Published in final edited form as:

Hum Mutat. 2009 May ; 30(5): 757–770. doi:10.1002/humu.20936.

Classifying *MLH1* and *MSH2* variants using bioinformatic prediction, splicing assays, segregation and tumor characteristics

Sven Arnold¹, Daniel D. Buchanan^{1,2}, Melissa Barker¹, Lesley Jaskowski¹, Michael D. Walsh^{1,2}, Genevieve Birney¹, Michael O. Woods³, John L. Hopper⁴, Mark A. Jenkins⁴, Melissa A. Brown⁵, Sean V. Tavtigian⁶, David E. Goldgar⁷, Joanne P. Young^{1,2}, and Amanda B. Spurdle^{1,2,*}

¹Genetics and Population Health Division, Queensland Institute of Medical Research, Brisbane, Australia

²School of Medicine, University of Queensland, Brisbane, Australia

³Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

⁴Centre for Genetic Epidemiology, University of Melbourne, Melbourne, Australia

⁵School of Medicine, and School of Molecular and Microbial Sciences, University of Queensland, Brisbane

⁶International Agency for Research on Cancer, Lyon, France

⁷Department of Dermatology, University of Utah, Salt Lake City, Utah, USA

Abstract

Reliable methods for predicting functional consequences of variants in disease genes would be beneficial in the clinical setting. This study was undertaken to predict, and confirm *in vitro*, splicing aberrations associated with mismatch repair (MMR) variants identified in familial colon cancer patients. Six programs were used to predict the effect of 13 *MLH1* and 6 *MSH2* gene variants on pre-mRNA splicing. mRNA from cycloheximide-treated lymphoblastoid cell lines of variant carriers was screened for splicing aberrations. Tumors of variant carriers were tested for microsatellite instability and MMR protein expression. Variant segregation in families was assessed using Bayes factor causality analysis. Amino acid alterations were examined for evolutionary conservation and physicochemical properties. Splicing aberrations were detected for ten variants, including a frameshift as a minor cDNA product, and altered ratio of known alternate splice products. Loss of splice sites was well predicted by splice site prediction programs SpliceSiteFinder (90%) and NNSPLICE (90%), but consequence of splice site loss was less accurately predicted. No aberrations correlated with ESE predictions for the nine exonic variants studied. Seven of eight missense variants had normal splicing (88%), but only one was a substitution considered neutral from evolutionary/physicochemical analysis. Combined with information from tumor and segregation analysis, and literature review, 16/19 variants were considered clinically relevant. Bioinformatic tools for prediction of splicing aberrations need improvement before use without supporting studies to assess variant pathogenicity. Classification

*Correspondence to: Amanda B Spurdle, PhD, Genetics and Population Health Division, Queensland Institute of Medical Research, 300 Herston Rd, Herston, Brisbane, 4029, Australia. Tel: +617 3362 0371; Fax: +617 3362 0105; Email: Amanda.Spurdle@qimr.edu.au..

of mismatch repair gene variants is assisted by a comprehensive approach which includes *in vitro*, tumor pathology, clinical, and evolutionary conservation data.

Keywords

MLH1; *MSH2*; MMR; splicing; bioinformatics; pathogenicity; tumor; cancer

Introduction

Lynch syndrome is an autosomal dominantly inherited disorder of cancer predisposition, caused by a functional defect in the DNA mismatch repair (MMR) complex. Identified mutations in the MMR genes *MLH1* (MIM# 120436) and *MSH2* (MIM# 609309) account for approximately 90% of Lynch syndrome families with detectable mutations, *MSH6* (MIM# 600678) for up to 10%, and *PMS2* (MIM# 600259) only very occasionally (Hampel, et al., 2005; Viel, et al., 1998; Wang, et al., 1999). Although Lynch syndrome families often exhibit a clinical phenotype of multiple cases of early onset colorectal and endometrial cancers across several generations (Jass, 1998; Lynch, et al., 1998), a more definitive diagnosis can be readily made on two important molecular pathological features of the tumour. Defects in MMR lead to tumour DNA microsatellite instability (MSI), detected in the laboratory as numerous insertion or deletion mutations in short repetitive sequence elements. In addition, loss of MMR protein expression is common, with no immunohistochemical (IHC) evidence of protein expression in >90% of cases with clearly pathogenic mutations. MSI or loss of MMR protein expression in the tumour of a young onset colorectal cancer case is indicative of Lynch syndrome (Caldes, et al., 2004; Lindor, et al., 2002). While IHC measurement of MMR protein appears to be the most sensitive and specific test for truncating and other immuno-unstable mutations (Southey, et al., 2005), this would obviously not be true for the subset of mutations encoding immuno-stable proteins with impaired function.

MMR gene truncation mutations considered to be pathogenic are identified in ~50% of colorectal cancer cases with presumptive Lynch Syndrome based on clinico-pathological features (Loughrey, et al., 2007). However MMR gene sequence variants of uncertain clinical significance (UVs) have been identified in a considerable proportion of such cases. UVs include rare nucleotide changes predicted to cause missense substitutions, small in-frame deletions, and sequence changes that may alter splicing due to their location near (*but not within*) consensus splice sites. Although individually rare, they collectively constitute a significant proportion of variants. At the time of study initiation, UVs had been identified in almost 10% of colorectal cancer case probands (n=37/383) ascertained through the Australasian site of the Colon Cancer Family Registry. Moreover, an analysis of the distribution of UVs by count and ascertainment revealed that UVs seen >10 times are detected in virtually equal numbers in clinic-based and population-based cases, but UVs detected once are *over-represented* at a ratio of 5:1 in “familial” clinic-based cases (Supp. Figure S1). Importantly, clinic-based participants in these series were screened for mutations irrespective of tumour IHC results. This suggests that rare UVs detected in individuals selected from family cancer clinics are more likely to be pathogenic. UVs present a challenge for genetic counselling, and thorough evaluation of their clinical significance will have direct impact on families concerned.

Estimates of pathogenicity are particularly difficult for UVs in MMR genes. They are individually rare, so classical family studies such as segregation analysis provide little power to provide precise estimates of disease risk on a per-family basis. Moreover, although multifactorial likelihood models incorporating genetic, tumour pathology, and evolutionary

data have been developed for the clinical evaluation of UVs in other high risk genes (Chenevix-Trench, et al., 2006; Goldgar, et al., 2004; Spurdle, et al., 2008b), a similar application for MMR genes is complicated by the fact that many clinical laboratories precede gene screening with IHC testing to prioritize the relevant MMR gene for full sequencing. That is, there is *a priori* an increased likelihood that *causal genetic variation* will be detected in the gene selected for screening, but there is also a greater chance that a neutral variant will occur *in cis* with an undetected mutation, and *appear* to be causal from family inheritance studies. While recombinant vector-based functional studies may provide a more direct test of compromised function of individual variants, these are time-consuming and require specialised laboratory skills. Therefore, use of non-laboratory-based bioinformatic criteria for prediction of the functional consequences of UVs would be beneficial in the clinical setting, where resources for functional studies are usually limited.

Splicing aberrations underlie an increasingly recognized group of mutations in high-risk genes, including MMR genes. Mutations in the two highly conserved intronic dinucleotides 5'GT and 3'AG flanking exons are well recognized to result in exclusion of at least the adjacent exon from the transcript, and in the clinical setting are generally classified as pathogenic on basis of sequence information alone. However, mutations in intronic nucleotides in close proximity to splice sites, and in cryptic sites associated with splicing can also produce a non-functional protein (Davoodi-Semiromi, et al., 2000). Furthermore, there is evidence that variants predicted to cause missense substitutions can alter exonic splice enhancer (ESE) sites at the nucleotide level, to cause aberrant splicing (Aretz, et al., 2004; Cravo, et al., 2002; McVety, et al., 2005; Sharp, et al., 2004; Zatkova, et al., 2004). While these differ from vector-based studies in that germline material can be tested for specific splicing aberrations predicted to occur, extensive testing using even these relatively simple assays generally falls outside the resources of clinical laboratories. We tested a panel of UVs *in vitro* for possible splicing aberrations, and compared the accuracy of a variety of bioinformatic approaches in predicting the observed splicing aberrations associated with each UV.

