

# The *MLH1* D132H variant is associated with susceptibility to sporadic colorectal cancer

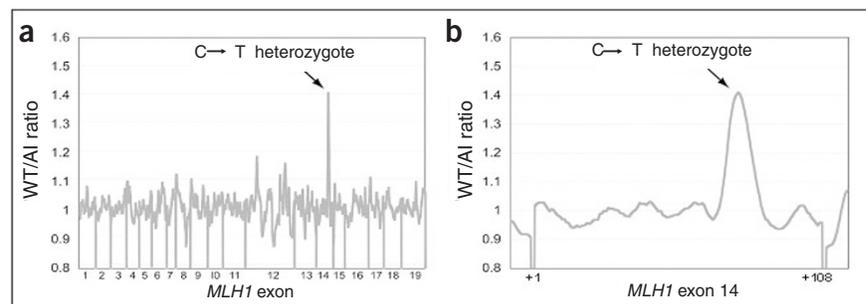
Steven M Lipkin<sup>1</sup>, Laura S Rozek<sup>2,3</sup>, Gad Rennert<sup>4</sup>, Wei Yang<sup>5</sup>, Peng-Chieh Chen<sup>1</sup>, Joseph Hacia<sup>6</sup>, Nathan Hunt<sup>6</sup>, Brian Shin<sup>1</sup>, Steve Fodor<sup>7</sup>, Mark Kokoris<sup>8</sup>, Joel K Greenson<sup>9</sup>, Eric Fearon<sup>2,9,10</sup>, Henry Lynch<sup>11</sup>, Francis Collins<sup>12</sup> & Stephen B Gruber<sup>2</sup>

Most susceptibility to colorectal cancer (CRC) is not accounted for by known risk factors. Because *MLH1*, *MSH2* and *MSH6* mutations underlie high-penetrance CRC susceptibility in hereditary nonpolyposis colon cancer (HNPCC), we hypothesized that attenuated alleles might also underlie susceptibility to sporadic CRC. We looked for gene variants associated with HNPCC in Israeli probands with familial CRC unstratified with respect to the microsatellite instability (MSI) phenotype. Association studies identified a new *MLH1* variant (415G→C, resulting in the amino acid substitution D132H) in ~1.3% of Israeli individuals with CRC self-described as Jewish, Christian and Muslim. *MLH1* 415C confers clinically significant susceptibility to CRC. In contrast to classic HNPCC, CRCs associated with *MLH1* 415C usually do not have the MSI defect,

which is important for clinical mutation screening. Structural and functional analyses showed that the normal ATPase function of *MLH1* is attenuated, but not eliminated, by the *MLH1* 415G→C mutation. The new *MLH1* variant confers a high risk of CRC and identifies a previously unrecognized mechanism in microsatellite-stable tumors. These studies suggest that variants of mismatch repair proteins with attenuated function may account for a higher proportion of susceptibility to sporadic microsatellite-stable CRC than previously assumed.

To identify variant alleles in *MLH1*, *MSH2* or *MSH6* with high sensitivity and specificity, we designed a new high-density oligonucleotide array (HNPCC Chip) that uses improved bioinformatic

**Figure 1** Variant allele detection by the HNPCC Chip resequencing array. (a) Graph of HNPCC Chip. Loss of hybridization analysis for full *MLH1* coding sequence and splice junctions. Data are the average of signals obtained for one sense and one antisense chip hybridization with the same individual's genomic DNA. Nucleotide substitutions in individuals with CRC are graphically shown as a signal peak above baseline (ratio of signal for wild-type DNA to DNA from affected individual; WT/AI). Downward spikes below graph were inserted *post hoc* to facilitate identification of exon boundaries. This DNA was heterozygous with respect to a C→T base substitution in *MLH1* exon 14. The boundaries of all 19 *MLH1* coding exons are labeled on the ordinate axis. (b) Higher magnification of HNPCC Chip signal analysis for *MLH1* exon 14 (108 bp). WT/AI, normalized ratio of chip hybridization signal for fluorescein-labeled wild-type reference DNA divided by the hybridization signal for streptavidin-biotin-phycoerythrin-labeled DNA from an affected individual. The exact position of the C→T substitution is indicated by the arrow in both panels.



