HUMAN MUTATION 29(6), 852–860, 2008

RESEARCH ARTICLE

Accurate Classification of *MLH1/MSH2* Missense Variants With Multivariate Analysis of Protein Polymorphisms–Mismatch Repair (MAPP-MMR)

Elizabeth C. Chao,[1] Jonathan L. Velasquez,[1] Mavee S.L. Witherspoon,[1] Laura S. Rozek,[2] David Peel,[1] Pauline Ng,[3] Stephen B. Gruber,[2] Patrice Watson,[4] Gad Rennert,[5,6] Hoda Anton-Culver,[1] Henry Lynch,[4] and Steven M. Lipkin[1]*

[1]Genetic Epidemiology Research Institute, University of California, Irvine, Irvine, California; [2]Department of Internal Medicine, Epidemiology, and Human Genetics, University of Michigan, Ann Arbor, Michigan; [3]J. Craig Venter Institute for Human Genetics, Rockville, Maryland; [4]Hereditary Cancer Institute, Creighton University School of Medicine, Omaha, Nebraska; [5]Department of Community Medicine and Epidemiology, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel; [6]CHS National Cancer Control Center, Haifa, Israel

Communicated by Marc Greenblatt

Lynch syndrome, also known as hereditary nonpolyposis colon cancer (HNPCC), is the most common known genetic syndrome for colorectal cancer (CRC). *MLH1/MSH2* mutations underlie approximately 90% of Lynch syndrome families. A total of 24% of these mutations are missense. Interpreting missense variation is extremely challenging. We have therefore developed multivariate analysis of protein polymorphisms–mismatch repair (MAPP-MMR), a bioinformatic algorithm that effectively classifies *MLH1/MSH2* deleterious and neutral missense variants. We compiled a large database (n > 300) of *MLH1/MSH2* missense variants with associated clinical and molecular characteristics. We divided this database into nonoverlapping training and validation sets and tested MAPP-MMR. MAPP-MMR significantly outperformed other missense variant classification algorithms (sensitivity, 94%; specificity, 96%; positive predictive value [PPV] 98%; negative predictive value [NPV], 89%), such as SIFT and PolyPhen. MAPP-MMR is an effective bioinformatic tool for missense variant interpretation that accurately distinguishes *MLH1/MSH2* deleterious variants from neutral variants. Hum Mutat 29(6), 852–860, 2008. **© 2008 Wiley-Liss, Inc.**

KEY WORDS: Lynch syndrome; colorectal cancer; variants of uncertain significance; cancer genetics; HNPCC; *MLH1*; *MSH*

INTRODUCTION

MLH1 (MIM# 120436) and *MSH2* (MIM# 609309) mutations underlie 90% of Lynch syndrome [Lynch and de la Chapelle, 2003] or hereditary nonpolyposis colon cancer (HNPCC). The clinical phenotype of *MLH1/MSH2* mutations ranges from Bethesda guidelines (BG) [Rodriguez-Bigas et al., 1997; Umar et al., 2004] to familial colorectal cancer (CRC) (defined as proband plus either one affected first-degree relative or two affected second-degree relatives), and even sporadic CRC [Peltomaki et al., 2004]. Part of the variation in clinical phenotypes is attributable to missense mutations. A total of 24% of Lynch syndrome mutations are missense [Peltomaki et al., 2004]. There is widespread consensus among cancer geneticists that distinguishing deleterious from neutral variants is challenging. Supporting evidence that a missense variant is functionally relevant and confers risk includes microsatellite instability (MSI), cosegregation with affected relatives, tumor immunohistochemistry (IHC), biochemical analyses, and case–control studies. Many missense variants have minimal or conflicting supporting evidence (termed variants of uncertain significance [VUS]).

SIFT (http://blocks.fhcrc.org/sift/SIFT.html) and PolyPhen (http://coot.embl.de/PolyPhen) are bioinformatic algorithms that use evolutionary history and physicochemical parameters to interpret missense variants. They are ~60 to 80% accurate in test studies [Chan et al., 2007; Chasman and Adams, 2001; Ng and Henikoff, 2001; Raevaara et al., 2005; Sunyaev et al., 2000b, 2001; Xi et al., 2004]. A recent study directly compared these, and other, algorithms, to interpret missense variants in several proteins, including *MLH1* and *MSH2* [Chan et al., 2007]. This study confirmed a high predictive value for algorithms that use evolutionary sequence conservation, with or without considering protein structural change, to predict the clinical consequences of missense variants [Chan et al., 2007]. A newer algorithm, multivariate analysis of protein polymorphisms (MAPP) [Stone and Sidow, 2005], utilizes a similar approach and is more

The Supplementary Material referred to in this article can be accessed at http://www.interscience.wiley.com/jpages/1059-7794/suppmat.

Received 7 June 2007; accepted revised manuscript 21 December 2007.

*Correspondence to: Steven M. Lipkin, MD, PhD, Director, Cancer Genetics Clinic, Division of Hematology-Oncology, University of California, Irvine, 204 Sprague Hall, ZC 4038, Irvine, CA 92697.
E-mail: slipkin@uci.edu

DOI 10.1002/humu.20735
Published online 27 March 2008 in Wiley InterScience (www.interscience.wiley.com).

© 2008 WILEY-LISS, INC.

accurate. Importantly, MAPP can be customized for specific proteins. We have therefore optimized MAPP specifically for *MLH1/MSH2* and developed MAPP-mismatch repair (MMR), an application that distinguishes *MLH1/MSH2* deleterious variants from neutral variants more effectively than any other current computational approach.

METHODS

Database of *MLH1/MSH2* Missense Variants

We compiled a database of all known *MLH1/MSH2* missense variants (Supplementary Table S1; available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat) from: the Mismatch Repair Genes Variant Database (www.med.mun.ca/mmrvariants); InSIGHT (www.insight-group.org); additional MEDLINE articles [Blasi et al., 2006; Lipkin et al., 2004]; Database of Single Nucleotide Polymorphisms (dbSNP, NCBI; www.ncbi.nlm.nih.gov); Myriad Genetics, unpublished data; Seattle SNP Discovery Project (SeattleSNPs; http://pga.gs.washington.edu); HapMap (www.hapmap.org); and missense variants from individuals with no personal or family history of cancer. All variants were reviewed and annotated with all supporting data available. Supporting data include: age at first CRC diagnosis, affected relatives with CRC, and missense mutation cosegregation/linkage, MSI status, IHC status, molecular epidemiology, and/or detailed biochemical functional analysis. Supporting data was used to classify missense variants as deleterious or neutral. Rigorous criteria were used to ascertain definitive classes of deleterious and neutral variants. To ensure specificity of neutral variant classification, any missense variant originally described in a CRC proband was limited to classification as a variant of uncertain significance or deleterious variant. Missense variants classified as deleterious satisfied ≥ 2 of the following criteria: 1) positive segregation analysis in >1 family member with CRC or endometrial cancer, or 2) abnormal result in ≥ 1 of the following biochemical studies: nuclear localization/*PMS2* colocalization, protein stability in cell culture, or functional assays; or 3) MSI-H in at least one tumor; 4) IHC-negative status; or 5) molecular epidemiology studies demonstrating significantly increased missense variant frequency in CRC cases vs. controls in ≥ 2 nonoverlapping cohorts. Variants that met only one of these criteria were classified as VUS. Criteria for neutral variants: 1) identification in groups of control subjects for SNP discovery (e.g., dbSNP), or 2) missense variants identified in control subjects with no personal or family history of CRC or endometrial cancer (Table 1). The entire set of deleterious variants was divided into two independent sets for algorithm training (n = 24) and validation (n = 25).