Materials and Methods

Written informed consent was obtained from all family and control participants. Recruitment studies and this laboratory-based study have been approved the Human Research Ethics Committee of participating Institutions.

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. cDNA references were NM_000249.2 for *MLH1*, and NM_000251.1 for *MSH2*.

Subject selection

MMR gene variants within consensus splice sites, or missense and small in-frame deletion variants, were identified by DHPLC and sequencing of *MLH1* and *MSH2* exonic regions in patients ascertained via the Australian site of the Cooperative Family Registry for Colorectal Cancer Studies (Colon CFR) (Jenkins, et al., 2006b). The majority of variant carriers were probands from families recruited through family cancer clinics, screened for mutations irrespective of tumour microsatellite instability characteristics, but a small subset were population-based probands recruited through a state-wide cancer registry and selected for mutation screening as a consequence of a microsatellite unstable colorectal tumour phenotype (4 families). Variants that were observed to co-occur with a known deleterious mutation were excluded from the study. A considerable proportion of these samples (69%) had also been previously analysed using MLPA for exonic deletions in *MLH1* and *MSH2*, using the Salsa MLPA Kit P003 (MRC-Holland, Holland). The final sample set

encompassed a total of 22 variants in 28 families, selected on the basis of lymphoblastoid cell line (LCL) availability from colorectal case probands for functional studies (See Table 1). At the time of study initiation, four of these 22 variants were considered deleterious from position alone but had no functional confirmation (*MLH1* c.1668-1G>A, *MLH1* c.1732-1G>A, *MLH1* c.1990-1G>A, *MSH2* c.1077-1G>T). The rest were designated UVs, including the more common variants *MSH2* c.943+3A>T and *MSH2* c.1906G>C p.Ala636Pro now generally considered pathogenic (See Table 2). The Colon CFR variant classifications at the start of this study were as follows: a deleterious mutation was any variant producing a stop codon or nonsense transcript, or having a change at the +1 or -1 position in an intron, or established from a thorough literature and International Society for Gastrointestinal Hereditary Tumours (InSiGHT) database search to portray the characteristics of a pathogenic mutation; a polymorphism was defined by prevalence within normal populations of 1% or greater; and remaining variants were considered to be unclassified. Colorectal case probands all reported white/Caucasian ancestry, with the exception of the proband carrying the *MLH1* c.1732-1G>A sequence variant.

Control screening

Variants were screened in 180 cancer-unaffected population controls reporting Caucasian ethnicity, ascertained as blood donors via the Australian Red Cross Blood Services (Marsh, et al., 2006), using direct sequencing of exonic regions containing variants under study. Variants found to be at polymorphic frequency (>1%) were not pursued further.

Tumor Characteristics

Tumor characteristics were determined for a subset of cancer-affected individuals using methods described previously. Briefly, formalin fixed, paraffin embedded tissue sections were stained for MMR proteins MLH1, MSH2, MSH6 and PMS2 (Lindor, et al., 2002). Tumors were analysed for MSI status (Lindor, et al., 2002) using 10 microsatellite markers (BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197 and MYCL), comparing to normal tissue as reference where possible, and classified according to number of markers demonstrating instability: MSI-H for ≥ 3 ; MSI-L for 1-2; MSS for 0 unstable markers. Tumours with microsatellite unstable phenotype at entry into this study of unclassified variants were from probands carrying the following variants: *MLH1* c.113A>G p.Asn38Ser (MSI-H and PMS2 loss only), *MLH1* c.1852_1854delAAG p.Lys618del (MSI-H and MLH1/PMS2 loss), *MSH2* c.913G>A p.Ala305Thr (MSI-L and normal protein expression), and *MSH2* c.942+3A>T (MSI-H and MSH2/MSH6 loss).

Splicing analysis

Lymphoblastoid cell lines (LCLs) were treated with cycloheximide (two hours at 100 μ g/ml) as previously described (Kirschner et. al. 2000), and RNA extracted using the Qiagen Mini RNA Extraction Kit (Qiagen, Hilden, Germany). First-strand cDNA was then synthesized from 2 μ g of RNA using the Eppendorf cMaster RT kit primed with random hexamers, following the standard protocol. cDNA was amplified in 25 μ l reactions using HotMaster Taq and buffer (Eppendorf, Hamberg, Germany) with 20pmol of each primer and 10ng of DNA, using established primer sets and cycling conditions verified to selectively amplify the target amplicon only (details available on request). cDNA template was sequenced directly after clean-up with Millipore Montage PCR96 Cleanup Plates (Millipore, Bedford, MA, USA) using the standard protocol adjusted for optimal recovery of small PCR products. 1 μ l of the cleaned product was used in for a 1/8-strength sequencing reaction using the BigDye Terminator v3.1 reagents and protocol (Applied Biosystems, Foster City, CA, USA) in a 12 μ l reaction, with 2pmol of the appropriate primer. Reactions were performed in both forward and reverse directions, except for control and relative screening for known variants. Sequencing product was cleaned with the DyeEx 96 Kit (Qiagen, Hilden, Germany) using

the standard protocol. The product was dried, resuspended in HiDi Formamide (Applied Biosystems, Foster City, CA, USA) and analysed on an ABI PRISM 3100 (Applied Biosystems, Foster City, CA, USA). In addition, 10µl of PCR product was run out on 2% agarose gels to visualize aberrant bands using ethidium bromide staining and UV light. Visible bands were excised manually and processed using QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) for sequencing, as above. Results were compared to reference sequence NM_000249.2 for *MLH1*, and NM_000251.1 for *MSH2*. Splicing aberrations were interpreted by comparison of observed results to carriers of other *MLH1* and *MSH2* variants screened at the same time.

Bioinformatic prediction of splicing aberrations

Default thresholds were used for all initial analyses.

Exonic variants were analysed for possible ESE disruption with:

ESEfinder 2.0 (<http://rulai.cshl.edu/tools/ESE>)

RESCUE-ESE (<http://genes.mit.edu/burgelab/rescue-ese>)

PESX (<http://cubweb.biology.columbia.edu/pesx/>)

Intronic variants were analysed using:

SpliceSiteFinder (<http://violin.genet.sickkids.on.ca/~ali/splicesitefinder.html>)

NNSPLICE V0.9 (http://www.fruitfly.org/seq_tools/splice.html)

NetGene2 (<http://www.cbs.dtu.dk/services/NetGene2/>)

For ESEfinder, increased custom thresholds were applied post-analysis (2 for SR protein ASF, and 3 for the other SR proteins), in an attempt to improve the sensitivity of ESE detection, as suggested in the literature (Gorlov, et al., 2004; Pettigrew, et al., 2005)

Segregation analysis

DNA was available from 3 to 53 family members per proband (average 16) for segregation analysis. DNA of relatives was screened for the family variant by direct sequencing using genomic DNA primers (details available on request), using the protocol described above. Results were compared to reference sequence NC_000003.10 for *MLH1*, and NC_000002.10 for *MSH2*. Information for analysis, including cancer status and type of cancer, age at interview/diagnosis, and pedigree structure, was available through the Colon CFR. Bayes factor analysis was performed by computing the likelihood ratio of the pedigree and genotype data under causality compared with neutrality using the approach described in Thompson et al (Thompson, et al., 2003). For this analysis, we assumed the age-specific relative risks estimated in Quehenberger et al (Quehenberger, et al., 2005), and derived penetrance functions using these relative risks and Australian (NSW) population-based age- and sex-specific incidence rates. The derived penetrance model included 14 liability classes, with unaffected male (4 classes: age groups <50, 50-59, 60-69, ≥70), unaffected female (4 classes: age groups <50, 50-59, 60-69, ≥70), affected colorectal cancer (4 classes: age groups <50, 50-59, 60-69, ≥70), endometrial cancer (any age), and minor cancer (small bowel/stomach/ovary/urinary tract, diagnosis <70 only). For individual variants, the probability of pathogenicity was derived from the Bayes scores, assuming a prior probability of pathogenicity of 0.5. This probability of pathogenicity based on segregation was then interpreted according to the 5 class variant classification system of Plon et al (Plon, et al.,

2008) (see below). It is implicit in the penetrance functions used that the term pathogenic indicates a variant has a high risk similar to the average MMR mutations estimated in Quehenberger et al (Quehenberger, et al., 2005), and that the methods used cannot exclude the possibility that a variant may be associated with a moderate or low risk of cancer.