<sup>1</sup>Divisions of Oncology and Epidemiology, Departments of Medicine and Biological Chemistry, University of California, Irvine, Irvine, California 92697, USA. Departments of <sup>2</sup>Internal Medicine and <sup>3</sup>Epidemiology, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>4</sup>Department of Community Medicine and Epidemiology Carmel Medical Center and Technion Faculty of Medicine Haifa, Israel. <sup>5</sup>Molecular Structure Section, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, USA. <sup>6</sup>The Institute for Genetic Medicine, University of Southern California, Los Angeles, California 90089, USA. <sup>7</sup>Affymetrix Corporation, Santa Clara, California 95051, USA. <sup>8</sup>BioCaptus, Bothell, Washington 99164, USA. Departments of <sup>9</sup>Pathology and <sup>10</sup>Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>11</sup>Hereditary Cancer Institute, Creighton University, Omaha, Nebraska 50035, USA. <sup>12</sup>Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland 20892, USA. Correspondence should be addressed to S.M.L. (slipkin@uci.edu) or S.B.G. (sgruber@med.umich.edu).

variant-detection algorithms. The HNPCC Chip covers ~14 kb of *MLH1*-*MSH2*-*MSH6* coding and flanking sequence. The HNPCC Chip effectively detected variants in these three mismatch repair (MMR) genes (Fig. 1), accurately detecting 60 of 60 unique missense or base substitution mutations (data not shown).

We used the HNPCC Chip to look for variants in 35 unrelated Israeli individuals with CRC who had one first-degree relative with CRC. We identified 36 unique variants (Supplementary Table 1 online) but only 1 classic MMR mutation. Of these 36 variants, 20 (55%) are represented in the SNPdb, Environmental Genome Project, Celera, SNP500 or HGBase SNP databases. Another six (17%) MMR variants are in mutation databases (ICG-HNPCC, Cardiff Mutation Database or Myriad Genetics Database and Medline), leaving ten (28%) new MMR variants in these intensively resequenced genes.

We tested association of MMR variant alleles with CRC in a set of 455 cases and 455 matched controls from the Molecular Epidemiology of Colorectal Cancer (MECC) study (Supplementary Table 2 online). Three variants showed allelic association with CRC. *MLH1* 415C (odds ratio (OR) = 5.0, 95% confidence interval (c.i.) = 0.6–42.8) and *MSH6* 1613C (OR = 1.87, 95% c.i. = 0.90–3.8) showed trends towards association with CRC (Table 1). In contrast, *MSH6* 2633C was over-represented in controls (OR = 0.35, 95% c.i. = 0.14–0.90). To minimize false positive associations, we validated these associations in a set of 844 individuals with CRC and 940 controls that did not overlap with the first set (Table 1). *MSH6* 1613C was equally frequent in cases and controls (OR = 1.0, 95% c.i. = 0.6–2.0; combined data set OR = 1.4, 95% c.i. = 0.9–2.2). For *MSH6* 2633C, the validation set showed no CRC risk (OR = 0.80, 95% c.i. = 0.4–1.4; combined data set OR = 0.60, 95% c.i. = 0.4–0.99). *MSH6* 2633T→C was originally identified in probands with familial and metachronous CRC<sup>1,2</sup>. It is annotated as a pathogenic mutation in the Celera, HGMD and ICG-HNPCC databases but also occurs in Ashkenazi control subjects<sup>3</sup>. We cannot exclude the possibility that this variant has a subtle effect, but our large

data set suggests that it is probably not associated with CRC risk. Similar results have been observed with putative rare *ATM* polymorphisms that, when tested in larger sets, are not associated with cancer susceptibility<sup>4</sup>.

