Chapter 4

Functional analysis of MSH2 unclassified variants found in suspected Lynch syndrome patients reveals pathogenicity due to attenuated mismatch repair

Eva Wielders, Jan Hettinger, Rob Dekker, C. Marleen Kets, Marjolijn Ligtenberg, Arjen Mensenkamp, Ans van den Ouweland, Judith Prins, Anja Wagner, Winand Dinjens, Hendrikus Jan Dubbink, Liselotte van Hest, Fred Menko, Frans Hogervorst, Senno Verhoef and Hein te Riele

*J Med Genet, 51, 245-253, 2014*
Introduction

Lynch syndrome (LS) is an autosomal dominant disorder caused by inherited defects in the mismatch repair (MMR) system. Individuals with LS have an up to ten times elevated risk of colorectal (CRC) and endometrial cancer compared to the general population. They furthermore have an increased risk of developing tumours of the small bowel, stomach, ovaries, pancreas, ureter and renal pelvis, bladder, biliary tract, brain and sebaceous glands.

The main function of MMR is recognition and repair of replication errors. MSH2 protein can form a heterodimeric complex with MSH6 (hMutSα) or MSH3 (hMutSβ), which can initiate the MMR process by binding to erroneously copied DNA. hMutSα binds single base-base mismatches or single unpaired bases whereas hMutSβ preferentially binds loops of multiple unpaired bases.

Mismatch-bound hMutS dimer recruits hMutLα, a dimer of MLH1 and PMS2, which directs further repair.

Most inherited MMR mutations in LS are found in MLH1 and MSH2 (70-80% of cases); mutations in MSH6 and PMS2 account for the remaining 20-30%. When LS is suspected, tumour
material is analysed for microsatellite instability (MSI) and/or expression of MMR proteins by immunohistochemistry (IHC). If a lack of expression is found, the corresponding gene(s) are analysed for mutations. When IHC shows no aberrancies or cannot be done, but the family history is strongly indicative for LS, mutation analysis of all four MMR genes is warranted. The majority of mutations found in LS families are abrogating mutations, such as frameshifts, truncations and deletions, which are unambiguously disease causing. However, also missense mutations are found, whose potential for pathogenicity is much more difficult to predict. These so-called Unclassified Variants (UVs) pose a serious problem to the clinic. As long as it is unclear whether a mutation affects the MMR system, the clinical suspicion of LS cannot be confirmed and no presymptomatic testing is available to identify at-risk family members.

The assess the pathogenicity of UVs, integrative approaches have been proposed to calculate likelihood ratios based on clinical data (segregation analysis, IHC, MSI, loss of heterozygosity) and in silico analysis considering evolutionary conservation and physicochemical differences between amino acids \cite{8-11}. In addition, functional assays have been developed based on in vitro reconstituted MMR reactions or cellular complementation of MMR deficiency by overexpression of variant MMR proteins, often in distantly related species such as yeast and Escherichia coli. While such approaches can accurately identify fully deleterious mutations, partial MMR defects may remain unnoticed \cite{12}.

We have therefore developed an alternative method to study the functional effects of missense mutations in MSH2 and MSH6 by recreating these variants at the endogenous genes in mouse embryonic stem cells (ESC) using oligonucleotide-directed gene modification \cite{13}. This allows us to study whether and to which extent the variant protein supports the most relevant MMR functions in a highly homologous system \cite{14} (murine MSH2 and MSH6 proteins share 94 and 96% homology with their human counterparts, respectively).

A defect in the MMR system is manifested by length alterations of mononucleotide- or dinucleotide repeat sequences (MSI). Additional functions of the MMR system include the prevention of recombination between homologous but not identical DNA sequences \cite{15,16} and the cellular response to methylating agents such as 6-thioguanine (6-TG) and N-methyl-N’-nitro-N-nitrosoguanidine (MNNG) that add methyl groups to bases in the DNA. MMR prohibits the incorporation of nucleotides opposite methylated guanines during DNA replication, which ultimately leads to double strand breaks \cite{17-19}. As a result, MMR-proficient cells are sensitive to methylating agents, while MMR-deficient cells are tolerant \cite{15,20}. 
Addressing these functions in ESCs expressing the variant allele provides an accurate and quantitative readout for MMR activity. Here we applied this method to study the functional consequences of three MSH2 missense mutations that have been found in Dutch suspected LS families and found evidence for partial MMR defects.

Results

Selection of UVs

We analysed three unclassified MSH2 variants detected in suspected LS families that are currently being counselled by clinical geneticists in the Netherlands. The main clinical findings are summarised in Table 1.