Evolutionary Conservation and Physicochemical Properties of Amino Acid Substitutions

MLH1 and *MSH2* protein multiple sequence alignments were built using M-Coffee (Wallace, et al., 2006), followed by minor hand editing, from sequences available at GenBank, the UC Santa Cruz genome browser, or Ensemble on January 2, 2008. The alignments contained *MLH1/MSH2* orthologs from several mammals, plus chicken, frog, pufferfish, zebrafish, tunicate, sea urchin, and starlet anemone. These alignments, or updated versions thereof, are available online at <<http://agvgd.iarc.fr/alignments.php>>. Sequence conservation at relevant positions in the alignment was assessed by calculating the Grantham Variation (GV), the fit between missense substitutions and the observed range of variation was assessed by calculating the Grantham Deviation (GD), and then the GV-GD coordinate of each substitution was reduced to a 1-dimensional graded score (Tavtigian, et al., 2008; Tavtigian, et al., 2006). The program SIFT was also used, with the same sequence alignments, to assess if amino acid substitutions were likely to affect function (Ng and Henikoff, 2002). Amino acid substitutions were also assessed for pathogenicity using the MAPP-MMR algorithm (Chao, et al., 2008), using the web-based link <http://mappmmr.blueankh.com>. The greatest possible substitution score was noted for the two amino acid deletion variants *MLH1* c.1852_1854delAAG p.Lys618del and *MSH2* c.571_573delCTC p.Leu191del.

Database and Literature Search, and Variant Interpretation

Previous reports of MMR gene variants under study were identified via the Mismatch Repair Genes Variant Database, which provides documentation of published reports of MMR gene variation (www.med.mun.ca/MMRvariants) (Woods, et al., 2007), and from an extensive review of functional analysis of mismatch repair unclassified variants (Ou, et al., 2007), and single recent extensive study of functional abrogation of *MLH1* variants (Takahashi, et al., 2007). The classification of variants as recorded on the InSiGHT MMR gene mutation database (<http://www.insight-group.org/>) was also assessed, as was information on likely pathogenicity, functional assays, and clinical reports on the MMR Gene Missense Mutation Database (<http://www.mmrmissense.net/db/query/query.aspx>). Each variant was placed in one of 5 classes describing their likely pathogenicity, according to the variant classification system of Plon et al (Plon, et al., 2008). For segregation data, class was derived directly from the probability score, as follows: class 5, definitely pathogenic, probability >0.99; class 4, likely pathogenic, probability 0.95-0.99; class 3, uncertain, probability 0.05-0.95; class 2, likely not pathogenic or of little clinical significance, probability 0.001-0.049; class 1, not pathogenic or of little clinical significance, probability <0.001. For *in vitro* splicing data, class was derived by interpretation of cDNA splicing products from variant carriers, according to the suggestions of Spurdle et al (Spurdle, et al., 2008a). For a subset of three variants only, there was insufficient evidence for classification from the current study, but interpretation of published information reporting abrogated function allowed their categorization as Class 5, Pathogenic.

Results

Three of the 22 variants were observed at polymorphic frequency: *MLH1* c.1039-8T>A (which co-occurred with *MLH1* c.198C>T, 9.6%), *MLH1* c.*35_37delCTT (3.3%) and *MSH2* c.965G>A p.Gly322Asp (1.7%). These were not studied further. The results from prediction and *in vitro* analysis of splicing aberrations for the remaining 19 variants are

shown in Table 1, and Figure 1 (a-f). Eight variants were associated with major splicing aberrations generally considered deleterious in a clinical setting (bold, with gray shading). These included the four variants at the +/-1 position and classified deleterious by the Colon-CFR on the basis of nomenclature at study initiation, another variant at the +2 position within the canonical splice junction GT-AG dinucleotides, and three intronic variants at the +5, +3 and -3 positions. In addition, two variants resulted in experimentally detectable splicing defects for which the clinical consequences are more difficult to predict (bold only): *MLH1* c.113A>G p.Asn38Ser caused a frameshift as a minor cDNA product; *MLH1* c.790+2_+3insT altered the ratio of known alternate splice products not observed in any other cyclohexamide-treated samples (Figure 1b).

Bioinformatic prediction of splice site loss for *in vitro*-confirmed splicing aberrations associated with variants at or near intron-exon boundaries was equally good for the splice site prediction programs SpliceSiteFinder (9/10) and NNSPLICE (9/10), but poorer for NetGene2 (6/10). However, while loss of splice sites leading to aberrations was well predicted, this was not the case for existence and alteration of ectopic splice sites, preventing accurate prediction of the *consequences* of splice site loss. Existence of ectopic splice sites was not predicted at all by NetGene2, and was not absolute and did not always correlate for the other two splice site prediction programs.

No splicing aberrations correlated directly with ESE predictions for the nine exonic variants studied. There was no consistency in the prediction across the three programs used, in terms of number of sites altered, and effect of the variant (site created, destroyed, increased or reduced). Overprediction of ESEs was least for PESX (potential ESE/ESS alterations predicted for three variants), and poorer for RESCUE-ESE (six variants predicted to destroy/create/alter ESEs). Prediction using standard thresholds was poorest for ESEfinder (eight variants predicted to alter ESEs), but improved slightly with increased thresholds, which eliminated two of the original predictions.

In conjunction with results from splicing studies, an assessment of the likely pathogenicity may be provided by assessing whether multiple tumors from variant carriers show the features expected for a pathogenic MMR gene variant, namely IHC MMR protein loss consistent with the gene in which the variant exists, and MSI-H phenotype. In addition, segregation of the variant with Lynch syndrome tumors is another feature used to assess pathogenicity of rare gene variants. Results from such analyses are shown in Table 2. Tumor protein expression data was available from variant carriers (probands or family members) for 18 of the 19 variants. The data were consistent with loss of expression of the relevant MMR protein for the vast majority of variants. For 10/18 variants (highlighted in bold), the loss of expression was observed for multiple tumors from carriers of a specific variant, indicating a causative relationship between variant and loss of MMR protein expression. Information in support of causality for another three variants was based on a single tumor only. Exceptions included PMS2 loss only for *MLH1* c.113A>G p.Asn38Ser (2/3 results) and *MLH1* c.198C>T (1 tumor screened), and normal MMR protein expression for *MSH2* c.913G>A p.Ala305Thr and *MLH1* c.2059C>T p.Arg687Trp (1 tumor screened each). MSI data was available for 15/19 variants, and was consistent with the variant being associated with MMR defects for all but one variant, *MSH2* c.913G>A p.Ala305Thr (1 tumor screened). MSI-H status was recorded for multiple tumors from carriers of 10 different variants (highlighted in bold). A general assessment of variant carrier status in genotyped family members affected with Lynch Syndrome associated cancers provided little information for four variants for which there was only a single family with only a single affected carrier. However, in the vast majority of the remaining families, all or most of the cancer-affected relatives carried the variant identified in the proband (noted in bold). Formal analysis of causality gave evidence of pathogenicity for four variants (probability ≥ 0.99),

and suggested causality for another five variants (probability 0.88-0.99). There was no convincing evidence against causality for the remaining variants, in part due to limited information.

Assessment of amino acid evolutionary conservation and physicochemical properties for the eight variants predicted to affect a single amino acid position complemented results from splicing analysis and supported results from causality analysis. There were no splicing aberrations observed for seven of the eight missense alteration variants. Of these seven variants, two (*MLH1* Lys618del and *MSH2* Pro622Leu) fell into the almost certainly pathogenic grade C65, two more (*MSH2* Leu191del and Ala305Thr) in the probably pathogenic grade C55, and two more (*MLH1* Arg687Trp and Thr117Met) in the possible pathogenic grades C25 and C15. SIFT predicted all six of these to affect protein function (score 0.00). The remaining missense substitution, *MSH2* Ala636Pro, occurred at a site demonstrating considerable cross species variability (Ala, Gly, Ser, and Gln were all observed in our alignment at this position), was scored in the likely neutral grade C0 by Align-GVGD, and received an innocuous score of 0.17 from SIFT. The minor product frameshift splicing aberration observed for the conservative substitution *MLH1* Asn38Ser occurred at an evolutionary constrained site, fell in the probably pathogenic Align-GVGD grade C45, and received a likely pathogenic SIFT score of 0.00. MAPP-MMR predicted two variants to be borderline deleterious, and the remainder as deleterious, including *MLH1* Asn38Ser, and *MSH2* Ala636Pro.

