, using an in vitro splicing assay. The results of the splicing assay strongly suggest the existence of an ESE in exon 3 of *MLH1*.

METHODS

The patient underwent genetic testing following informed consent. This study is the result of our attempt to identify the causative mutation.

Splicing construct

Silent, missense, and nonsense mutations in the ESE and flanking sequence, and the patient's mutation, c.213_215delAGA, were all created by site directed polymerase chain reaction (PCR) mutagenesis.

Wild type or mutant exon 3 was cloned into the EcoRI and BamHI sites of the multiple cloning site of the exon trapping plasmid pSPL3 (Invitrogen, Burlington, Ontario, Canada), shown in fig 1. An EcoRI site present 65 bp upstream of exon 3 in intron 2 and a BamHI site introduced by PCR 235 bp downstream of exon 3 were used. Plasmids were grown in subcloning efficiency DH5α chemically competent *E coli*.

Abbreviations: CRC, colorectal cancer; ESE, exon splicing enhancer; HNPCC, hereditary non-polyposis colorectal cancer; NAS, nonsense associated altered splicing; PTT, protein truncation test

Figure 1 A plasmid containing a splice donor site and a splice acceptor site was used to assay exon 3 inclusion. (A) The region surrounding the multiple cloning site of the exon trapping vector pSPL3. An enlarged view of the insert is shown below. Mutant and wild type exon 3 sequences were cloned into the EcoRI and BamHI sites of the pSPL3 plasmid. Arrows above the vector represent the primers used to analyse the splicing products. (B) The sequence surrounding the putative exon splicing enhancer (ESE) is shown for all the inserts. ESE bases are in bold. Mutated bases are underlined. bp, base pair.

(Invitrogen). Ampicillin (Invitrogen) was used as the selection agent, at 100 µg/ml.

Transfection

Cos-1 cells, grown to approximately 85% confluency in supplemented Dulbecco's modified Eagle medium (Invitrogen) in six-well plates, were transfected with either 0 DNA (control) or 4 µg of plasmid DNA per well using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol.

Total RNA was extracted after 24 hours using Trizol Reagent (Invitrogen) and resuspended in 20 µl.

Analysis of splicing products

For each clone, reverse transcription PCR was undertaken on 0.5 µl of RNA using primer PR and SuperScriptII reverse transcriptase (Invitrogen) according to the manufacturer's protocol.

PCR was carried out on 4 µl of cDNA using primers SF and SR and Taq polymerase (Invitrogen) according to the manufacturers' protocol. The following thermocycling conditions were used: three minutes at 95°C; five cycles of 30 seconds at 95°C, 30 seconds at 63°C, and 30 seconds at 72°C; and 30 cycles of 30 seconds at 95°C, 30 seconds at 58°C, and 30 seconds at 72°C.

PCR products were analysed by electrophoresis on a 6% native polyacrylamide gel.

The positions of the primers used are shown in fig 1.

RESULTS

Mutations inserted into the putative ESE disrupt normal splicing

The outcome of the splicing assay was determined by analysis of the PCR products. The PCR yields a product of 276 bp when exon 3 is included in the transcript, and a product of 177 bp when exon 3 is skipped, as shown in fig 2. A single

Figure 2 Several point mutations inserted into the putative exon splicing enhancer in exon 3 of *MLH1* disrupt normal splicing. Products of the splicing analysis were run on a 6% native polyacrylamide gel. The numbers accompanying the deletion names represent the clone number.

Figure 3 The 3' splice site of exon 3 of *MLH1* differs slightly from the consensus splice signal. The splice sites used in normal splicing are highlighted in black boxes. A cryptic 5' splice site is highlighted in a grey box. The Quebec deletion, c.213_215delAGA, and the Finnish deletion, c.210_213delAGAA, have a similar effect on the resulting sequence. Both deletions disrupt a putative exon splicing enhancer (ESE), highlighted in bold characters. Sequence homologous to the ESE is underlined. The GAAGAU ESE motif is present in the Finnish sequence, but it overlaps the splice site.