The index patient of the first family (Figure 1A), referred to the Department of Human Genetics of the Radboud University Medical Center, had colorectal cancer at the age of 51. The tumour showed MSI at mononucleotide but not at dinucleotide markers (Table 1) and normal IHC staining for MLH1 and PMS2 (not shown) and MSH2 and MSH6 (Figure 2A, a). Lymphocyte DNA sequencing of MSH2, MLH1, MSH6 and PMS2 detected a variant nucleotide in MSH2: c.493T>G resulting in a tyrosine to aspartic acid substitution at position 165 (p.Tyr165Asp), further indicated as Y165D. A brother, also carrier of the UV, developed colon cancer at the age of 42. His tumour showed MSI at both types of repeats and absence of MSH2 and MSH6 (Figure 2A, b). The different MSI and IHC results in the two families raise doubt as to whether the Y165D mutation was responsible for tumour development. In silico analysis gave varying predictions, sorting intolerant from tolerant (SIFT), polymorphism phenotyping (Polyphen) and Align-GVGD predicting a polymorphism, the high Grantham score and multi-variate analysis of protein polymorphism-mismatch repair (MAPP-MMR) suggesting pathogenicity (Table 2). Based on the dissimilarities of tumours in mutation carriers and the ambiguity of the prediction programmes, the UV was classified as a variant of uncertain significance.

The second variant is the MSH2 c.2068C>G mutation, resulting in amino acid substitution p.Gln690Glu (Q690E). This UV was found in lymphocyte DNA of a female who developed colon cancer at the age of 41 and was referred to the Department of Clinical Genetics of the Erasmus Medical Center in Rotterdam (Figure 1B and Table 1). The tumour showed MSI and MSH2 and MSH6 were absent (Figure 2B). Next to the MSH2 missense mutation, two single base substitutions, both known poly-
morphisms, were detected in MSH6: IVS5+14A>T (rs2020911)\textsuperscript{21,22} and a c.268G>A mutation resulting in a glycine to glutamic acid substitution at position 39 (rs1042821)\textsuperscript{23,24}. The patient’s brother, who was diagnosed with CRC at the age of 50, also carried the UV. Also his tumour showed MSI and absent MSH2 and MSH6 staining. However, in this family, also a non-carrier developed CRC at young age and some carriers developed a tumour only at high age of which one was microsatellite-stable (MSS) and showed MSH2 and MSH6 staining (Figure 1B and Table 1). Of the four daughters of mutation carrier II-4, one died of heart disease (not shown in the pedigree because their mutation status is unknown). \textit{In silico} analyses gave varying predictions (Table 2) providing no unambiguous indication for the pathogenicity of this variant.

<table>
<thead>
<tr>
<th>Patient</th>
<th>sex</th>
<th>UV</th>
<th>location of cancer</th>
<th>Age at Diagnosis (years)</th>
<th>MSI mono</th>
<th>MSI di</th>
<th>Presence of MLH1/ PMS2</th>
<th>Presence of MSH2/6</th>
</tr>
</thead>
<tbody>
<tr>
<td>YDa III-3</td>
<td>male</td>
<td>+</td>
<td>colon transversum</td>
<td>51</td>
<td>3/3</td>
<td>0/3</td>
<td>present</td>
<td>present</td>
</tr>
<tr>
<td>YDa III-4</td>
<td>male</td>
<td>+</td>
<td>colon transversum</td>
<td>42</td>
<td>3/3</td>
<td>3/3</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>YDb</td>
<td>male</td>
<td>+</td>
<td>colon descendens</td>
<td>38</td>
<td>3/3</td>
<td>3/3</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>QE III-1</td>
<td>female</td>
<td>+</td>
<td>colon</td>
<td>41</td>
<td>5/5</td>
<td>ND</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>QE III-2</td>
<td>male</td>
<td>+</td>
<td>rectum</td>
<td>50</td>
<td>5/5</td>
<td>ND</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>QE III-4</td>
<td>male</td>
<td>-</td>
<td>colon</td>
<td>48</td>
<td>0/5</td>
<td>ND</td>
<td>present</td>
<td>present</td>
</tr>
<tr>
<td>QE II-4</td>
<td>male</td>
<td>+</td>
<td>coecum</td>
<td>79</td>
<td>5/5</td>
<td>ND</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>QE II-6</td>
<td>male</td>
<td>+</td>
<td>sigmoid</td>
<td>80</td>
<td>0/5</td>
<td>ND</td>
<td>present</td>
<td>present</td>
</tr>
<tr>
<td>MV III-2</td>
<td>female</td>
<td>+</td>
<td>colon descendens</td>
<td>48</td>
<td>0/2</td>
<td>0/5</td>
<td>present</td>
<td>present</td>
</tr>
</tbody>
</table>

LS Lynch syndrome; MSI, microsatellite instability; UV, unclassified variants

The index patients are indicated in bold.

| Table 2. \textit{in silico} analysis of MSH2 codon substitutions* |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| UV             | SIFT \textsuperscript{25} | PolyPhen \textsuperscript{26} | Align GVGD \textsuperscript{27} | Grantham \textsuperscript{28} | MAPP-MMR \textsuperscript{8} |
| MSH2-Y165D     | 0.22                        | 0.003                       | C0                          | 160                         | 8.75                         |
| MSH2-Q690E     | 0.07                        | 0.998                       | C25                         | 29                          | 10.23                        |
| MSH2-MB13V     | 0.05                        | 0.24                        | C0                          | 21                          | 7.79                         |
| Neutral\textsuperscript{#} | ≥0.05                       | <0.5                        | C0                          | <60                         | <3                           |
| Deleterious\textsuperscript{#} | <0.05                      | >2                          | C65                         | >100                        | >5                           |

* using web-based applications \textsuperscript{29}.
# reference values for neutral and deleterious variants.