An assessment of variant classification as recorded on the InSiGHT database (<http://www.insight-group.org/>) showed that 8 variants were reported, one considered unlikely to be pathogenic, and 7 considered pathogenic. A literature review facilitated by accessing the Mismatch Repair Genes Variant Database (www.med.mun.ca/MMRvariants) identified at least one published report for 14 variants, and the MMR Gene Missense Mutation Database ([Uhttp://www.mmrmissense.net/db/query/query.aspx](http://www.mmrmissense.net/db/query/query.aspx)) was accessed for a review of functional studies of missense variants specifically. For the *MSH2* c.965G>A p.Gly322Asp variant found at polymorphic frequency (>1%) in our control series, there was a single report of reduced function - measured as increased mutation rate in a yeast reporter gene assay (Drotschmann, et al., 1999). There was evidence for functional consequences of the variant(s) under study for 7/20 variants pursued in this study as possibly deleterious (Casey, et al., 2005; Luce, et al., 1995; Otway, et al., 2005; Ou, et al., 2007; Takahashi, et al., 2007), but only three variants had previously been subjected to RNA analysis (from lymphoblastoid cells (Otway, et al., 2005) or LCLs (Luce, et al., 1995), or by conversion analysis - separation of alleles into hybrids and subsequent analysis of mRNA expression and cDNA sequence changes (Casey, et al., 2005)).

For the four variants with no previous report in the InSiGHT, MMR Variant or MMR Gene Missense Mutation databases, three were likely pathogenic or pathogenic from splicing analysis, tumor and/or segregation data (*MLH1* c.113A>G; *MSH2* c.1077-1G>T, *MSH2* c.571_573delCTC). Although the three reports for the *MLH1* c.198C>T variant (Ghimenti, et al., 1999; Palicio, et al., 2002; Wehner, et al., 1997) suggested this variant was not pathogenic, this appeared to be based solely on the fact it is a synonymous substitution, with no functional, tumor or segregation analysis reported.

There were numerous previous reports for *MLH1* variants c.350C>T p.Thr117Met and c.1582_1854delAAG p.Lys618del, supporting the pathogenic classification for these variants by InSiGHT. Importantly, functional studies suggest that these missense alterations result in functional abrogation (Ou, et al., 2007; Takahashi, et al., 2007), as does *MLH1* c.2059C>T p.Arg687Trp, consistent with the fact that no splicing aberrations were detected for these variants in our study. Similarly *MSH2* c.1865C>T p.Pro622Leu and c.1906G>C

p.Ala636Pro were generally considered pathogenic, have been shown to be functionally compromised (Ou, et al., 2007), and showed no splicing aberrations in this study.

The single report for *MLH1* c.116+5 G>C used conversion analysis to identify a 227 nucleotide inclusion of intron 1 (Casey, et al., 2005), as we have done here using relatively straightforward experimental methods. The single report for *MLH1* c.208-3C>G provided evidence from functional analysis of LCL-derived cDNA for an in-frame deletion of exon 3 (Otway, et al., 2005), confirmed in this study. The variant was identified in an individual from a family fulfilling Amsterdam Criteria, but no other information was provided regarding its likely pathogenicity (Otway, et al., 2005). Several reports indicated that *MLH1* c.589-2A>G was associated with features of a high-risk mutation, and a single study using an *in vitro* transcription/translation assay identified this variant to be associated with protein truncation (Luce, et al., 1995), as we would predict from our splicing analyses.

Evidence from reports for *MLH1* c.1668-1G>A, c.1732-1G>A and c.1990-1G>A, and *MSH2* c.942+3A>T, were consistent with pathogenicity, as expected from our splicing analysis showing that these variants are associated with exon deletions. The single report for *MLH1* c.790+2_+3insT included the same family as in this study (Jenkins, et al., 2006a), but did not report estimates of risk for this variant alone.

MSH2 c.913G>A p.Ala305Thr has been reported to be pathogenic by InSIGHT, based on the submission relating to unpublished data of Wijnen et al. However, the very limited tumor data from a published study by the same authors (Wijnen, et al., 1997), and this study, indicate that this variant is not obviously associated with the features of a high-risk mutation other than its identification in an Amsterdam family. Additional functional studies may provide more insight into the clinical relevance of this alteration.

The overall evidence from this and previous studies suggested that 14 variants were high-risk pathogenic and another two likely pathogenic. Overall, 14 variants had aberrant splicing or protein function, including two variants not previously described. In addition, we detected three variants at polymorphic frequency in unaffected controls, one of which had not previously been reported as polymorphic on InSiGHT or in the literature, and another for which a single report suggested functional abrogation.

Discussion

This study was undertaken to assess utility of various methods for assessment of splicing aberrations associated with rare variants in MMR genes. The principle aim was to correlate bioinformatic predictions to splicing aberrations detected *in vitro*, and so provide information for future prioritization of functional studies of such variants. We included variants with varying likelihood of causing splicing aberrations - variants in consensus splice sites (high), variants near consensus splice sites (moderate), and variants predicted to cause missense alterations, many of which are known to have functional consequences relating to protein stability, binding or function (low-nil).

Our results show that use of prediction programs for assessing effects of variants within canonical splice sites are useful to assess if an aberration is likely. However, use of several programs is preferable to predict the precise nature of the splicing aberration(s) that result from a genetic variant (in-frame deletion, frameshift etc), particularly to assess the presence of ectopic splice sites. This indicates that, at present, prediction programs for detection of aberrations associated with variants in canonical splice sites should not be considered a replacement for *in vitro* studies, but rather a means to prioritize variants for *in vitro* analysis and confirmation of any associated splicing defects. Furthermore, although our study does not provide data to allow direct comparison of splicing in LCLs and splicing in the target

disease tissue (colon or endometrial epithelial cells), it does provide support for the utility of splicing assays performed on RNA extracted from LCLs as a means to infer functional consequences of variants due to the detection of loss of splice sites as predicted bioinformatically. This is further supported by the fact that many of the variants shown to have splicing aberrations *in vitro* also had high probability of pathogenicity based on Bayes causality scores, implying that the splicing aberrations observed in LCLs do translate directly into risk

ESEfinder & RESCUE-ESE predictions were poor, consistent with published findings from other *in vitro* studies (Auclair, et al., 2006; Chenevix-Trench, et al., 2006; Lastella, et al., 2006). Use of increased thresholds appears to reduce overprediction, as we have previously shown for ESEfinder (Pettigrew, et al., 2005), but cannot always be easily applied. However, it appears that PESX may be a more conservative tool, as shown in one other study of *MLH1* and *MSH2* variants (Lastella, et al., 2006). While use of a higher threshold (perhaps as high as 6) may improve prediction of ESE/ESS sites by PESX, this would need experimental validation in a large sample set that preferably includes proven ESE alterations leading to splicing aberrations.

Another filter suggested to prioritize UVs within predicted ESEs for *in vitro* analysis is the location of variants and their associated ESEs within 125bp of an intron:exon boundary (Pettigrew, et al., 2005), based on molecular evidence for ESE positions relative exon ends (Cartegni and Krainer, 2002; Majewski and Ott, 2002). This filter was irrelevant for the exonic variants studied here, since they all were located within this limit, but may prove useful in future studies of exonic variants. Other possibilities to be considered to formally improve prediction of splicing aberrations outside of the consensus splice sites could include amino acid evolutionary alignment and physicochemical alterations to identify aberrations at the level of protein stability and protein-protein interaction. It is interesting to note that the 6/7 exonic sequence variants *without* associated splicing aberrations had alterations predicted between almost certainly or possibly pathogenic from the A-GVGD combined evolutionary and physicochemical analysis, and were also predicted to affect protein function by a SIFT position specific scoring matrix analysis, and the MAPP-MMR bioinformatic algorithm.