MAPP-MMR is modified from multivariate analysis of protein polymorphisms (MAPP) (http://mendel.stanford.edu/SidowLab/downloads.html) [Stone and Sidow, 2005]. At each amino acid position in a multisequence alignment of *MLH1* or *MSH2* orthologs, MAPP-MMR infers the evolutionary relationships of the protein sequence by likelihood analysis. Weights for each sequence position are quantified for the physiochemical properties of all amino acids occurring at that position for the following parameters: hydropathy, polarity, charge, side-chain volume, free energy in α-helix conformation, and free energy in β-sheet conformation. For each physiochemical property parameter at that position, the mean and standard deviation are calculated based on each amino acid appearing in the alignment. The summary of these measures is an estimate of the physicochemical constraint at each position, and a standard normalized score for each new amino acid is calculated using the observed mean and variance for that position. A large standard normalized score indicates that the amino acid analyzed is dissimilar from the amino acids appearing in the ortholog alignment, and hence is more likely to affect protein function. For a more detailed description of the MAPP algorithm, please see Stone and Sidow [2005].

To develop MAPP-MMR we customized input sequences, evolutionary conservation gap-weight threshold, and impact score threshold. Initially, default parameters for MAPP were used (gap weight threshold = 50% and impact score threshold = 5.0 [Stone and Sidow, 2005]). Sequence alignments and unrooted phylogenetic trees were constructed using ClustalW (www.ebi.ac.uk/Tools/clustalw). Initial training and optimization sequence sets included: 1) all NCBI Blast hits with input *MLH1* (gi:4557757) and *MSH2* (gi:4557761); 2) all unique *MLH1/MSH2* orthologs and homologs; 3) *MLH1/MSH2* eukaryotic orthologs only; 4) *MLH1/MSH2* vertebrate orthologs only; and 5) nonredundant prokaryotic and eukaryotic orthologs. MAPP-MMR scores from these sequence sets were calculated for all missense variants. Accuracy of MAPP-MMR predictions was calculated based on impact scores for the training subset of deleterious variants and all neutral variants. Initially, sequence set ♯5 performed optimally and was used for further optimization.

Gap weight represents an important measure of amino acid conservation in a multisequence alignment [Ng and Henikoff,

TABLE 1. **Criteria for *MLH1/MSH2* Missense Variant Classification Based on Experimental and Published Data**

Deleterious variants	Variant of uncertain significance (VUS)	Neutral variants
Found in proband with either early onset CRC or first degree relative with CRC	Found in proband with either early onset CRC or 1st degree relative with CRC	Identified in a control population of healthy individuals
AND	AND	AND
≥2 of the following	Not meeting criteria for deleterious variants	No evidence of pathogenicity in any published report or any publicly available SNP database
MSI-H	OR	
OR	Conflicting evidence regarding variant assignation as deleterious	
Cosegregation /linkage of the variant with disease in ≥1 family member		
OR		
Biochemical evidence identifying functional mismatch repair deficiency including negative immunohistochemistry		
OR		
Molecular epidemiology study in ≥2 nonoverlapping groups demonstrating significantly increased CRC risk		

2001]. To optimize predictive accuracy we next varied gap weight threshold from 25 to 75%. A gap weight threshold of 67% improved MAPP-MMR sensitivity without sacrificing specificity. MAPP-MMR was further tested with two additional sequence inputs (numbering continued from previous paragraph): 6) nonredundant prokaryotic and eukaryotic orthologs for domains with gap weight <67% (Supplementary Table S2) and eukaryotic orthologs only for domains with >67% gap weight; and 7) nonredundant prokaryotic and eukaryotic orthologs for domains with gap weight <67% and eukaryotic orthologs only for domains with >67% gap weight, and eukaryotic orthologs only for the *MLH3/PMS2/PMS1* interaction domain (amino acids 492–680) and the *MSH3/MSH6* interaction domain (amino acids 378–625). Sequence set #7 performed optimally and was used for all further missense variant analysis.

Varying the MAPP-MMR classification threshold from 2.0 to 10.0, we generated a receiver operating characteristic (ROC) curve, and calculated the area under the curve (AUC) to equal 0.945 (Fig. 1). This curve identified an optimal threshold of 4.55 for classifying deleterious and neutral variants (Fig. 1). All MAPP-MMR training and optimization was carried out on a randomly selected subset of deleterious (n = 24 of 49) and all neutral variants (n = 26). The remaining nonoverlapping independent set of deleterious variants (n = 25) was used for validation of the algorithm (Table 2). MAPP-MMR code is available upon request from the authors.

In Silico Analyses

All MAPP-MMR, SIFT [Ng and Henikoff, 2001, 2002, 2003] and PolyPhen [Ramensky et al., 2002; Sunyaev et al., 2001] comparisons used input sequences *MLH1* (gi:4557757), *MSH2* (gi:4557761), and Supplementary Table S1 missense variants. Because they are used in subsequent analyses, Supplementary Table S1 missense variants carried in families from the Creighton Registry and familial CRC subjects were excluded from MAPP-MMR comparisons with SIFT/PolyPhen. Thresholds of <0.05 (SIFT) and >2.0 (PolyPhen) were used to classify deleterious variants and ≥0.05 (SIFT) and ≤2.0 (PolyPhen) to classify neutral variants [Ng and Henikoff, 2001, 2002; Ramensky et al., 2002; Sunyaev et al., 2000a, 2001]. Chi-squared tests compared significant differences in classification accuracy for sensitivity/specificity/negative predictive value (NPV)/positive predictive value (PPV) of missense deleterious and neutral variants. MAPP-MMR scores for all possible *MLH1/MSH2* amino acid substitutions are publicly available in a web-based format (www.healthaffairs.uci.edu/biochem/faculty/lipkin.html).