MAPP-MMR, multivariate analysis of protein polymorphisms-mismatch repair; PolyPhen, polymorphism phenotyping; SIFT, sorting intolerant from tolerant; UV, unclassified variant.
Figure 1. Pedigrees of the suspected Lynch syndrome families. Pedigrees of families carrying the MSH2-Y165D variant (A) the MSH2-Q690E variant (B) or the MSH2-M813V variant (C). For privacy reasons, the pedigrees have been adjusted. The index patients are indicated with an arrow. Family members are indicated by Roman (generation) and Arabic numbers (individual). + and − symbols refer to presence or absence of the UV when known. Black symbols indicate that cancer has been confirmed. Striped boxes refer to anamnestic cases. Besides each specified family member, the type of cancer and the age of cancer diagnosis is indicated: Br, breast cancer; C, colon cancer; L, lung cancer; Li, liver cancer; Pr, prostate cancer; R, rectal cancer and Uro, urothelial cancer.

The index patient of the third family was diagnosed at the Netherlands Cancer Institute with carcinoma of the sigmoid at 48 years of age. A hyperplastic polyp was removed 12 years later (Figure 1C and Table 1). The carcinoma was MSS and all four MMR proteins were present (Figure 2C). Also
her brother developed a carcinoma of the sigmoid at age 49 but no additional clinical data is available. The mother of the patient died at age 79 without cancer. Of her six siblings, four got cancer in their sixties and seventies (Figure 1C); according to family history their mother had died of cancer at young age. Mutation analysis in the index patient revealed an A>G change in MSH2 at position 2437, resulting in the amino acid substitution p.MSH2-M813V (M813V) 30. Most in silico analyses suggested this UV to be neutral but the MAPP-MMR score was indicative for pathogenicity (Table 2). Taken together, although the tumour of the index patient showed no signs of MMR deficiency, pathogenicity of the UV could not be excluded.

Functional analysis of UVs
The selected UVs were introduced into the genome of mouse ESCs using oligo targeting 13 (Supplementary Figure 1) and the presence of the mutations was confirmed by sequencing (Figure 3). Next, we inactivated the wild-type allele using conventional Msh2 targeting. Correct targeting of the wild-type allele was verified by Southern blotting and mutation specific (quantitative) PCR. For further details see Materials and Methods and Supplementary Figures 2-4.

We first determined the MSH2, MSH3 and MSH6 protein levels using western blotting to see whether the mutations affected protein stability and dimer formation (Figure 3). The MSH2 levels in Msh2YD/- and Msh2Q690E/- cells were markedly reduced compared with Msh2+/- and Msh2MV/- cells, which showed similar levels of MSH2 protein. The levels of MSH3 and MSH6 provide information about dimer formation since the stability of these proteins depends on the presence of MSH2 31. Although the MSH3 and MSH6 levels were both reduced in the Msh2YD/- and Msh2Q6/- cell lines, these proteins were still present in all mutant cell lines, whereas they were virtually absent in Msh2/- cells. This indicates that all three MSH2 variants were able to bind and stabilise MSH3 and MSH6.

To study the MMR capacity of the three variant MSH2 proteins, we first determined the mutator phenotype using two assays addressing MSI and mutagenesis at the Hprt gene, respectively. For both assays, three single cells per cell line were expanded to 10^9 cells to allow for approximately 30 population doublings during which mutations could accumulate. The frequency of MSI was then determined by measuring the length of two mononucleotide and two dinucleotide repeats in approximately 30 single cell clones derived from each expanded culture. In Msh2/- cells, the levels of mononucleotide and dinucleotide repeat instability were similar, whereas in Msh6/- cells mononucleotide repeats were more unstable than dinucleotide repeats (Figure 4A). The MSH2-Q690E protein appeared to be fully inactive in the repair of mononucleotide slippage errors since the slippage frequency was as high as in MSH2-deficient and MSH6-deficient cells.
Figure 2. Immunohistochemical analysis of human tumours. Immunohistochemical staining of tumours from the index patients. (A) MSH2-Y165D variant: MSH2 and MSH6 staining in the colorectal tumour from the patient from family 1 (indicated with (a)); In the patient from family 2 (indicated with (b)), MSH2 and MSH6 stainings are absent from the nuclei in the tumour tissue. (B) MSH2-Q690E variant: MLH1 and PMS2 are present in the tumour and MSH2 and MSH6 are absent. (C) MSH2-M813V variant: MSH2, MSH6, MLH1 and PMS2 are all present in the tumour.
Figure 3. Generation of mutant Msh2 ESC lines. Sequence analysis of (A) Msh2+ and Msh2YD genomic DNA, (B) Msh2+ and Msh2QE genomic DNA and (C) Msh2+ and Msh2MV genomic DNA. Single letter amino acid codes are given below the sequence. (D) Whole cell lysates were analysed for MSH2, MSH6 and MSH3 by Western blotting. γ-Tubulin was used as a loading control. +, wild-type allele; -, knockout allele; YD, QE, MV, mutant alleles.