We have also shown how assessment of tumor characteristics (IHC MMR protein expression, MSI status), and particularly segregation analysis, can be helpful in assessing variant pathogenicity. For eight variants in this study, including two with potentially equivocal interpretation of splicing findings, *MLH1* c.113A>G and c.790+2_+3insT, the segregation analysis alone would permit classification as pathogenic or likely pathogenic. We suggest that sequence variants be evaluated using the compilation of tumor and segregation data in combination with improved bioinformatic prediction methods to predict likely splicing alteration or physicochemical properties of amino acid substitutions. Such methods would assess likely pathogenicity and *also* identify the likely aberrations associated with UVs, to prioritize them for appropriate assays establishing their effect on protein function.

In summary, we have shown that bioinformatic tools for prediction of splicing aberrations need considerable improvement before they may be used without the support of functional studies as a tool to assess pathogenicity of variants in these and other high-risk genes. Our study provided evidence for splicing aberrations associated with ten variants, seven of which had no such data previously available, but which might have been considered to be pathogenic or possibly pathogenic from position alone. We have also shown that, given the availability of resources in terms of tumor characteristics and family segregation information, relatively simple approaches can be used to provide information about the

likely clinical importance of MMR gene UVs, particularly if used in conjunction with functional studies. An important advance in this field of research would be the development of a more formal approach to quantitate likely clinical significance of the large number of MMR gene unclassified sequence variants reported by testing laboratories, such as the multifactorial likelihood analysis methods developed for evaluation of *BRCA1* and *BRCA2* gene variants (Chenevix-Trench, et al., 2006; Goldgar, et al., 2004; Spurdle, et al., 2008b). This would provide a quantitative assessment of risk for improved genetic counseling and management of patients and their family members (Plon, et al., 2008).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the many families who have participated in the research programs, and the Australian Red Cross Blood Services (ARCBS) donors who participated as healthy controls in this study. We are grateful to Rachel Morris and the staff at ARCBS for their assistance with the collection of risk factor information and blood samples, and Joanne Young, Melissa Barker, Melanie Higgins, Kimberley Hinze, Felicity Lose and members of the Molecular Cancer Epidemiology Laboratory for their assistance with collection and processing of blood samples.

Funding: This research was supported by: the Australian National Health and Medical Research Council; the National Cancer Institute; National Institutes of Health under RFA # CA-95-011; through cooperative agreement with Australasian Colorectal Cancer Family Registry (U01 CA097735); by a subaward agreement from the Mayo Clinic, Rochester, MN, funded by the National Institutes of Health Breast Cancer Specialized Program in Research Excellence grant P50 CA116201; and by the Canadian Institutes of Health Research Interdisciplinary Health Research Team Grant (#CRT-43821) and a National Cancer Institute of Canada research fellowship (#13493).

References

- Aretz S, Uhlhaas S, Sun Y, Pagenstecher C, Mangold E, Caspari R, Moslein G, Schulmann K, Propping P, Friedl W. Familial adenomatous polyposis: aberrant splicing due to missense or silent mutations in the APC gene. *Hum Mutat.* 2004; 24(5):370–80. [PubMed: 15459959]
- Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S, et al. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Hum Mutat.* 2006; 27(2):145–54. [PubMed: 16395668]
- Caldes T, Godino J, Sanchez A, Corbacho C, De la Hoya M, Lopez Asenjo J, Saez C, Sanz J, Benito M, Ramon YCS, et al. Immunohistochemistry and microsatellite instability testing for selecting MLH1, MSH2 and MSH6 mutation carriers in hereditary non-polyposis colorectal cancer. *Oncol Rep.* 2004; 12(3):621–9. [PubMed: 15289847]
- Cartegni L, Krainer AR. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet.* 2002; 30(4):377–84. [PubMed: 11925564]
- Casey G, Lindor NM, Papadopoulos N, Thibodeau SN, Moskow J, Steelman S, Buzin CH, Sommer SS, Collins CE, Butz M, et al. Conversion analysis for mutation detection in MLH1 and MSH2 in patients with colorectal cancer. *Jama.* 2005; 293(7):799–809. [PubMed: 15713769]
- Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, Ng P, Gruber SB, Watson P, Rennert G, Anton-Culver H, et al. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). *Hum Mutat.* 2008; 29(6):852–60. [PubMed: 18383312]
- Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T, et al. Genetic and histopathologic evaluation of *BRCA1* and *BRCA2* DNA sequence variants of unknown clinical significance. *Cancer Res.* 2006; 66(4):2019–27. [PubMed: 16489001]
- Cravo M, Afonso AJ, Lage P, Albuquerque C, Maia L, Lacerda C, Fidalgo P, Chaves P, Cruz C, Nobre-Leitao C. Pathogenicity of missense and splice site mutations in hMSH2 and hMLH1

- mismatch repair genes: implications for genetic testing. *Gut*. 2002; 50(3):405–12. [PubMed: 11839723]
- Davoodi-Semiromi A, Lanyon GW, Davidson R, Connor MJ. Aberrant RNA splicing in the hMSH2 gene: molecular identification of three aberrant RNA in Scottish patients with colorectal cancer in the West of Scotland. *Am J Med Genet*. 2000; 95(1):49–52. [PubMed: 11074494]
- Drotschmann K, Clark AB, Tran HT, Resnick MA, Gordenin DA, Kunkel TA. Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. *Proc Natl Acad Sci U S A*. 1999; 96(6):2970–5. [PubMed: 10077621]
- Ghimenti C, Tannergard P, Wahlberg S, Liu T, Giulianotti PG, Mosca F, Fornaciari G, Bevilacqua G, Lindblom A, Caligo MA. Microsatellite instability and mismatch repair gene inactivation in sporadic pancreatic and colon tumours. *Br J Cancer*. 1999; 80(1-2):11–6. [PubMed: 10389971]
- Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet*. 2004; 75(4):535–44. [PubMed: 15290653]
- Gorlov IP, Gorlova OY, Frazier ML, Amos CI. Missense mutations in cancer suppressor gene TP53 are colocalized with exonic splicing enhancers (ESEs). *Mutat Res*. 2004; 554(1-2):175–83. [PubMed: 15450416]
- Hampel H, Stephens JA, Pukkala E, Sankila R, Aaltonen LA, Mecklin JP, de la Chapelle A. Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. *Gastroenterology*. 2005; 129(2):415–21. [PubMed: 16083698]
- Jass JR. Diagnosis of hereditary non-polypoid colorectal cancer. *Histopathology*. 1998; 32(6):491–7. [PubMed: 9675586]
- Jenkins MA, Baglietto L, Dowty JG, Van Vliet CM, Smith L, Mead LJ, Macrae FA, St John DJ, Jass JR, Giles GG, et al. Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study. *Clin Gastroenterol Hepatol*. 2006a; 4(4):489–98. [PubMed: 16616355]
- Jenkins MA, Macrae F, Young J, Barker M, Leggett B, Goldblatt J, Kirk J, Tucker K, Suthers G, Young G, et al. The Colon Cancer Family Registry, an international resource for studying the genetic and molecular epidemiology of colorectal cancer. *Gastrointestinal Cancer Update*. 2006b; 46:14–20.
- Lastella P, Surdo NC, Resta N, Guanti G, Stella A. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. *BMC Genomics*. 2006; 7:243. [PubMed: 16995940]
- Lindor NM, Burgart LJ, Leontovich O, Goldberg RM, Cunningham JM, Sargent DJ, Walsh-Vockley C, Petersen GM, Walsh MD, Leggett BA, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *J Clin Oncol*. 2002; 20(4):1043–8. [PubMed: 11844828]
- Loughrey MB, Waring PM, Tan A, Trivett M, Kovalenko S, Beshay V, Young MA, McArthur G, Boussioutas A, Dobrovic A. Incorporation of somatic BRAF mutation testing into an algorithm for the investigation of hereditary non-polypoid colorectal cancer. *Fam Cancer*. 2007; 6(3):301–10. [PubMed: 17453358]
- Luce MC, Marra G, Chauhan DP, Laghi L, Carethers JM, Cherian SP, Hawn M, Binnie CG, Kam-Morgan LN, Cayouette MC, et al. In vitro transcription/translation assay for the screening of hMLH1 and hMSH2 mutations in familial colon cancer. *Gastroenterology*. 1995; 109(4):1368–74. [PubMed: 7557107]
- Lynch HT, Smyrk T, Lynch JF. Molecular genetics and clinical-pathology features of hereditary nonpolyposis colorectal carcinoma (Lynch syndrome): historical journey from pedigree anecdote to molecular genetic confirmation. *Oncology*. 1998; 55(2):103–8. [PubMed: 9499183]
- Majewski J, Ott J. Distribution and characterization of regulatory elements in the human genome. *Genome Res*. 2002; 12(12):1827–36. [PubMed: 12466286]
- Marsh A, Healey S, Lewis A, Spurdle AB, Kedda MA, Pupo GM, Lakhani SR, Chenevix-Trench G. Mutation analysis of five candidate genes. *Breast Cancer Res Treat*. 2006 Dec 23 [epub ahead of print].