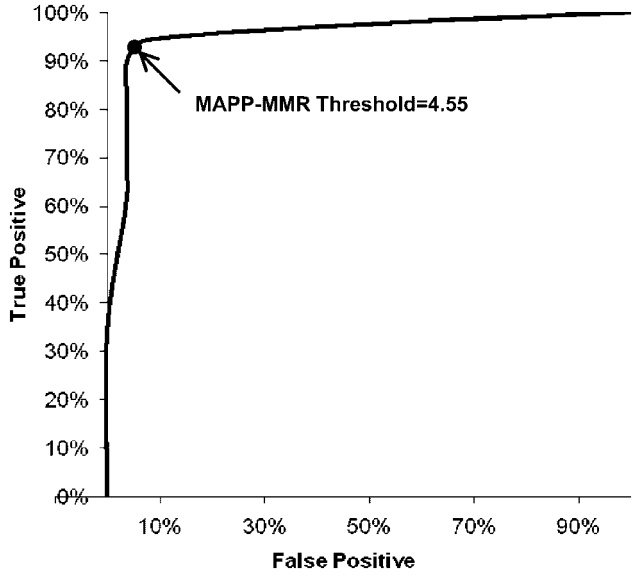

FIGURE 1. **ROC curve for MAPP-MMR accuracy. True-positive and false-positive rates (%) are plotted for a range of MAPP-MMR categorical thresholds from 2.0 to 10.0 to distinguish missense deleterious (n = 49) and neutral (n = 26) variants. AUC = 0.945.**

Creighton Hereditary Cancer Institute Registry Families

A total of 21 families previously evaluated in the Creighton Hereditary Cancer Institute (HCI) Cancer Genetics Clinic with identifiable deleterious *MLH1/MSH2* missense variants were studied. All families meet revised BG and have been previously reported [Lynch, 1999; Peltomaki et al., 2004; Stella et al., 2001; Wagner et al., 2003; Watson et al., 2003, 2004].

MSH2 G322D Case–Control Study

For the relatively common missense variant *MSH2 G322D*, 1,867 CRC case and 1,918 age-, gender-, and ethnically-matched control subjects from the MECC study [Poynter et al., 2005] were genotyped using TaqMan (Applied Biosystems, Foster City, CA) as described [Greenson et al., 2003; Gruber et al., 2002; Shin et al., 2005] (Supplementary Table S3).

Familial CRC Subjects

In the Southern California CRC Cohort of CRC subjects younger than age 55 from Orange County, San Diego County, and Imperial County in Southern California [Anton-Culver et al., 2003; Peel et al., 2000], 142 American CRC probands with familial CRC regardless of MSI status were identified. The clinical and epidemiological features of these subjects are presented in Supplementary Table S4. From this group, germline DNA biospecimens from 91 individuals were available. *MLH1/MSH2* coding exons were sequenced in these 91 individuals for this study.

Control Subject *MLH1/MSH2* Sequencing

A total of 65 American control subjects from the University of California (UC) Irvine Environmental Health Effects Study [Bernstein et al., 2004; Semenza et al., 2001] matched to the first 65 of 91 Southern California Colorectal Cancer Cohort Familial CRC probands were sequenced for this study. Control sequencing stopped after the first 65 subjects because of the highly

TABLE 2. **MAPP-MMR Performance With Increasing Stringency of Deleterious Variant Classification Criteria**

Deleterious variants satisfy	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Number of deleterious variants correctly classified	Number of neutral variants correctly classified
≥1 criteria	94	96	99	83	77/82	25/26
≥2 criteria (All)	94	96	98	89	46/49	25/26
≥2 criteria (training)	92	96	96	93	22/24	25/26
≥2 criteria (validation)	96	96	96	96	24/25	25/26
≥3 criteria	100	96	97	100	29/29	25/26

PPV, positive predictive value; NPV, negative predictive value.

significant difference in MAPP-MMR predicted deleterious variants between familial CRC cases and control groups (P = 0.003, chi-squared test). All *MLH1/MSH2* coding exons were sequenced. MAPP-MMR predicted neutral missense variants were found in case and control sequencing at approximately equal frequencies and are listed in Supplementary Table S4 along with clinical characteristics.

Statistical Methods

Mean MAPP-MMR scores were calculated for all groups: deleterious variants, VUS, neutral variants, Creighton HCI families, and familial CRC Cohort. A two-sided Student's *t*-test and one-way analysis of variance (ANOVA) was used to compare these means. MAPP-MMR performance was compared to that of SIFT and PolyPhen using chi squared analysis. Genotyping for *MSH2 G322D* was carried out for the entire MECC cohort of 3,785 individuals. Odds ratios were calculated for the entire cohort, cases vs. controls, and adjusted for ethnicity and family history. The frequency of MAPP-MMR predicted deleterious and truncating mutations was compared in the American Familial CRC cohort using the chi-squared test.

RESULTS

We compiled a database of all known *MLH1/MSH2* missense variants (n = 301) and used rigorous criteria to classify them into three categories: A) deleterious variants (n = 49); B) VUS (n = 226); or C) neutral variants (n = 26) (Supplementary Table S1; Table 1). We used MAPP-MMR to predict the impact of these missense variants (Fig. 2). Higher MAPP-MMR scores correlate with greater predicted deleterious impact on protein function. Consistent with this observation, the mean MAPP-MMR score for deleterious variants is higher than VUS or neutral variants (16.5 vs. 10.8 vs. 3.5, P < 0.001, one-way ANOVA) (Fig. 3). Similarly, MAPP-MMR scores are inversely correlated with CRC age of onset, which is younger in carriers of deleterious variants (P = 0.0006, two-sided Student's *t*-test with Welch correction).

The MAPP-MMR missense variant scores divide into three primary distributions. A total of 96% of neutral variants cluster in the lowest scoring distribution and 94% of deleterious variants cluster in the higher two distributions, while SIFT/PolyPhen scores for deleterious and neutral variants overlap significantly. The optimal MAPP-MMR threshold to segregate deleterious and neutral variants is 4.55 (Fig. 1). We directly compared the accuracy of MAPP-MMR with SIFT and PolyPhen to classify *MLH1/MSH2* deleterious and neutral variants. Initially, an exploratory set of 24 deleterious variants was used to develop MAPP-MMR. For validation of the MAPP-MMR algorithm, a direct comparison was made between SIFT, PolyPhen, and MAPP-MMR accuracy on a second, nonoverlapping set of 25 deleterious variants. In this independent set, MAPP-MMR outperformed SIFT and PolyPhen with improved sensitivity, specificity, PPV, and NPV. On the entire set of deleterious variants, the accuracy of classification for MAPP-MMR was significantly higher, (sensitivity, 94%; specificity, 96%; PPV, 98%; NPV, 89%), than SIFT or PolyPhen (Table 3). As expected, when