The level of dinucleotide repeat instability, however, was only about half of the level found in MSH2-deficient cells indicating some residual activity of MSH2-Q690E. The MSH2-Y165D protein had retained partial repair activity towards mononucleotide slippage errors and almost full activity for dinucleotide slippage events. The level of mononucleotide repeat instability in cells expressing the MSH2-M813V variant was not different from that in wild-type cells and neither was the amount of instability of the dinucleotide repeats. To measure other mutation events, cells from the expanded cultures were exposed to 6-TG at a dose that is lethal to ESCs unless they have obtained an inactivating mutation in the Hprt gene. The MSH2 mutant cell lines generated a lower number of 6-TG-resistant colonies than the MSH2-deficient line, although the mutation frequency varied between the three variant cell lines (Figure 4A). The number of 6-TG-resistant colonies was zero (Msh2MV/- cell line) or comparable to that of Msh2+/- cells (Msh2QE/- cell line), suggesting full functionality of these two variants in this assay. Only the Msh2YD/- cells showed a partial defect in this repair assay with a mutation frequency about half of that of MSH2-deficient cells.

Finally, we studied the ability of the variant MSH2 proteins to mediate the toxicity of methylating DNA damage. Repeated attempts of MMR to repair erroneous incorporation of a thymidine opposite a methylated guanine (MeG) causes formation of lethal double strand breaks. As a result, MMR proficient cells are sensitive to methylating agents such as MNNG and 6-TG, while MMR deficient cells are
Figure 4. Functional analysis of Msh2 mutant ESC lines (A) White and black bars show the average percentage of length alterations at mono- and dinucleotide repeats, respectively (left Y-axis). The grey bars show the average number of 6-TG resistant colonies per 10^6 plated cells as measured in the Hprt assay (right Y-axis). (B) Survival of mutant and control cell lines exposed to MNNG. (C) Survival of mutant and control cell lines exposed to 6-TG. Error bars show standard errors from independent experiments. +, wild-type allele; -, knockout allele; YD, QE, MV, mutant alleles.
tolerant (Figure 4B and C). Upon exposure to increasing MNNG concentrations in a colony survival assay we saw that the Msh2<sup>YD/−</sup> and Msh2<sup>QE/−</sup> cell lines were equally resistant as Msh2<sup>+/−</sup> cells, which indicates a defect in the recognition or processing of MeG-T mismatches. The Msh2<sup>MV/−</sup> cells did not show such a defect since they were as sensitive to MNNG exposure as the Msh2<sup>+/+</sup> control cell line (Figure 4B). When cells were exposed to 6-TG we again observed that Msh2<sup>MV/−</sup> cells were as sensitive as MSH2-proficient cells (Figure 4C). In contrast to MNNG however, the Msh2<sup>YD/−</sup> and Msh2<sup>QE/−</sup> cell lines were partially sensitive to 6-TG, suggesting that these mutant MSH2 proteins had retained some functionality in this assay.

**Discussion**

In order to support genetic counselling, we have applied our method of MMR gene UV analysis to three MSH2 missense mutations found in suspected LS families that are currently being seen in our clinical genetics centres. Each of these variants is rare: the Y165D variant was detected in two families among 182 families with a proven MMR gene defect seen at the Radboud University Medical Center; the Q690E and M813V variants were detected in only one family and none of the variants was present in the 1000 Genomes database. Segregation analysis and clinical parameters (MSI, IHC) did not allow deciding on the likelihood of pathogenicity of the three variants. Therefore all family members were advised biannual colonoscopy based on the international guidelines for the clinical diagnosis of suspected LS.

In human and mouse MSH2, 92% of the amino acids are identical and the positions of the three variant amino acids studied here are conserved in human, mouse, chicken and fish, but not in fruit fly and yeast. The results of our analyses are summarised in Table 3.