- McVety S, Li L, Gordon PH, Chong G, Foulkes WD. Disruption of an exon splicing enhancer in exon 3 of MLH1 is the cause of HNPCC in a Quebec family. *J Med Genet.* 2005
- Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* 2002; 12(3):436–46. [PubMed: 11875032]
- Otway R, Tetlow N, Hornby J, Doe WF, Kohonen-Corish MR. Gene symbol: MLH1. Disease: Hereditary nonpolyposis colorectal cancer. *Hum Genet.* 2005; 116(6):534. [PubMed: 15988827]
- Ou J, Niessen RC, Lutzen A, Sijmons RH, Kleibeuker JH, de Wind N, Rasmussen LJ, Hofstra RM. Functional analysis helps to clarify the clinical importance of unclassified variants in DNA mismatch repair genes. *Hum Mutat.* 2007; 28(11):1047–1054. [PubMed: 17594722]
- Palicio M, Balmana J, Gonzalez S, Blanco I, Marcuello E, Peinado MA, Julia G, Germa JR, Lopez Lopez JJ, Brunet J, et al. Mismatch repair gene analysis in Catalan families with colorectal cancer. *J Med Genet.* 2002; 39(6):E29. [PubMed: 12070261]
- Pettigrew C, Wayte N, Lovelock PK, Tavtigian SV, Chenevix-Trench G, Spurdle AB, Brown MA. Evolutionary conservation analysis increases the colocalization of predicted exonic splicing enhancers in the BRCA1 gene with missense sequence changes and in-frame deletions, but not polymorphisms. *Breast Cancer Res.* 2005; 7:R929–R939. [PubMed: 16280041]
- Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FBL, Hoogerbrugge N, Spurdle AB, Tavtigian SV. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human Mutation.* 2008 in press.
- Quehenberger F, Vasen HF, van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. *J Med Genet.* 2005; 42(6):491–6. [PubMed: 15937084]
- 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(3):272. [PubMed: 15300854]
- Southey MC, Jenkins MA, Mead L, Whitty J, Trivett M, Tesoriero AA, Smith LD, Jennings K, Grubb G, Royce SG, et al. Use of molecular tumor characteristics to prioritize mismatch repair gene testing in early-onset colorectal cancer. *J Clin Oncol.* 2005; 23(27):6524–32. [PubMed: 16116158]
- Spurdle AB, Couch FJ, Hogervorst FBL, Radice P, Sinilnikova OM, Group IUGVW. Prediction and Assessment of Splicing Alterations: Implications for clinical testing. *Human Mutation.* 2008a in press.
- Spurdle AB, Lakhani SR, Healey S, Parry S, Da Silva LM, Brinkworth R, Hopper JL, Brown MA, Babikyan D, Chenevix-Trench G, et al. Clinical classification of BRCA1 and BRCA2 DNA sequence variants: the value of cytokeratin profiles and evolutionary analysis--a report from the kConFab Investigators. *J Clin Oncol.* 2008b; 26(10):1657–63. [PubMed: 18375895]
- Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. *Cancer Res.* 2007; 67(10):4595–604. [PubMed: 17510385]
- Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Human Mutation.* 2008 in press.
- Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet.* 2006; 43(4):295–305. [PubMed: 16014699]
- Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet.* 2003; 73(3):652–5. [PubMed: 12900794]
- Viel A, Novella E, Genuardi M, Capozzi E, Fornasari M, Pedroni M, Santarosa M, De Leon MP, Della Puppa L, Anti M, et al. Lack of PMS2 gene-truncating mutations in patients with hereditary colorectal cancer. *Int J Oncol.* 1998; 13(3):565–9. [PubMed: 9683794]
- Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* 2006; 34(6):1692–9. [PubMed: 16556910]
- Wang Q, Lasset C, Desseigne F, Saurin JC, Maugard C, Navarro C, Ruano E, Descos L, Trillet-Lenoir V, Bosset JF, et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and

- hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. *Hum Genet.* 1999; 105(1-2):79–85. [PubMed: 10480359]
- Wehner M, Buschhausen L, Lamberti C, Kruse R, Caspari R, Propping P, Friedl W. Hereditary nonpolyposis colorectal cancer (HNPCC): eight novel germline mutations in hMSH2 or hMLH1 genes. *Hum Mutat.* 1997; 10(3):241–4. [PubMed: 9298827]
- Wijnen J, Khan PM, Vasen H, van der Klift H, Mulder A, van Leeuwen-Cornelisse I, Bakker B, Losekoot M, Moller P, Fodde R. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. *Am J Hum Genet.* 1997; 61(2):329–35. [PubMed: 9311737]
- Woods MO, Williams P, Careen A, Edwards L, Bartlett S, McLaughlin JR, Younghusband HB. A new variant database for mismatch repair genes associated with Lynch syndrome. *Hum Mutat.* 2007; 28(7):669–73. [PubMed: 17347989]
- Zatkova A, Messiaen L, Vandenbroucke I, Wieser R, Fonatsch C, Krainer AR, Wimmer K. Disruption of exonic splicing enhancer elements is the principal cause of exon skipping associated with seven nonsense or missense alleles of NF1. *Hum Mutat.* 2004; 24(6):491–501. [PubMed: 15523642]

Figure 1.

PCR amplification of LCL-derived cDNA from variant carriers

a. The target amplicon incorporates c.-28 (5'UTR) to c.1426 of *MLH1*. Lanes A (variant *MSH2* c.913G>A) and D (variant *MLH1* c.1668-1G>A) display normal wildtype product, with two additional fainter products representing native alternately-spliced mRNA (exon 9 exclusion, exon 9 and 10 exclusion). Lane B (variant *MLH1* c.116+5G>C) shows the inclusion of 227bp (uppermost product, Arrow 1) and a concatenated wildtype/variant product (above wildtype product). Lane C (variant *MLH1* c.208-3C>G) shows the exclusion of exon 3 (Arrow 2) overlying the native exon 9 exclusion product.

b. The target amplicon incorporates c.441 to c.1054 of *MLH1*. Lanes A (variant *MSH2* c.571_573delCTC), B (variant *MSH2* c.942+3A>T) and D (variant *MSH2* c.913G>A) display normal wildtype product, with fainter product (Arrow 1) comprising a concatenation of wildtype and native alternately-spliced mRNA excluding exon 9. Lane C (variant *MLH1* c.790+2_+3insT) shows increased concatenation of wildtype and exon 9 exclusion (Arrow 1, Lane C). Genotyping for the presence of an *in cis* 655A>G polymorphism in exon 8 confirmed that no wildtype mRNA expressed by the variant allele, and that all normal length mRNA in both upper bands (Arrows 1 and 2, Lane C) was derived from the wildtype allele. The product indicated by Arrow 3 comprises variant-derived product exclusive of exon 9, and that indicated by Arrow 4 comprises a variant-derived product exclusive of both exons 9 and 10.