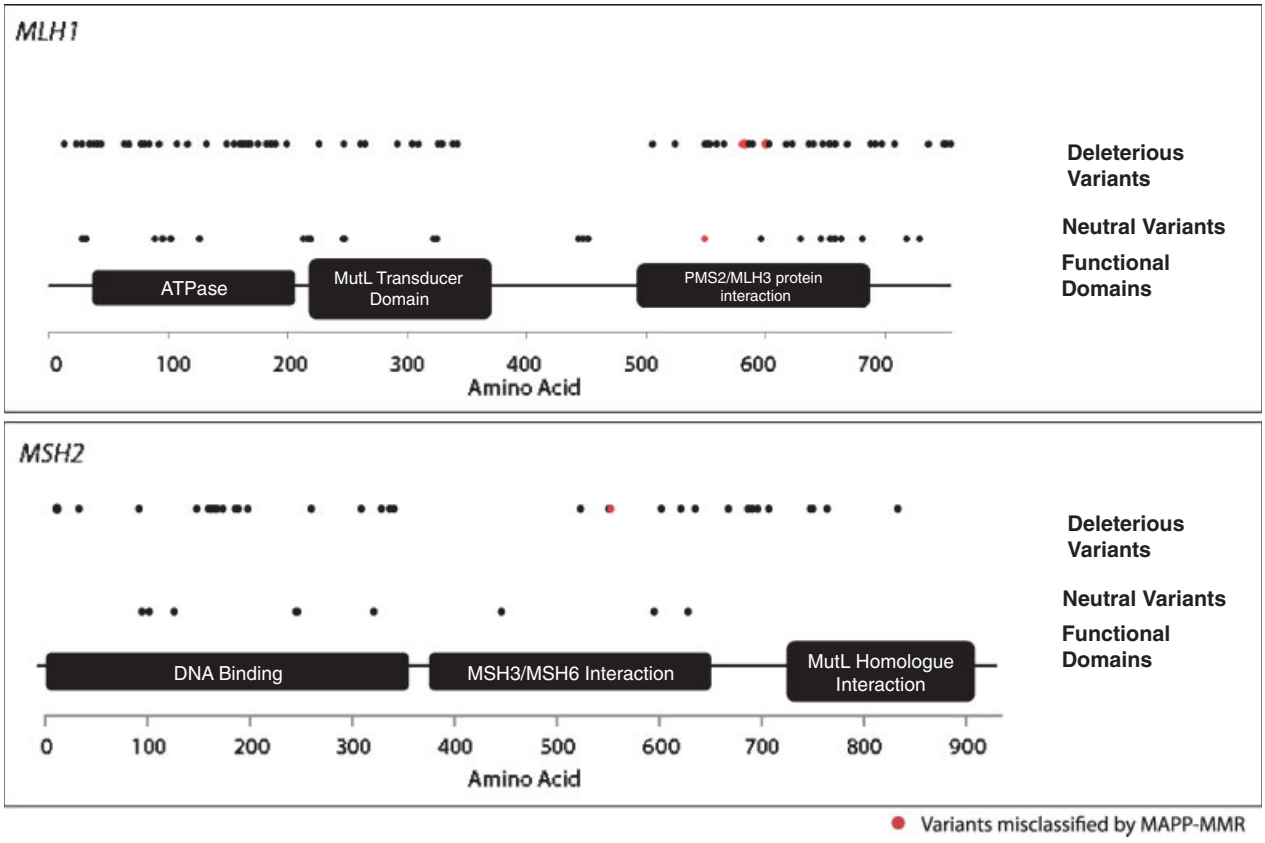

FIGURE 2. **Position of deleterious and neutral variants in *MLH1* and *MSH2*. Deleterious and neutral variants (SupplementaryTable S1) plotted on the *MLH1/MSH2* coding sequences. Each black diamond denotes a unique missense variant that is correctly classified by MAPP-MMR. Red circles denote misclassified missense variants. Functional domains of *MLH1/MSH2* are denoted [Guerrette et al., 1999; Kondo et al., 2001; Lin et al., 2004; Mohd et al., 2006; Park et al., 2006; Schmutte et al., 2001]. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]**

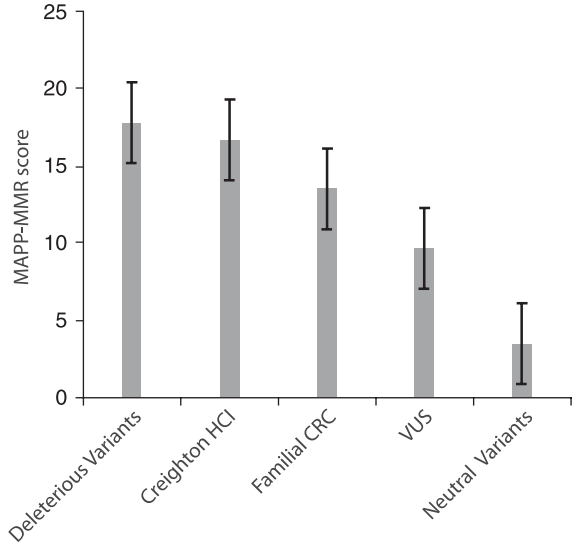

FIGURE 3. **Mean MAPP-MMR scores in different groups of CRC subjects. Mean score with standard error of the mean (SEM) error bars are shown for the following groups: Creighton HCI ascertained families (Table 4; Families 1–19), familial CRC subjects (Table 5), deleterious variants, VUS, and neutral variants (Supplementary Table S1; Table 1) are indicated.**

TABLE 3. **Accuracy of MAPP-MMR Classification of *MLH1/MSH2* Missense Mutations Compared With SIFT and PolyPhen***

	PolyPhen (P value)	SIFT (P value)	MAPP-MMR
Sensitivity	58% (<0.001)	82% (0.035)	94%
Specificity	81% (0.041)	81% (0.041)	96%
Number of deleterious variants correctly classified	29/50	42/50	46/49
Number of neutral variants correctly classified	21/26	21/26	25/26

*P values for comparison with MAPP-MMR using chi-square test are noted in parentheses. The criteria for missense deleterious and neutral variants are defined in Table 1, and the list of all variants is presented in Supplementary Table S1.

more validating criteria are used to classify deleterious variants more rigorously, MAPP-MMR sensitivity increases at the expense of statistical power (Table 2).