product of 276 bp was observed for wild type exon 3 and for mutation c.210A→C, predicted to change a lysine (K) residue to an asparagine (N) in the protein, indicating normal splicing. A single product of 177 bp was observed for the patient's mutation, c.213_215delAGA, and the silent mutation, c.216T→C, indicating complete exclusion of exon 3 in the transcript. For the missense mutation c.214G→C, expected to result in the change of a glutamate (E) to a glutamine (Q) in the protein, and the nonsense mutation c.214G→T, expected to change a glutamate to a stop codon (X) at the protein level, two bands were present. Both the 276 bp and the 177 bp products were produced, indicating partial exon skipping. The 276 bp band was much fainter than the 177 bp band, and was apparently absent in clone c.214G→T-3 in the photograph, indicating that activity of the ESE was greatly reduced by these two mutations.

DISCUSSION

The deletion of bases 213–215 AGA in *MLH1* disrupts an exon splicing enhancer

A 3 bp in-frame deletion, c.213_215delAGA, results in the loss of codon 71 and causes skipping of exon 3 during mRNA splicing. Another 3 bp deletion resulting in the deletion of codon 71, previously described as c.211_213delGAA, has been reported¹⁹ and is shown in fig 3. In that case, the effect of the loss of codon 71 on protein function was studied by introducing the deletion into cDNA by PCR mutagenesis, cloning the cDNA, and then expressing the protein in *Spodoptera frugiperda* 9 cells. The mutant protein was found to have a reduced association with PMS2, and complexes of mutant *MLH1* and normal PMS2 were unable to complement MMR deficient cells. However, because the mutation was inserted directly into the cDNA, the effect of c.211_213delGAA on splicing was not seen. Our results show that deletion of codon 71 causes the in-frame skipping of exon 3 in the mRNA, resulting in the loss of 33 amino acids. Thus, while the glutamic acid encoded by codon 71 may be important to the function of *MLH1*, the pathogenicity of c.211_213delGAA, like that of c.213_215delAGA, is probably caused by the disruption of an ESE and not by the simple loss of one codon. In fact, as the resulting sequence is the same for both deletions, both should actually be named

c.213_215delAGA according to the guidelines for mutation nomenclature outlined by the Human Genome Variation Society (<http://www.genomic.unimelb.edu.au/mdi/>), which states that variations should be assigned arbitrarily to the 3'-most position possible.

Small mutations can result in significant changes in the way RNA is processed when they occur in elements that regulate splicing, such as ESEs. Here, missense, nonsense, and silent mutations introduced into the same region as c.213_215delAGA all reduced the amount of normal transcript. The existence of a nuclear reading frame recognition system that induces nonsense associated altered splicing (NAS) has been proposed to explain exon skipping in exons harbouring premature stop codons.^{4, 20} However, both a nonsense and a missense mutation introduced at the same location have the same effect on splicing, suggesting that disruption of an ESE, and not NAS, causes exon skipping in the case of the c.214G→T mutation.

A similar mutation in the same region, c.210_213delAGAA, has been described in a patient in Finland and is also associated with exon skipping.⁹ Owing to the proximity of the Finnish deletion to the 3' splice site of exon 3, exon skipping in that case was attributed to disruption of the splice site.⁹ As shown in fig 3, the resulting sequence from both deletions is very similar. The 3' splice site of exon 3 differs slightly from the consensus sequence, and there are two 5' splice sites in exon 3, which supports our hypothesis that normal splicing of exon 3 may require the presence of regulatory sequences. This is also illustrated in fig 3.

Potential clinical implications of altered splicing

The discovery that the missense mutation c.214G→C in exon 3 reduces, but does not abolish, normal splicing suggests that different ESE alterations could be associated with different phenotypes. Mutations that completely abrogate normal splicing may be associated with more typical HNPCC, while low level production of *MLH1* transcripts encoding a mutated but functional exon 3 could result in reduced penetrance. Assays of *MLH1* function have been developed,^{21–23} and could be used to determine if the mutant protein harbouring a glutamate to glutamine substitution resulting from the c.214G→C mutation is functional.