c. The target amplicon incorporates c.1392 to c.1912 of *MLH1*. Lanes A (variant *MSH2* c.1906G>A), C (variant *MSH2* c.942+3A>T) and D (variant *MSH2* c.913G>A) display normal wildtype band weights. In lane B (variant *MLH1* c.1668-1G>A), the band indicated by Arrow 1 comprises a concatenation of wildtype and exon 15 exclusion products, while the product indicated by Arrow 2 comprises exon 15 exclusion only.

d. The target amplicon incorporates c.1392 to c.2278 of *MLH1*. Lanes A (variant *MSH2* c.913G>A), B (variant *MSH2* c.1865C>T) and D (variant *MSH2* c.1906G>C) display normal wildtype product. In lane B (variant *MLH1* c.1732-1G>A), the band indicated by Arrow 1 comprises a concatenation of wildtype and exon 16 exclusion products, while the band indicated by Arrow 2 comprises a exon 16 exclusion product only.

e. The target amplicon incorporates c.1781 to c.2278 of *MLH1*. Lanes A (variant *MSH2* c.1865C>T), B (variant *MSH2* c.1906G>C) and D (variant *MSH2* c.571_573delCTC) display normal wildtype products. In lane C (variant *MLH1* c.1990-1 g>A) the band indicated by Arrow 1 comprises a concatenation of wildtype and exon 18 exclusion product, while Arrow 2 comprises the exon 18 exclusion product only.

f. The target amplicon incorporates c.627 to c.1296 of *MSH2*. Lanes A (variant *MLH1* c.1732-1G>A), B (variant *MLH1* c.208-3C>G) and D (variant *MLH1* c.116+5G>C) display normal wildtype products. In lane C (variant *MSH2* c.942+3A>T), the band indicated by Arrow 1 comprises a concatenation of wildtype and exon 5 exclusion product, while the band indicated by Arrow 2 comprises the exon 5 exclusion product only.

Table 1
Informatic prediction and *in vitro* analysis of splicing aberrations associated with *MLH1* or *MSH2* variant

Exon	Variant	Variant - Alternative Nomenclature	ESEfinder	RESCUE-ESE	PESX	SpliceSiteFinder	NNSPLICE V 0.9.	NetGene2	Results from cDNA analysis ^d
			Searches for ESEs only. Threshold alteration possible. Score provided - allows for subsequent filtering. Sites that pass increased threshold filtering are noted in red.	Searches for ESEs only. No threshold alteration possible, no score provided	Searches for ESE and ESS sites. Default thresholds provide classification but no score (postanalysis filtering not possible): Weak > 2.62 but < 3; Very weak < 2.62 but > 2.2	In-house algorithm, uses & accounts for strength of upstream branchpoint (possibly better prediction of acceptor sites).	No use of branchpoint sequence; Standard threshold = 0.4. Has wider coverage (possibly better prediction of donor sites)	In-house algorithm, cannot alter thresholds and extent of coverage is not clear.	Results for sequencing and gel electrophoresis of cDNA products
1	c.113A>G p.Asn38Ser	113A>G Asn38Ser	SRp40 site created abutting splice site (2.9)	4/4 ESEs destroyed.	Weak ESE destroyed. Very weak ESS created.	Primary splice site score 76. Ectopic splice site at c.116+227 has score 79. SpliceSiteFinder does not analyze effect of this substitution.	Primary splice site reduced (0.93 > 0.74). Ectopic splice site at c.116+227 unaltered (0.97).	No difference reported.	Normal profile as major product. Inclusion of 227nt of intron 1 causes frameshift, as minor product (~10%)
1	c.116+5G>C	IVS1+5G>C	---	---	---	Primary splice site destroyed (76 > 62). Ectopic splice site at c.116+227 (79) unaltered.	Primary splice site (0.93) destroyed. Ectopic splice site at c.116+227 (0.97) unaltered.	Primary splice site destroyed.	Inclusion of 227nt of intron 1 causes frameshift
2	c.198C>T	198C>T Thr66Thr	SF2 site reduced (4.9 > 2.8). SRp40 site created (3.6)	No ESE found.	No ESE or ESS found.	SpliceSiteFinder does not analyze effect of this substitution.	NNSPLICE does not analyze effect of this substitution.	No difference reported.	C>T base substitution only.
3	c.208-3C>G	IVS2-3C>G	---	---	---	Primary splice site decreased (83 > 71). Creates ectopic splice site (78.2) 2nt upstream with possible branch support.	Primary splice site (0.67) destroyed. Ectopic splice site 2nt upstream not recognised.	No sites recognised (wildtype or variant sequence).	In-frame deletion of exon 3

Exon	Variant	Variant - Alternative Nomenclature	ESEfinder	RESCUE-ESE	PESX	SpliceSiteFinder	NNSPLICE V 0.9.	NetGene2	Results from cDNA analysis ^a
4	c.307-29C>A	IVS3-29C>A	---	---	---	Not recognized as a viable branchpoint.	NNSPLICE does not analyze effect of this substitution.	No difference reported.	Normal profile
4	c.350C>T p.Thr117Met	350C>T Thr117Met	SRp40 site increased (4.6 > 5.4)	4/4 ESEs destroyed. 2 ESEs created.	Weak ESE destroyed.	---	---	---	C>T base substitution only
8	c.589-2A>G	IVS7-2A>G	---	---	---	Primary splice site (88.3) destroyed. Ectopic splice site 4nt downstream is not recognized.	Primary splice site (0.68) destroyed. Ectopic splice site 4nt downstream is not recognized.	No sites recognised (wildtype or variant sequence, or ectopic site).	Use of ectopic acceptor site 4 nt downstream causes frameshift
9	c.790+2G>T	IVS9+2 ₋ +3insT	---	---	---	Splice site (78.3) destroyed.	Splice site (0.96) destroyed.	Splice site destroyed.	No full-length product produced by variant allele, only known products alternate splice products missing Ex 9 & Ex9/10
15	c.1668-1G>A	IVS14-1 G>A	---	---	---	Splice site (93.3) destroyed.	Splice site (0.96) destroyed.	Splice site destroyed.	Out-of-frame deletion of exon 15
16	c.1732-1G>A	IVS15-1G>A	---	---	---	Splice site (88) moved to 1nt downstream (80). Identical branchpoint used by ectopic site.	Splice site (0.99) moved to 1 nt downstream (0.74)	Splice site destroyed.	In-frame deletion of exon 16
16	c.1852_1854delAAG p.Lys618del	1852_1854delAAG Lys618del	No ESE found.	3/4 overlapping ESEs destroyed, 1/4 moved. 2/4 have duplicate ESEs 3bp upstream due to AAG repeat.	2/3 overlapping ESEs destroyed (scores >4).	---	---	In-frame deletion of AAG	In-frame deletion of AAG
18	c.1990-1G>A	IVS17-1G>A	---	---	---	Splice Site (90.6) destroyed.	Splice site (0.57) destroyed.	Splice site destroyed.	In-frame deletion of exon 18
18	c.2059C>T p.Arg687Trp	2059C>T Arg687Trp	SF2/ASF site (2.6) destroyed.	1 ESE created.	1 ESE, 1 weak ESS created.	---	---	---	C>T base substitution only

Exon	Variant	Variant - Alternative Nomenclature	ESEfinder	RESCUE-ESE	PESX	SpliceSiteFinder	NNSPLICE V 0.9.	NetGene2	Results from cDNA analysis ^a
3	c.571_573delCTC p.Leu191del	571_573delCTC Leu191del	SC35 site created (3.4)	No ESEs found.	1 very weak ESE unaltered. 1 very weak ESS created.	---	---	---	In-frame deletion of CTC
5	c.913G>A p.Ala305Thr	913G>A Ala305Thr	SC35 site reduced (3.0 > 2.6). SRp55 site reduced (4.7 > 4.1). SRp40 site created (5.2).	No ESEs found.	1 weak ESE destroyed.	---	---	---	G>A base substitution only
5	c.942+3A>T	IVS5+3A>T	---	---	---	Splice site reduced (79.9 > 69.9)	Splice Site marginally increased (0.96 to 0.97)	No sites recognised (wildtype or variant sequence).	In-frame deletion of exon 5
7	c.1077-1G>T	IVS6-1G>T	---	---	---	Primary splice site destroyed (88.7). Ectopic splice site, c. 1077-83 (83.3)	Primary splice site (0.91) destroyed.	No sites recognised (wildtype or variant sequence).	In-frame deletion of exon 7 as major product. Use of ectopic acceptor site c. 1077-83 causes an out-of-frame insertion of 83nt. as minor product (at least 10%)
12	c.1865C>T p.Pro622Leu	1865C>T Pro622Leu	SF2/ASF site (2.4) destroyed. SRp40 site increased (2.8 > 4.3).	1 ESE created.	3 ESEs destroyed.	---	---	---	C>T base substitution only
12	c.1906G>C p.Ala636Pro	1906G>C Ala636Pro	SRp40 site (3.9) destroyed. SRp55 site (3.6) destroyed. SRp40 site created (3.0).	2/2 ESEs destroyed. 2 ESEs created.	1 very weak ESE unaltered. 1 very weak ESS created.	---	---	---	G>C base substitution only.