The BG [Umar et al., 2004] and bioinformatic models incorporate past medical history (PMH), family medical history (FMH), and MSI/IHC to estimate risk of carrying a deleterious variant, but do not interpret missense variants subsequently identified. We reasoned that if MAPP-MMR is accurate in missense variant interpretation, MAPP-MMR predicted deleterious variants should be found in families at high risk of carrying deleterious variants by these models. After development, and to further test its accuracy, we used MAPP-MMR to classify missense deleterious variants in 21 families with *MLH1/MSH2* missense variants evaluated for Lynch syndrome by Henry Lynch. These were excluded from the model development stage (Table 4). Two families carry both missense deleterious variants and genomic deletions, and are discussed separately. For the 19 remaining families with 15 unique variants we applied the identical criteria as above to classify variants incorporating findings regarding the presence/absence of missense variants in other family members and MSI/IHC. A total of 17 families had missense variants classified as deleterious and two families as VUS (Table 4). Using the identical approach and parameters employed to analyze the missense variant database with these 19 families, MAPP-MMR had 100% accuracy in correct classification (17 deleterious variants and two VUS) (Table 4).

To prospectively test MAPP-MMR classification accuracy we analyzed *MSH2 G322D*. This missense variant has been interpreted by different groups alternately as deleterious [Liu et al., 1996; Maliaka et al., 1996] or benign [Froggatt et al., 1996; Herfarth et al., 1997] (In InSiGHT *MSH2 G322D* is listed in both pathogenic and nonpathogenic categories). *MSH2 G322D*'s score is 3.24 (predicted neutral variant). To rigorously test MAPP-MMR's classification, we quantified *MSH2 G322D* frequency in 3,785 population-based CRC cases and matched controls (Supplementary Table S3). *MSH2 G322D* is not associated with increased CRC risk (odds ratio [OR], 0.90; 95% confidence interval [CI], 0.66–1.21). Adjusting for first degree family history of CRC or ethnicity did not appreciably alter risk ($OR_{G322D+\text{family history}}$, 0.92, [95% CI, 0.68–1.24]; $OR_{G322D+\text{ethnicity}}$, 0.88 [95% CI, 0.66–1.19]).

Next, we applied MAPP-MMR to quantify more precisely the contribution of *MLH1/MSH2* missense deleterious variants and truncating mutations to familial CRC (defined as proband and one first degree relative or two second degree relatives with CRC). A total of 15 to 20% of all CRC is familial [Burt and Peterson, 1996; Burt, 1997], and many of these probands do not meet BG [Barnetson et al., 2006; Hampel et al., 2005a, 2005b; Pinol et al., 2005]. To prospectively test MAPP-MMR's predictive ability, we sequenced germline *MLH1/MSH2* from 91 familial CRC subjects from Southern California [Peel et al., 2000]. In addition to 34 predicted neutral variants and one truncating mutation, MAPP-MMR predicted 10 deleterious missense variants in familial CRC subjects (12.1%) (Table 5). Two (*MSH2 T552P* and *MLH1 T117M*) were previously identified as deleterious variants (Supplementary Table S1). Tumor location is known for 8 out of 10 subjects. CRCs are right-sided in 7 out of 8 (87%). MSI status was performed on six subjects, and 4 out of 6 are MSI-L or MSI-H (66%). For three familial CRC probands, a DNA sample from a first degree relative with CRC was available, and in 3 out of 3 cases the predicted missense variant cosegregated. Therefore, MAPP-MMR identified patients enriched in Lynch syndrome features, and the predicted deleterious variants cosegregated in affected siblings. For comparison, we sequenced 65 controls matched to the above American Familial CRC cases and identified no truncating or MAPP-MMR predicted *MLH1/MSH2* deleterious variants ($P = 0.003$, chi-squared).

As expected, MAPP-MMR predicted neutral variants were found in sequencing of both familial CRC cases and controls (Supplementary Table S4A). *MLH1* I219V was found at similar frequencies in case and controls ($P > 0.05$, chi-squared) and the only predicted neutral variant with allele frequency high enough for analysis for molecular epidemiology. The remaining predicted neutral variants were evaluated in aggregate and had similar frequencies in cases and controls ($P > 0.05$, chi-squared). These patients had clinical features similar to those seen in sporadic CRC (Supplementary Table S4B). The mean MAPP-MMR score is 13.5 for familial CRC subjects with predicted deleterious variants. Consistent with the clinical phenotype, the familial CRC mean MAPP-MMR score is less than that of Lynch syndrome missense deleterious variants (16.5) and higher than neutral variants (3.5, $P = 0.038$, two-sided Student's *t*-test) (Fig. 3).

TABLE 4. Families From the Creighton University Hereditary Cancer Institute With *MLH1/MSH2* Deleterious and Neutral Variants*

Family ID	Gene	Variant	Age of onset (years)	Gender	BG	MAPP-MMR	MAPP-MMR prediction	Classification
14	MSH2	Ile154Met	36	F	+	2.63	Neutral	VUS
1	MLH1	Ala29Ser	51	M	+	3.59	Neutral	VUS
11	MLH1	Val506Ala	28	M	+	5.56	Deleterious	Deleterious
12	MLH1	Val506Ala	26	F	+	5.56	Deleterious	Deleterious
13	MSH2	Gly149Asp	37	F	+	5.76	Deleterious	Deleterious
9	MLH1	Arg226Leu	50	F	+	6.09	Deleterious	Deleterious
16	MSH2	Thr552Pro	40	M	+	8.71	Deleterious	Deleterious
4	MLH1	Glu102Lys	49	F	+	12.9	Deleterious	VUS
8	MLH1	Val185Gly	40	F	+	14.54	Deleterious	Deleterious
10	MLH1	Gly244Asp	52	M	+	15.65	Deleterious	VUS
19	MSH2	Ser577Leu	46	M	+	15.65	Deleterious	Deleterious
17	MSH2	Ala636Pro	52	M	+	17.23	Deleterious	Deleterious
5	MLH1	Thr117Met	41	F	+	21.05	Deleterious	Deleterious
6	MLH1	Thr117Met	45	F	+	21.05	Deleterious	Deleterious
7	MLH1	Thr117Met	47	M	+	21.05	Deleterious	Deleterious
18	MSH2	Asn671Lys	50	F	+	25.31	Deleterious	VUS
2	MLH1	Gly67Arg	33	M	+	36.52	Deleterious	Deleterious
3	MLH1	Gly67Arg	40	F	+	36.52	Deleterious	Deleterious
15	MSH2	Arg338Gly	45	F	+	36.52	Deleterious	Deleterious
20	MSH2	Asn583Ser	35	M	+	2.42	Neutral	VUS
21	MLH1	Glu578Gly	39	F	+	4.85	Borderline	Deleterious

*The gene and missense variant identified in the family, the proband age of first CRC and gender are given. Deleterious and neutral variant classification using criteria specified in Table 1. Families below the horizontal line have both genomic deletions and missense mutations (see Results and Discussion sections).
+, family satisfies revised Bethesda Guidelines; MAPP-MMR, MAPP-MMR score for missense mutations in column 7; VUS, variant of uncertain significance.