<table>
<thead>
<tr>
<th>MSH genotype</th>
<th>UVs</th>
<th>MSH2 levels</th>
<th>Mutation rate</th>
<th>Damage response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSI-mono 0.3%</td>
<td>MSH2 sensitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSI-di 1.1%</td>
<td>MSH2 sensitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hprt (x10&lt;sup&gt;-6&lt;/sup&gt;) 7.1</td>
<td>MSH2 tolerant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6TG 14.3%</td>
<td>MSH2 tolerant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MNNG 7.5%</td>
<td>MSH2 tolerant</td>
</tr>
</tbody>
</table>

Table 3. MMR activity of variant MSH2 protein

The bold text is indicative of a MMR defect.

MMR, mismatch repair, MNNG, N-Methyl-N’-nitro-N-nitrosoguanidine; MSI, microsatellite instability; UV, unclassified variant.
Upon introduction of the MSH2-Y165D UV in the endogenous gene of mouse ESCs, we noticed that the MSH2 protein level was markedly reduced compared with \( Msh2^{+/+} \) cells. MSH3 and MSH6 were also reduced in this cell line but not absent, indicating that the proteins could still interact and residual levels of MutS\( \alpha \) and MutS\( \beta \) were formed. This is consistent with the location of residue Y165 in the so-called connector domain that is believed to mediate allosteric signalling, but not dimerisation.\(^{32}\) This domain is 92% identical between human and mouse MSH2, 14/15 substitutions remaining within the same side-chain group. Reduced MutS\( \alpha \) and MutS\( \beta \) levels are in accordance with the partial MMR defect in \( Msh2^{YD/-} \) ESCs. Interestingly, the results from our MSI analysis matched those from the tumour of the index patient in family a: instability of mononucleotide but not of dinucleotide markers. In \( Msh2^{YD/-} \) ESCs the frequencies of MSI at mononucleotide repeats and mutations in \( Hprt \) were clearly elevated with respect to wild-type cells but twofold to threefold lower than in fully \( Msh2^{-/-} \) deficient cells. This partial MMR defect was also observed in the clonogenic survival assay where we saw increased, but not complete, tolerance to 6-TG. Our experimental results did not easily explain the different MSI and IHC results in tumours from different mutation carriers. The reduced level of MSH2-Y165D protein we observed may allow immunohistochemical detection only under optimal conditions that may not have been reached in all cases. Alternatively, an additional (epi)genetic event may have further reduced MSH2-Y165D protein level in the two tumours in which MSH2 could not be detected; this may also explain MSI at dinucleotide repeats.

Based on our clinical and experimental data we conclude that the p.MSH2-Tyr165Asp substitution is most likely disease causing, although it may be a reduced penetrance allele. This conclusion contrasts with the PolyPhen, SIFT and Align GVGD predictions, which classified the mutation as benign. However, these prediction models included the protein sequence from the outbred Sprague-Dawley rat strain, which appears to be the only sequenced mammal that has an aspartic acid at position 165. Recently, the Y165D mutation was found to be completely inactive in an \textit{in vitro} MMR assay,\(^{12}\) although our \textit{in vivo} assays addressing several MMR functions are indicative for some residual activity.

The experimental results obtained for the p.MSH2-Q690E mutation also showed a partial defect, although more severe than the Y165D mutation. This may relate to the location of Q690 in the highly conserved ATPase domain (97% identity between human and mouse). MSH2-Q690E protein levels were lower than MSH2-Y165D levels, but again MSH3 and MSH6 were visible, indicating proper dimerisation. Mononucleotide instability in \( Msh2^{QE/-} \) ESCs was as high as in \( Msh2^{-/-} \) cells and dinucleotide instability was
about half of the level found in Msh2/- cells. Surprisingly, the mutation rate in the Hprt assay was very low. Similar to the Msh2YD/- ESCs, Msh2QE/- ESCs were resistant to MNNG and partially resistant to 6-TG. Our results correspond with the clinical data. The absence of MSH2 and MSH6 in tumour material is consistent with the very low protein levels in our ESCs. Furthermore, three out of four tumours from the UV carriers showed instability of all five mononucleotide markers tested. When we combine our experimental results with the clinical data, the p.MSH2-Q690E mutation is most likely the pathogenic mutation in this family, although it may be less penetrant than a full knockout allele. Indeed, two carriers of the UV did not develop an MSI tumour: whereas one carrier (II-2) did not develop cancer at all, the adenocarcinoma that developed in family member II-6 was MSS and showed staining for all four MMR proteins and was therefore likely sporadic. However, it is also possible that development of MSI tumours in these individuals was hampered by specific modifier genes. Also the two MSH6 single base substitutions that were found in the lymphocyte DNA from the index patient, IVS5+14A>T and c.268G>A resulting in a glycine to glutamic acid substitution at position 39 could play a role. The first is a known polymorphism for which no correlation with CRC risk has been found. Instead, homozygosity for the G39E substitution has been associated with a modestly increased risk for colorectal cancer, and a small increase in colon cancer incidence has been reported among male carriers of this mutation (odds ratio 1.27). It is therefore possible that the MSH6-G39E variant affects tumour predisposition in carriers of the MSH2-Q690E mutation. However, to exert such an effect, the two mutations must reside on the same chromosome since MSH2 and MSH6 are closely located on chromosome 2: if the mutations were on different chromosomes, an LOH event eliminating the wild-type MSH2 allele would also eliminate the mutant MSH6 allele. Therefore, variable inheritance of the MSH6-G39E mutation unlikely explains the variable tumour incidence and characteristics in carriers of the MSH2-Q690E mutation.