^a In bold with grey highlight indicate major aberrations likely to be considered pathogenic. Result in bold only indicates splicing aberration is minor product or altered level of naturally occurring splice product.

w and splicing analysis.

Bayes Causality Score ^b	Probability of Pathogenicity, based on Bayes Score ^b	Grantham Variation ^c	Grantham Deviation ^c	A-GYGD Class ^c	SIFT Class	MAPP-MMR prediction	Classification and Evidence from InSIGHT ^d	Classification and Evidence from Literature Search ^e	Results from cDNA splicing analysis ^f	Conclusion Regarding Variant Classifications ^g
27.47	0.96	0.00	46.24	C45	0.00	13.40	NR	NR	Normal profile as major product. Inclusion of 227nt of intron 1 causes frameshift, as minor product (~10%)	Class 4 Variant - Likely Pathogenic
21.68	0.96	N/A	N/A	N/A	N/A	N/A	NR	P - Conversion analysis shows that variant associated with use of cryptic splice site and production of longer cDNA	Inclusion of 227nt of intron 1 causes frameshift	Class 5 Variant - Pathogenic
0.69	0.41	N/A	N/A	N/A	N/A	N/A	not P	not P - Variant is synonymous substitution.	C>T base substitution only.	Class 3 Variant - Uncertain
7.49	0.88	N/A	N/A	N/A	N/A	N/A	NR	P - Variant associated with in-frame deletion of exon 3	In-frame deletion of exon 3	Class 5 Variant - Pathogenic
1.02	0.51	N/A	N/A	N/A	N/A	N/A	NR	NR	Normal profile	Class 3 Variant - Uncertain

Hum Mutat. Author manuscript; available in PMC 2010 May 1.

Bayes Causality Score ^b	Probability of Pathogenicity, based on Bayes Score ^b	Grantham Variation ^c	Grantham Deviation ^c	A-GVGD Class ^c	SIFT Class	MAPP-MMR prediction	Classification and Evidence from InSIGHT ^d	Classification and Evidence from Literature Search ^e	Results from cDNA splicing analysis ^f	Conclusion Regarding Variant Classification ^g
1043193.34	>0.99	0.00	81.04	C15	0.00	21.05	P	P - Variant occurs in Amsterdam families, segregates with disease, and is associated with reduced dimerization and abrogated function	C>T base substitution only.	Class 5 Variant - Pathogenic
1.23	0.55	N/A	N/A	N/A	N/A	N/A	P	P - Variant is associated with MSI, early diagnosis, multiple HNPCC tumors, and splicing aberration leading to truncation	Use of ectopic acceptor site 4 nt downstream causes frameshift	Class 5 Variant - Pathogenic
312.75	>0.99	N/A	N/A	N/A	N/A	N/A	NR	P - Variant identified in an Amsterdam family, included in this study.	No full-length product produced by variant allele, only known alternate splice products missing Ex 9 & Ex9/10	Class 5 Variant - Pathogenic
0.55	0.35	N/A	N/A	N/A	N/A	N/A	NR	P - Variant associated with tumor MSI and/or MMR protein loss in Bethesda families.	Out-of-frame deletion of exon 15	Class 5 Variant - Pathogenic

Hum Mutat. Author manuscript; available in PMC 2010 May 1.

Bayes Causality Score ^b	Probability of Pathogenicity, based on Bayes Score ^b	Grantham Variation ^c	Grantham Deviation ^c	A-GVGD Class ^c	SIFT Class	MAPP-MMR prediction	Classification and Evidence from InSIGHT ^d	Classification and Evidence from Literature Search ^e	Results from cDNA splicing analysis ^f	Conclusion Regarding Variant Classification ^g
3.97	0.80	N/A	N/A	N/A	N/A	N/A	NR	P - Variant associated with tumor MSI and MMR protein loss in Bethesda and Amsterdam families	In-frame deletion of exon 16	Class 5 Variant - Pathogenic
579.90	>0.99	0.00	126.46	C65	0.00	>9.740	P	P - associated with cancer phenotype in Bethesda and Amsterdam families, and abrogated function.	In-frame deletion of AAG	Class 5 Variant - Pathogenic
No score possible	N/A	N/A	N/A	N/A	N/A	N/A	NR	P - Variant associated with MSI and MMR protein loss in Amsterdam family	In-frame deletion of exon 18	Class 5 Variant - Pathogenic
0.81	0.45	56.60	101.29	C25	0.00	3,870 - borderline deleterious	NR	P - Variant associated with cancer phenotype, tumor MSI and MMR protein loss in Amsterdam families, and abrogated function	C>T base substitution only	Class 5 Variant - Pathogenic
19.81	0.95	14.30	Very High	C55	0.00	>39,390	NR	NR	In-frame deletion of CTC	Class 4 Variant - Likely Pathogenic

Hum Mutat. Author manuscript; available in PMC 2010 May 1.

Bayes Causality Score ^b	Probability of Pathogenicity, based on Bayes Score ^b	Grantham Variation ^c	Grantham Deviation ^c	A-GVGD Class ^c	SIFT Class	MAPP-MMR prediction	Classification and Evidence from InSIGHT ^d	Classification and Evidence from Literature Search ^e	Results from cDNA splicing analysis ^f	Conclusion Regarding Variant Classification ^g
0.52	0.34	0.00	58.02	C55	0.00	3.550 - borderline deleterious	P	Unclassified Variant - Variant identified in an individual from an Amsterdam family, but tumor examined did not display MSI	G>A base substitution only	Class 3 Variant - Uncertain
27.66	0.97	N/A	N/A	N/A	N/A	N/A	P	P - associated with cancer phenotype in Bethesda and Amsterdam families.	In-frame deletion of exon 5	Class 5 Variant - Pathogenic
2.84	0.74	N/A	N/A	N/A	N/A	N/A	NR	NR	In-frame deletion of exon 7 as major product. Use of ectopic acceptor site c. 1077-83 causes an out-of-frame insertion of 83nt, as minor product (at least 10%)	Class 5 Variant - Pathogenic
67134.69	>0.99	0.00	97.78	C65	0.00	26.92	P	P - associated with cancer phenotype in Bethesda and Amsterdam families, and	C> base substitution only	Class 5 Variant - Pathogenic

Hum Mutat. Author manuscript; available in PMC 2010 May 1.

Bayes Causality Score ^b	Probability of Pathogenicity, based on Bayes Score ^b	Grantham Variation ^c	Grantham Deviation ^c	A-GVGD Class ^c	SIFT Class	MAPP-MMR prediction	Classification and Evidence from InSIGHT ^d	Classification and Evidence from Literature Search ^e	Results from cDNA splicing analysis ^f	Conclusion Regarding Variant Classification ^g
1.19	0.54	133.90	1.62	C0	0.17	8.730	P	abrogated function. P - associated with cancer phenotype in Bethesda and Amsterdam families, and abrogated function. Ashkenazi founder mutation.	G>C base substitution only.	Class 5 Variant - Pathogenic

Hum Mutat. Author manuscript; available in PMC 2010 May 1.

elvis, central nervous system, small bowel, sebaceous skin,
ability of pathogenicity is derived as (Bayes score)/(Bayes score
ld with
single codon deletions (1). Studies of the BRCA1 and BRCA2
tion is minor product or altered level of alternate splice product