TABLE 5. Familial CRC Probands With MAPP-MMR Predicted Deleterious Variants*

Gene	Variant	Age	MSI status	MAPP-MMR	Location	Histology	Stage	Differentiation
MLH1	Tyr 293 Asp	49	MSI-H	25.3	Colon transverse	Adenocarcinoma	Stage II	Moderate
	Glu 268 Gly	41	MSS	6.02	Colon transverse	Adenocarcinoma	Stage I	Moderate
	Tyr 280 Asp	49	NA	27.48	Cecum	Adenocarcinoma	Stage II	Poor
	Arg 725 Cys	47	MSS	6.39	Colon transverse	Adenocarcinoma	Stage II	Moderate
	Thr 117 Met	43	NA	21.05	NA			
MSH2	Thr 441 Pro	56	MSI-L	5.9	Cecum	Adenocarcinoma	Stage II	Poor
	Thr 552 Pro	40	MSI-H	8.71	Hepatic flexure	Adenocarcinoma	Stage III	Moderate
	Thr 552 Pro	40	NA	8.71	Colon transverse	Adenocarcinoma	Stage II	Moderate
	Thr 552 Pro	41	MSI-L	8.71	Colon descending	Adenocarcinoma	Stage II	Well
	Thr 552 Pro	42	MA	8.71	NA			

*Proband age at CRC onset, or Southern California Colorectal Cancer Cohort ascertainment, MSI status, and MAPP-MMR scores are indicated. CRC location in colorectum, histology, stage, and tumor type are also indicated.
MSS, microsatellite stable; MSI-L, microsatellite low status; MSI-H, microsatellite high status; NA, not available.

DISCUSSION

More than 300 unique *MLH1/MSH2* missense variants have been reported. There is consensus among cancer geneticists that interpretation of missense variants is challenging. MSI analysis is important, with an estimated 80% sensitivity and ~90% specificity for detecting *MLH1/MSH2* mutations [Yan et al., 2000]. However, MSI testing is expensive and tumors sometimes cannot be obtained. IHC is useful, but deleterious missense variants can cause nonfunctional proteins that are still expressed and scored as wild type in this assay [Liu et al., 1999; Wahlberg et al., 2002]. Molecular epidemiology is helpful, but most missense variants are extremely rare. Biochemical analysis of missense variants is important, but the cost is prohibitive for clinical application. A recent study compared head-to-head current computational algorithms designed to interpret missense variants in several proteins that are associated with human genetic disease [Chan et al., 2007]. This comparison included 28 *MLH1* and 21 *MSH2* missense variants and confirmed a high predictive value for algorithms that use evolutionary sequence conservation (with or without incorporating protein structural change) to predict the clinical consequences of missense variants [Chan et al., 2007]. A newer algorithm, MAPP, utilizes a similar approach and is more accurate [Stone and Sidow, 2005]. Importantly, MAPP can be customized for specific proteins. We have therefore optimized MAPP specifically for *MLH1/MSH2* and developed MAPP-MMR for missense variant interpretation in these clinically important genes.

To develop and validate our variant prediction algorithm we constructed a database of all known *MLH1/MSH2* variants and all publicly available supporting evidence, in numerous cases supplemented by additional unpublished data. As comprehensive evidence is not systematically reported in the literature for all variants, 75% of variants were classified as VUS. This point highlights the importance of reporting as much clinical, epidemiologic, and biochemical data available as possible on new variants.

In general, variants with more independent lines of evidence supporting deleterious effects are more likely to be deleterious. While the same trend is seen whether MAPP-MMR uses 1, 2, or 3 criteria to classify variants as deleterious or neutral, to decrease false positives MAPP-MMR was trained and validated on deleterious variants meeting ≥ 2 criteria (Table 1). For variants meeting ≥ 2 criteria, 94% of deleterious variants are correctly classified and 6% are incorrectly classified as neutral variants (Table 1). For variants meeting ≥ 3 criteria, 100% of deleterious

variants are correctly classified. However, the number of VUS increases. Therefore, clinicians must use their individual judgment as to how they want to combine available evidence with MAPP-MMR scores in their interpretation of novel *MLH1/MSH2* missense variants seen in the clinic in the future. It is also important to note that higher MAPP-MMR scores correlate with greater deleterious impact, a parameter that can also be incorporated into clinical interpretation of novel variants. The overall accuracy of MAPP-MMR is 94.6%. However, for variants with borderline scores of 3.0 to 5.0 the overall accuracy is only 88% (15 of 17 variants). Therefore we have made added a provisional cautionary warning to these variants as "borderline."

Missense variants can impair proteins not only by affecting structure–function, but also at the DNA level because of mRNA splicing. Based on evolutionary conservation, MAPP-MMR identified several missense variants as deleterious mutations whose mechanism is impaired splicing. Splicing is an important mechanism because biochemical approaches expressing *MLH1/MSH2* variant cDNAs will misclassify deleterious splicing mutations as neutral variants. Similarly, because some biochemical assays express *MLH1/MSH2* in yeast and insect cells, posttranslational modification processing signals conserved in mammals may similarly be misclassified using only the biochemical approach.

For VUS, MAPP-MMR generates hypotheses as to which are neutral and which are deleterious (Supplementary Table S5). *MSH2 G322D* is common enough that we were able to test its MAPP-MMR classification using molecular epidemiology. Unfortunately, most of the remaining VUS are rare and not amenable to this approach.

The development of risk prediction models based on PMH and FMH to predict the probability of detecting Lynch syndrome mutations is exciting [Balmana et al., 2006; Barnetson et al., 2006; Chen et al., 2006; Wijnen et al., 1998]. However, these algorithms can cause missense variant interpretation biases. In the example of *MLH1 T117M* (Table 4), its high MAPP-MMR score (21.05) provided useful confirmatory evidence that it is truly a deleterious variant. While additional supporting evidence from the literature exists for *MLH1 T117M*, for many missense mutations supporting evidence is inadequate (Supplementary Table S5) and MAPP-MMR scores may be the only information available. Finally, *MLH1/MSH2* variants identified in high-risk probands may cause their overinterpretation as deleterious. For example, variants inconsistent with highly penetrant Lynch syndrome mutations (*MSH2 N583S* and *MLH1 E578G*) and *MSH2* genomic deletions were identified in two families (Table 4). Both families were correctly classified by BG as high-risk for a deleterious variant. However, these algorithms cannot distinguish whether the variant or genomic deletion is the causative deleterious mutation. MAPP-MMR predicted *MSH2 N583S* as benign (score 2.42), and *MLH1 E578G* (score 4.85) as a borderline deleterious variant. In both families the MAPP-MMR scores are inconsistent with the probands' CRC onset at 35 and 39 years of age and affected first degree relatives and predicted the existence of a cryptic mutation not identified by *MLH1/MSH2* sequencing, such as a genomic deletion. Because of cost, many laboratories will serially sequence *MLH1/MSH2* first, and perform MLPA/Southern analysis only if no mutation is identified. Therefore, it is anticipated that in some patients VUS will be interpreted as mutations and additional analyses foregone unless the missense mutation's predicted impact is assessed. MAPP-MMR may be particularly useful in these situations. Recently, similar approaches to combine mutation risk prediction and missense variant classification have recently been developed for *BRCA1/BRCA2* [Chenevix-Trench et al., 2006; Goldgar et al., 2004; Phelan et al., 2005; Wu et al., 2005]. As with these *BRCA1/BRCA2* missense mutation classification algorithms, future studies will be required to validate MAPP-MMR with additional datasets, and to define the optimal combination of mutation risk prediction models and MAPP-MMR, for the highest accuracy of missense variant classification.