Although family history suggested a strong predisposition for colorectal cancer, the p.MSH2-M813V substitution is unlikely to be the cause. Consistent with microsatellite stability and the presence of all four MMR proteins in the tumour from the index patient, we saw no reduction in protein level and no functional MMR defects in our mouse Msh2MV/- ESCs. Of note, the mutation underlying the human M813V variant was analysed in SpliceSite-Finder-like, MaxEntScan, NNSPLICE, GeneSplicer and Human Splice Finder, but none of them indicated an effect on splicing. We therefore classify this variant as non-pathogenic.

In summary, we have recreated three MSH2 UVs found in Dutch families and analysed their effect on different MMR functions. By targeting the highly
homologous mouse MMR genes and subsequent inactivation of the wild-type allele in ESCs, we closely mimicked the human situation. Combined with the clinical data, our experimental results provide strong indications for the pathogenicity of two of these UVs. We have previously shown that in all our assays a proven deleterious mutant, MSH2-P622L, behaved indistinguishable from a full MSH2 knockout allele \(^{14}\). Therefore, the intermediate phenotypes of MSH2-Y165D and MSH2-Q690E in our assays are real manifestations of partial MMR defects. Attenuated MMR activity may remain unnoticed in in vitro MMR assays using purified proteins or cellular overexpression systems. Although it is unlikely that the mutations we have examined affect gene regulation or splicing, it is unfortunate that we have not been able to formally exclude this possibility by analysis of cDNA prepared from tumour or normal tissue.

We believe our approach is a valuable alternative to existing functional assays and will facilitate UV classification and subsequent counseling. However, we realize translating laboratory data to clinical practice is challenging. Recently, classification systems have been proposed that predominantly rely on clinical information. As this information is often insufficient or inconsistent, Plon et al. have proposed five classes of variants based on the likelihood of pathogenicity \(^{37}\). The authors proposed that proper interpretation of all aspects used to classify variant alleles requires a panel of experts covering the range of expertise required. While this classification system will certainly provide guidance to clinical geneticists, we believe that state of the art patient care dictates full exploitation of scientific knowledge. The MMR system has been extensively studied for decades in bacteria, yeast and higher eukaryotes. This has generated a wealth of knowledge about the function and functioning of MMR genes and yielded ample protocols for experimentally testing the MMR capacity of individual MMR gene variants. The inclusion of functional assays in any classification system for MMR gene variants is therefore highly desirable, a viewpoint also advocated by Rasmussen et al. \(^{38,39}\). With this report we intend to contribute to this discussion and to lower the barrier to clinical implementation of functional tests. Extending the panel of experts proposed by Plon et al.\(^{37}\) by a molecular biologist ensures full advantage to be taken from the outcome of functional assays.

**Acknowledgements**

This work was supported by grants from the Dutch Cancer Society (NKI 2004-3084 and NKI 2009-4477).
Materials and Methods

Molecular analyses of human tumours

For MSI analysis, genomic DNA was isolated from formalin-fixed paraffin-embedded tissue and normal tissue, essentially as described by Hoogerbrugge et al. 40. Microsatellite markers BAT40, BAT26, D1S158, D2S123, BAT25, D5S346, D9S63, D17S250 and D18S58 were used in various combinations. According to international guidelines, a tumour was considered MSI-high when at least two out of five markers showed length alterations. IHC was performed on formalin-fixed, paraffin-embedded tumour and normal tissues, using antibodies against MLH1 (Pharmingen code: 51-1327gr and Novocastra/Leica clone ES05), PMS2 (Pharmingen code: 556415 and Epitomics clone EPR3947), MSH2 (Oncogene Research Products code: NA26 and Pharmingen clone G219-1129) and MSH6 (Transduction Laboratories code: G70220 and Pharmingen clone 44/MSH6).

Mutation analyses in humans

DNA extracted from peripheral blood lymphocytes was analysed for MSH2, MSH6, MLH1 and PMS2 mutations by a capillary sequence analysis and multiplex ligation-dependent probe amplification (MLPA, MRC-Holland kit P003-B1). IHC, MSI analyses and sequencing of MMR genes were performed in the context of a referral to genetic counselling, for diagnostic purposes and after consent had been obtained.