Conclusions

The importance of studying RNA

Consensus binding motifs for several SR proteins have been established, and web based tools are available to identify potential ESEs in RNA sequences. ESEfinder (<http://rulai.cshl.edu/tools/ESE/>) identifies potential binding motifs for four of the 10 known human SR proteins.²⁴ However, not all sequences shown to have ESE activity are known consensus SR protein binding motifs. ESEfinder does not recognize the ESE identified here in exon 3 of *MLH1*. Another tool, called RESCUE-ESE (<http://genes.mit.edu/burgelab/rescue-ese/>), has been developed to identify RNA sequences with known ESE activity.²⁵ RESCUE-ESE recognises the ESE in exon 3 of *MLH1* identified here, but also two other sequences that may or may not represent true ESEs. Both tools have been shown to have a high rate of false positive results.^{26–27} Therefore, at present, the best way to determine the effect of a sequence variant on splicing is to study the cDNA.

Disruption of ESE sequences in *MLH1* can cause HNPCC

In this study, we have demonstrated the presence of an ESE in exon 3 of *MLH1* and shown that a three base pair in-frame deletion in this ESE is the cause of HNPCC in a Quebec family. Disruption of an ESE in exon 12 of *MLH1* has also been associated with HNPCC.⁸ Furthermore, four different pathogenic mutations affecting codon 659 of *MLH1* have been reported to date, and all four result in skipping of exon 17,²⁸ suggesting the presence of an ESE in exon 17. Taken together, these data suggest that disruption of ESEs in *MLH1* represents a distinct category of mutation contributing to HNPCC.

ACKNOWLEDGEMENTS

We would like to thank the Cancer Research Society and the Judy Steinberg Trust for their financial support of this project. We also thank Dr Bernard Lemieux and Dr Gail Ouellette of the Centre hospitalier universitaire de Sherbrooke for referring the patient to us.

Authors' affiliations

S McVety, Department of Human Genetics, McGill University, Montreal, Quebec, Canada

L Li, Programme in Cancer Genetics, Department of Oncology and Human Genetics, McGill University

P H Gordon, Department of Surgery, Division of Colorectal Surgery, Sir Mortimer B Davis Jewish General Hospital, McGill University

G Chong, Department of Diagnostic Medicine, SMBD Jewish General Hospital, McGill University

W D Foulkes, Cancer Prevention Centre, SMBD-Jewish General Hospital, McGill University

Conflicts of interest: none declared.

Correspondence to: Dr William Foulkes, Department of Medical Genetics, Room C-107.1, Sir M B Davis-Jewish General Hospital, 3755 Côte Ste Catherine Rd, Montreal, Quebec, Canada H3T 1E2; william.foulkes@mcgill.ca