ACKNOWLEDGMENTS

We thank Susan Neuhausen, Ph.D., and Chad Garner, Ph.D., of the University of California, Irvine, for critical reading of this manuscript.

REFERENCES

Anton-Culver H, Ziogas A, Bowen D, Finkelstein D, Griffin C, Hanson J, Isaacs C, Kasten-Sportes C, Mineau G, Nadkarni P, Rimer B, Schildkraut J, Strong L, Weber B, Winn D, Hiatt R, Nayfield S. 2003. The Cancer Genetics Network: recruitment results and pilot studies. Community Genet 6:171–177.

Balmana J, Stockwell DH, Steyerberg EW, Stoffel EM, Deffenbaugh AM, Reid JE, Ward B, Scholl T, Hendrickson B, Tazelaar J, Burbidge LA, Syngal S. 2006. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. JAMA 296:1469–1478.

Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. 2006. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 354:2751–2763.

Bernstein JL, Langholz B, Haile RW, Bernstein L, Thomas DC, Stovall M, Malone KE, Lynch CF, Olsen JH, Anton-Culver H, Shore RE, Boice JD, Jr., Berkowitz GS, Gatti RA, Teitelbaum SL, Smith SA, Rosenstein BS, Borresen-Dale AL, Concannon P, Thompson WD. 2004. Study design: evaluating gene-environment interactions in the etiology of breast cancer - the WECARE study. Breast Cancer Res 6:R199–R214.

Blasi MF, Ventura I, Aquilina G, Degan P, Bertario L, Bassi C, Radice P, Bignami M. 2006. A human cell-based assay to evaluate the effects of alterations in the MLH1 mismatch repair gene. Cancer Res 66:9036–9044.

Burt R, Peterson GM. 1996. In: Young GP, Rozen P, Levin B. Prevention and early detection of colorectal cancer. Philadelphia: W.B. Saunders.

Burt RW. 1997. Screening of patients with a positive family history of colorectal cancer. Gastrointest Endosc Clin N Am 7:65–79.

Chan PA, Duraisamy S, Miller PJ, Newell JA, McBride C, Bond JP, Raevaara T, Ollila S, Nystrom M, Grimm AJ, Christodoulou J, Oetting WS, Greenblatt MS. 2007. Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). Hum Mutat 28:683–693.

Chasman D, Adams RM. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 307:683–706.

Chen S, Wenyi Wang, Shing Lee, Khedoudja Nafa, Johanna Lee, Kathy Romans, Patrice Watson, Stephen B. Gruber, David Euhus, Kenneth W. Kinzler, John Hopper, Steven Gallinger, Noralane Lindor, Graham Casey, Nathan Ellis, Francis M. Giardiello, Colon Cancer Family Registry, Kenneth Offit, Giovanni Parmigiani. 2006. Prediction of germline mutations and cancer risk in hereditary nonpolyposis colorectal cancer. JAMA 296:1479–1487.

Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T, Scholl T, Bekessy A, Marsh A, Lovelock P, Wong M, Tesoriero A, Renard H, Southey M, Hopper JL, Yannoukakos K, Brown M, Easton D, Tavtigian SV, Goldgar D, Spurdle AB. 2006. Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. Cancer Res 66:2019–2027.

Froggatt NJ, Joyce JA, Evans DG, Lunt PW, Koch DJ, Ponder BJ, Maher ER. 1996. MSH2 sequence variations and inherited colorectal cancer susceptibility. Eur J Cancer 32A:178.

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. 2004. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 75:535–544.

Greenson JK, Bonner JD, Ben-Yzhak O, Cohen HI, Miselevich I, Resnick MB, Trougouboff P, Tomsho LD, Kim E, Low M, Almog R, Rennert G, Gruber SB. 2003. Phenotype of microsatellite unstable colorectal carcinomas: well-differentiated and focally mucinous tumors and the absence of dirty necrosis correlate with microsatellite instability. Am J Surg Pathol 27:563–570.

Gruber SB, Ellis NA, Scott KK, Almog R, Kolachana P, Bonner JD, Kirchhoff T, Tomsho LP, Nafa K, Pierce H, Low M, Satagopan J, Rennert H, Huang H, Greenson JK, Groden J, Rapaport B, Shia J, Johnson S, Gregersen PK, Harris CC, Boyd J, Rennert G, Offit K. 2002. BLM heterozygosity and the risk of colorectal cancer. Science 297:2013.

Guerrette S, Acharya S, Fishel R. 1999. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. J Biol Chem 274:6336–6341.

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J, Panescu J, Fix D, Lockman J, Comeras I, de la Chapelle A. 2005a. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 352:1851–1860.

Hampel H, Stephens JA, Pukkala E, Sankila R, Aaltonen LA, Mecklin JP, de la Chapelle A. 2005b. Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. Gastroenterology 129:415–421.

Herfarth KK, Kodner IJ, Whelan AJ, Ivanovich JL, Bracamontes JR, Wells Jr SA, Goodfellow PJ. 1997. Mutations in MLH1 are more frequent than in MSH2 in sporadic colorectal cancers with microsatellite instability. Genes Chromosomes Cancer 18:42–49.

Kondo E, Horii A, Fukushige S. 2001. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. Nucleic Acids Res 29:1695–1702.

Lin DP, Wang Y, Scherer SJ, Clark AB, Yang K, Avdievich E, Jin B, Werling U, Parris T, Kurihara N, Umar A, Kucherlapati R, Lipkin M, Kunkel TA, Edelmann W. 2004. An Msh2 point mutation uncouples DNA mismatch repair and apoptosis. Cancer Res 64:517–522.

Lipkin SM, Rozek LS, Rennert G, Yang W, Chen PC, Hacia J, Hunt N, Shin B, Fodor S, Kokoris M, Greenson JK, Fearon E, Lynch H, Collins F, Gruber SB. 2004. The MLH1 D132H variant is associated with susceptibility to sporadic colorectal cancer. Nat Genet 36:694–699.

Liu B, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, Jass JR, Dunlop M, Wyllie A, Peltomaki P, de la Chapelle A, Hamilton SR, Vogelstein B, Kinzler KW. 1996. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. Nat Med 2:169–174.