Generation of codon substitutions in Msh2

The oligonucleotide-directed gene modification (‘oligo targeting’) procedure was carried out in wild-type ESCs, essentially as described 13. Briefly, we transfected ESCs with a single-stranded DNA oligonucleotide (Supplementary Figure 1) after establishing transient down regulation of MLH1. Transfected cells were expanded and plated on 96-well plates at 5000 cells/well. Mutation-specific PCR was used to identify pools containing mutated cells as described 14. A positive pool was subcloned using limiting dilution in subsequent pools of 1000, 100 and 1 cell/well. Once a clonal cell line was established, cDNA was made using an oligo dT primer. We then amplified a cDNA fragment containing the modified codon by PCR. The resulting PCR product was cloned into the pGEM-T Easy vector (Promega) and sequenced to verify the presence of the mutation using vector primers T7 and SP6. Primer sequences are available upon request.

Inactivation of the wild-type allele

For inactivation of the wild-type allele in Msh2YD/+ and Msh2MV/+ ESCs, we used a targeting vector derived from Claij and Te Riele 41, in which exon 12 was removed by deleting a SnaB1-Swa1 fragment (Supplementary Figures 2 and 4). In Msh2QE/+ cells, the wild-type allele was inactivated using a targeting vector in which exon 12 was disrupted by a hygromycin resistance marker (Supplementary Figure 3) 15. Details of the procedure are given in the legends to Supplementary Figures 2, 3 and 4.

Western blot analysis

Cells were lysed in a buffer containing 150 mM NaCl, 50 mM Hepes pH 7.5, 5 mM EDTA, 0.1% NP-40, 5 mM NaF, 0.5 mM vanadate, 20 mM β-glycerolphosphate and 1 tablet complete protease inhibitor cocktail (Roche) per 50 ml. Protein concentrations were measured using the Pierce BCA protein assay kit (Thermo scientific) and equal amounts of
each sample were separated by 3-8% Tris-Acetate gels (NuPAGE) using the NuPAGE electrophoresis system and transferred to nitrocellulose membrane. We used rabbit polyclonal antibodies as primary antibodies to detect MSH2 (1:500) and MSH6 (1:500), and mouse monoclonal antibodies to detect MSH3 (1:50) and γ-Tubilin (GTU-88, Sigma-Aldrich). Peroxidase-conjugated goat anti-rabbit IgG and goat anti-mouse IgG (BioSource International) were used as a secondary antibody. Signals were visualised with enhanced chemiluminescence.

**Hprt mutation assay**

Single cells of the mutant cell lines $Msh2^{YD/-}$, $Msh2^{QE/-}$ and $Msh2^{MV/-}$, and the control cell lines $Msh2^{+/-}$, $Msh2^{+/+}$ and $Msh2^{-/-}$ were expanded to $10^9$ cells. Each expanded clone was plated onto three 150 mm gelatine-coated tissue culture plates at $1.5 \times 10^6$ cells/plate. The next day, 6-TG was added at a final concentration of 10 μg/ml. After 10 days, the resistant colonies were counted.

**MSI assay**

A minimum of three single-cell clones of each mutant and control cell line were expanded to $10^9$ cells. Of each of the expanded cultures we generated 32 subclones and isolated genomic DNA. The length of two dinucleotide microsatellite markers (D7Mit17 and D14Mit15) as well as two mononucleotide markers (mBAT-26 and mBAT37) was analysed by PCR analysis.

**Sensitivity to MNNG and 6-TG**

The mutant cell lines, $Msh2^{YD/-}$, $Msh2^{QE/-}$ and $Msh2^{MV/-}$ and the control cell lines, $Msh2^{+/-}$, $Msh2^{+/+}$ and $Msh2^{-/-}$ were plated onto irradiated mouse embryonic fibroblast feeder layers at a density of 500 cells/1.8 cm$^2$. The next day we treated the cells for 1 hour with 0-40 μM MNNG or 6-TG and after 4 days we counted the number of surviving colonies. In the case of MNNG exposure, cells were cultured in the presence of 40 μM $O^6$-benzylguanine, starting from 1 hour prior to MNNG exposure until counting of the colonies. $O^6$-benzylguanine inhibits the removal of methyl groups from the $O^6$ position of guanine by endogenous $O^6$-methylguanine-methyltransferase activity.
References


Supplementary Figure 1. Msh2 targeting oligonucleotides.