Received 13 February 2005

Revised version received 17 May 2005

Accepted for publication 20 May 2005

REFERENCES

- 1 de la Chapelle A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer* 2004;**4**:769–80.
- 2 Wagner A, Barrows A, Wijnen JT, van der KH, Franken PF, Verkuijlen P, Nakagawa H, Geugien M, Jaghmohan-Changur S, Breukel C, Meijers-Heijboer H, Morreau H, Van Puijtenbroek M, Burn J, Coronel S, Kinarski Y, Okimoto R, Watson P, Lynch JF, de la Chapelle A, Lynch HT, Fodde R. 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* 2003;**72**:1088–100.
- 3 Blencowe BJ. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 2000;**25**:106–10.
- 4 Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002;**3**:285–98.
- 5 Manley JL, Tacke R. SR proteins and splicing control. *Genes Dev* 1996;**10**:1569–79.
- 6 Schaal TD, Maniatis T. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol* 1999;**19**:261–73.
- 7 Roesser JR, Liittschwager K, Leff SE. Regulation of tissue-specific splicing of the calcitonin/calcitonin gene-related peptide gene by RNA-binding proteins. *J Biol Chem* 1993;**268**:8366–75.
- 8 Stella A, Wagner A, Shito K, Lipkin SM, Watson P, Guanti G, Lynch HT, Fodde R, Liu B. A nonsense mutation in *MLH1* causes exon skipping in three unrelated HNPCC families. *Cancer Res* 2001;**61**:7020–4.
- 9 Renkonen E, Lohi H, Jarvinen HJ, Mecklin JP, Peltomäki P. Novel splicing associations of hereditary colon cancer related DNA mismatch repair gene mutations. *J Med Genet* 2004;**41**:e95.
- 10 Gorlov IP, Gorlova OY, Frazier ML, Amos CI. Missense mutations in hMLH1 and hMSH2 are associated with exonic splicing enhancers. *Am J Hum Genet* 2003;**73**:1157–61.
- 11 Pagani F, Stuaní C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 2003;**12**:1111–20.
- 12 Pagani F, Buratti E, Stuaní C, Baralle FE. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J Biol Chem* 2003;**278**:26580–8.
- 13 Caputi M, Kendzior RJ, Beemon KL. A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. *Genes Dev* 2002;**16**:1754–9.
- 14 Shiga N, Takeshima Y, Sakamoto H, Inoue K, Yokota Y, Yokoyama M, Matsuo M. Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J Clin Invest* 1997;**100**:2204–10.
- 15 Eperon IC, Makarova OV, Mayeda A, Munroe SH, Caceres JF, Hayward DG, Krainer AR. Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* 2000;**20**:8303–18.
- 16 Lavigne A, La Branch, Kornblihtt AR, Chabot B. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev* 1993;**7**:2405–17.
- 17 Zahler AM, Neugebauer KM, Lane WS, Roth MB. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* 1993;**260**:219–22.
- 18 Zahler AM, Roth MB. Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proc Natl Acad Sci USA* 1995;**92**:2642–6.
- 19 Rævaara TE, Timoharju T, Lonnqvist KE, Kariola R, Steinhoff M, Hofstra RM, Mangold E, Vos YJ, Nystrom-Lahti M. Description and functional analysis of a novel in frame mutation linked to hereditary non-polyposis colorectal cancer. *J Med Genet* 2002;**39**:747–50.
- 20 Wang J, Hamilton JL, Carter MS, Li S, Wilkinson MF. Alternatively spliced TCR mRNA induced by disruption of reading frame. *Science* 2002;**297**:108–10.
- 21 Maquat LE. NASTY effects on fibrillin pre-mRNA splicing: another case of ESE does it, but proposals for translation-dependent splice site choice live on. *Genes Dev* 2002;**16**:1743–53.
- 22 Quaresima B, Alifano P, Tassone P, Avvedimento EV, Costanzo FS, Venuta S. Human mismatch-repair protein MutL homologue 1 (*MLH1*) interacts with *Escherichia coli* MutL and MutS in vivo and in vitro: a simple genetic system to assay *MLH1* function. *Biochem J* 2003;**371**:183–9.
- 23 Yuan ZQ, Gottlieb B, Beitel LK, Wong N, Gordon PH, Wang Q, Puisieux A, Foulkes WD, Trifiro M. Polymorphisms and HNPCC: PMS2-*MLH1* protein interactions diminished by single nucleotide polymorphisms. *Hum Mutat* 2002;**19**:108–13.
- 24 Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003;**31**:3568–71.
- 25 Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science* 2002;**297**:1007–13.
- 26 Lastella P, Resta N, Miccolis I, Quagliarella A, Guanti G, Stella A. Site directed mutagenesis of hMLH1 exonic splicing enhancers does not correlate with splicing disruption. *J Med Genet* 2004;**41**:e72.
- 27 Sharp A, Pichert G, Lucassen A, Eccles D. RNA analysis reveals splicing mutations and loss of expression defects in *MLH1* and *BRCA1*. *Hum Mutat* 2004;**24**:272.
- 28 Nystrom-Lahti M, Holmberg M, Fidalgo P, Salovaara R, de la Chapelle A, Jiricny J, Peltomäki P. Missense and nonsense mutations in codon 659 of *MLH1* cause aberrant splicing of messenger RNA in HNPCC kindreds. *Genes Chromosomes Cancer* 1999;**26**:372–5.