Liu T, Tannergard P, Hackman P, Rubio C, Kressner U, Lindmark G, Hellgren D, Lambert B, Lindblom A. 1999. Missense mutations in hMLH1 associated with colorectal cancer. Hum Genet 105:437–441.

Lynch HT. 1999. Hereditary nonpolyposis colorectal cancer (HNPCC). Cytogenet Cell Genet 86:130–135.

Lynch HT, de la Chapelle A. 2003. Hereditary colorectal cancer. N Engl J Med 348:919–932.

Maliaka YK, Chudina AP, Belev NF, Alday P, Bochkov NP, Buerstedde JM. 1996. CpG dinucleotides in the hMSH2 and hMLH1 genes are hotspots for HNPCC mutations. Hum Genet 97:251–255.

Mohd AB, Palama B, Nelson SE, Tomer G, Nguyen M, Huo X, Buermeyer AB. 2006. Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. DNA Repair (Amst) 5:347–361.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Res 11:863–874.

Ng PC, Henikoff S. 2002. Accounting for human polymorphisms predicted to affect protein function. Genome Res 12:436–446.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814.

Park JG, Kim DW, Hong CW, Nam BH, Shin YK, Hong SH, Kim IJ, Lim SB, Aronson M, Bisgaard ML, Brown GJ, Burn J, Chow E, Conrad P, Douglas F, Dunlop M, Ford J, Greenblatt MS, Heikki J, Heinimann K, Lynch EL, Macrae F, McKinnon WC, Moeslein G, Rossi BM, Rozen P, Schofield L, Vaccaro C, Vasen H, Velthuizen M, Viel A, Wijnen J. 2006. Germ line mutations of mismatch repair genes in hereditary non-polyposis colorectal cancer patients with small bowel cancer: International Society for Gastrointestinal Hereditary Tumours Collaborative Study. Clin Cancer Res 12:3389–3393.

Peel DJ, Ziogas A, Fox EA, Gildea M, Laham B, Clements E, Kolodner RD, Anton-Culver H. 2000. Characterization of hereditary nonpolyposis colorectal cancer families from a population-based series of cases. J Natl Cancer Inst 92:1517–1522.

Peltomaki P, Vasen, H. 2004. Mutations associated with HNPCC predisposition—update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 20:269–276.

Phelan CM, Dapic V, Tice B, Favis R, Kwan E, Barany F, Manoukian S, Radice P, van der Luijt RB, van Nesselrooij BP, Chenevix-Trench G, kConFab, Caldes T, de la Hoya M, Lindquist S, Tavtigian SV, Goldgar D, Borg A, Narod SA, Monteiro AN. 2005. Classification of BRCA1 missense variants of unknown clinical significance. J Med Genet 42:138–146.

Pinol V, Castells A, Andreu M, Castellvi-Bel S, Alenda C, Llor X, Xicola RM, Rodriguez-Moranta F, Paya A, Jover R, Bessa X. 2005. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 293:1986–1994.

Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS, Low M, Greenson JK, Rennert G. 2005. Statins and the risk of colorectal cancer. N Engl J Med 352:2184–2192.

Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lonnqvist KE, Holinski-Feder E, Sutter C, McKinnon W, Duraisamy S, Gerdes AM, Peltomaki P, Kohonen-Ccorish M, Mangold E, Macrae F, Greenblatt M, de la Chapelle A, Nystrom M. 2005. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology 129:537–549.

Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900.

Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, Srivastava S. 1997. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 89:1758–1762.

Schmutte C, Sadoff MM, Shim KS, Acharya S, Fishel R. 2001. The interaction of DNA mismatch repair proteins with human exonuclease I. J Biol Chem 276:33011–33018.

Semenza JC, Ziogas A, Largent J, Peel D, Anton-Culver H. 2001. Gene-environment interactions in renal cell carcinoma. Am J Epidemiol 153:851–859.

Shin BY, Chen H, Rozek LS, Paxton L, Peel DJ, Anton-Culver H, Rennert G, Mutch DG, Goodfellow PJ, Gruber SB, and others. 2005. Low allele frequency of MLH1 D132H in American colorectal and endometrial cancer patients. Dis Colon Rectum 48:1723–1727.

Stella A, Wagner A, Shito K, Lipkin SM, Watson P, Guanti G, Lynch HT, Fodde R, Liu B. 2001. A nonsense mutation in MLH1 causes exon skipping in three unrelated HNPCC families. Cancer Res 61:7020–7024.

Stone EA, Sidow A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res 15:978–986.

Sunyaev S, Ramensky V, Bork P. 2000a. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet 16:198–200.

Sunyaev SR, Lathe 3rd WC, Ramensky VE, Bork P. 2000b. SNP frequencies in human genes an excess of rare alleles and differing modes of selection. Trends Genet 16:335–337.

Sunyaev S, Ramensky V, Koch I, Lathe 3rd W, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. Hum Mol Genet 10:591–597.

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 96:261–268.

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

Wahlberg SS, Schmeits J, Thomas G, Loda M, Garber J, Syngal S, Kolodner RD, Fox E. 2002. Evaluation of microsatellite instability and immunohistochemistry for the prediction of germ-line MSH2 and MLH1 mutations in hereditary nonpolyposis colon cancer families. Cancer Res 62:3485–3492.

Watson P, Narod SA, Fodde R, Wagner A, Lynch JF, Tinley ST, Snyder CL, Coronel SA, Riley B, Kinarsky Y, Lynch HT. 2003. Carrier risk status changes resulting from mutation testing in hereditary non-polyposis colorectal cancer and hereditary breast-ovarian cancer. J Med Genet 40:591–596.

Watson P, Ashwathnarayan R, Lynch HT, Roy HK. 2004. Tobacco use and increased colorectal cancer risk in patients with hereditary nonpolyposis colorectal cancer (Lynch syndrome). Arch Intern Med 164:2429–2431.

Wijnen JT, Vasen HF, Khan PM, Zwinderman AH, van der Klift H, Mulder A, Tops C, Moller P, Fodde R. 1998. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med 339:511–518.

Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian SV, Deffenbaugh A, Goldgar D, Couch FJ. 2005. Functional evaluation and cancer risk assessment of BRCA2 unclassified variants. Cancer Res 65:417–426.

Xi T, Jones IM, Mohrenweiser HW. 2004. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics 83:970–979.

Yan H, Papadopoulos N, Marra G, Perrera C, Jiricny J, Boland CR, Lynch HT, Chadwick RB, de la Chapelle A, Berg K, Eshleman JR, Yuan W, Markowitz S, Laken SJ, Lengauer C, Kinzler KW, Vogelstein B. 2000. Conversion of diploidy to haploidy. Nature 403:723–724.