Msh2 targeting oligonucleotides (upper case) hybridized to their complementary genomic sequence (lower case); mismatching bases in the oligonucleotides are shown in red. The codon alteration is underlined.
Supplementary Figure 2. Generation of the Msh2<sup>YD/-</sup> ESC line. (A) Targeting vector used to inactivate the wild-type allele (adapted from 41). "neo" indicates the selectable neomycin resistance gene and exons are indicated as black boxes with exon numbers in white. The grey area between SnaB1 and Swa1 was deleted leading to almost complete removal of exon 12 and hence inactivation of Msh2 upon homologous recombination. Recombination between lox sites (black triangles) was therefore not necessary. (B) Southern blot analysis of Msh2<sup>+/+</sup> and Msh2<sup>YD/-</sup> cell lines showing the appearance of an EcoR1 fragment indicative for targeting of one of the alleles, using the probe shown in panel A. (C) qPCR analysis of Msh2 cDNA to show inactivation of the wild-type allele. cDNA was produced exclusively from the non-targeted allele using a primer laying in the grey area that is deleted in the inactivated allele (dotted arrow). The asterisk indicates the position of the YD mutation in exon 3 and the arrows show the position of two primer pairs specific for the Y (wild-type) and D (mutant) alleles (right pair is forward, F; left pair is reversed, R). The graph shows the relative wild-type (white bars) and MSH2-Y165D (black bars) mRNA levels in Msh2<sup>+/+</sup>, Msh2<sup>YD/+</sup> and Msh2<sup>YD/-</sup> cells as measured using wild-type and mutation specific F and R primer sets, respectively. The strongly reduced signal from the wild-type primer pairs in Msh2<sup>YD/-</sup> cells demonstrates targeting of the wild-type Msh2 allele.
**Supplementary Figure 3. Generation of the Msh2\(^{QE/-}\) ESC line.** (A) Targeting vector used to inactivate the wild type allele (described by de Wind et al.\(^{15}\)). "hyg" indicates the selectable hygromycin gene and exons are shown as black boxes with exon numbers in white. (B) Southern blot analysis of the Msh2\(^{+/+}\) and Msh2\(^{QE/-}\) cell lines showing the appearance of an EcoR1 fragment indicative for targeting of one of the alleles, using the probe shown in panel A. (C) Mutation specific PCR to show inactivation of the wild type allele. In the schematic, the hygromycin gene is shown in grey, the position of the mutation in exon 13 is indicated with an asterisk and arrows show the position of the wild-type (wt) or mutation-specific (m) primer pairs used in the PCR reaction (lower panel). The presence of the 2 kb hyg gene precludes effective amplification, indicating that in Msh2\(^{QE/-}\) cells the wild-type allele was targeted. (D) qPCR analysis of Msh2 cDNA to show inactivation of the wild type allele. In the schematic the hygromycin cassette is shown in grey, the asterisk indicates the QE mutation and the arrows indicate two primer pairs specific for the Q (wild-type) and E (mutant) alleles (right pair is forward, F; left pair is reversed, R). The graph shows the relative wild type (white bars) and MSH2-Q690E (black bars) mRNA levels of wild type, Msh2\(^{QE/+}\) and Msh2\(^{QE/-}\) cells as measured using wild-type and mutation specific F and R primer sets, respectively. Since insertion of the Hyg gene destabilizes Msh2 mRNA, reduction of the signal with the wt F and R primer sets in Msh2\(^{QE/-}\) cells is indicative for inactivation of the wild-type allele.
Supplementary Figure 4. Generation of the Msh2\textsuperscript{MV/−} ESC line. (A) Targeting vector used to inactivate the wild-type allele (adapted from 41). “neo” indicates the selectable neomycin resistance gene and exons are indicated as black boxes with exon numbers in white. The grey area between SnaB1 and Swa1 was deleted leading to almost complete removal of exon 12 and hence inactivation of Msh2 upon homologous recombination. Recombination between lox sites (black triangles) was therefore not necessary. (B) Southern blot analysis of Msh2\textsuperscript{+/+} and Msh2\textsuperscript{MV/−} cell lines showing the appearance of an EcoR1 fragment indicative for targeting of one of the alleles, using the probe shown in panel A. (C) Mutation specific PCR to show inactivation of the wild type allele. In the schematic, the deleted region is shown in grey, the position of the mutation in exon 14 is indicated with an asterisk and arrows show the position of the wild-type (wt) or mutation-specific (m) primer pairs used in the PCR reaction (lower panel). The left lane shows the DNA size marker. A lower band due to the exon 12 deletion was only seen with the wt primer set in Msh2\textsuperscript{MV/−} cells, indicative for targeting of the wild-type allele. (D) qPCR analysis of Msh2 cDNA to show inactivation of the wild type allele. The grey area in the schematic shows the region that is deleted in the inactivated allele. The arrows indicate the position of the primers used for the quantitative PCR and the asterisk that of the MV mutation. The graph shows the relative wild type (white bars) and MSH2-M813V (black bars) mRNA levels in wild type, Msh2\textsuperscript{MV/+} and Msh2\textsuperscript{MV/−} cells as measured using wild type and mutation specific primer sets. F and R indicate whether the Forward or Reversed primer set was used. Deletion of the SnaB1/Swa1 fragment includes about half of the following intron, which is expected to greatly hamper proper slicing. Therefore reduction of the signal with the wt F and R primer sets in Msh2\textsuperscript{MV/−} cells is indicative for inactivation of the wild-type allele.