In the validation set, the *MLH1* 415C allele was more frequent in CRC cases (OR = 4.6, 95% c.i. = 1.5–13.9; combined data set OR = 4.6, 95% c.i. = 1.7–12.3). These results suggest that *MLH1* accounts for ~1.3% of CRC in Israel. We identified *MLH1* 415C in Ashkenazi Jews, Muslim and Christian Arabs, Sephardic Jews, Druze Christians and Bedouins (self-described ethnicities; Fig. 2 and Table 2). Analysis of flanking and intragenic microsatellite markers identified no conserved *MLH1* 415C haplotype and suggested that Jewish carriers of the *MLH1* 415C allele are unrelated. The average age of CRC onset in carriers of *MLH1* 415C is 70.1 years, compared with 20–40 years in individuals with HNPCC. Only 6 of 21 carriers of *MLH1* 415C had a first-degree relative affected with HNPCC-spectrum tumors (Fig. 2). Multiple primary tumors are also rare. Only one proband had a second primary malignancy (papillary thyroid cancer), and none had metachronous CRC (Fig. 2 and Supplementary Table 3 online). Classic inactivating mutations in *MLH1* that underlie HNPCC cause MSI. Notably, in the 11 CRCs from carriers of *MLH1* 415C available for MSI testing, 9 (82%) did not show MSI (Supplementary Table 4 online). Analogous to the attenuated *APC* allele encoding the amino acid substitution I1307K<sup>5</sup> and the high-penetrance syndrome familial adenomatous polyposis, the clinical characteristics of the 21 carriers of the *MLH1* 415C allele with CRC in this study more closely resemble those of individuals with sporadic CRC than those of individuals with typical HNPCC (Fig. 2 and Table 2).

Mutations in MMR genes are thought to cause CRC by increasing rates of single-base substitutions and frameshift mutations (MSI) or preventing the initiation of apoptosis<sup>6</sup>. Structure-function and comparative sequence analyses defined essential, evolutionarily conserved MMR domains. In *Escherichia coli*, a MutS dimer binds mismatched DNA and then forms an ATP-binding-dependent ternary complex with dimeric MutL<sup>7</sup>. In mammals, MutS homologs (MSH proteins) bind to mismatched DNA. MutL homologs (MLH and PMS proteins) interact with MSH proteins and help to catalyze different MMR functions in an ATPase-dependent manner.

Sequence comparisons of MLH1 from different species show that D132H is highly conserved (Fig. 3). The MLH1 N terminus has an ATPase function<sup>8</sup>. To determine the functional effects of the MLH1 D132H substitution, we expressed the N terminus of MLH1 (amino acids 1–344) containing the ATPase and ATP binding domains with and without this sequence variant in bacteria as a glutathione S-transferase (GST) fusion protein. Consistent with previous studies, GST alone had essentially no ATPase activity, whereas the wild-type MLH1 N terminus fused to GST had measurable activity<sup>8–10</sup>. The MLH1 D132H substitution attenuated, but did not eliminate, ATPase activity (Fig. 3).

To understand more precisely the mechanisms by which MLH1 D132H disrupts protein function, we mapped the substitution onto the crystal structures of MLH1 homologs *E. coli* MutL and human PMS2. Asp132 is situated in an evolutionarily conserved  $\beta$ -hairpin structure that is part of the ATP binding and hydrolysis domain and buttresses the 'ATP lid'. In the PMS2 structure, a pair of charged residues His139 (equivalent to MLH1 Asp132) and Asp119 (equivalent to MLH1 His112 at the end of the ATP lid) interact to stabilize the ATP lid. In MLH1, these charged residues are swapped, and MLH1 Tyr157, which substitutes a sterically smaller serine residue in PMS2, bridges the interaction between MLH1 Asp132 and His112. Replacing MLH1 Asp132 with histidine interrupts the interactions between Asp132 and His112

**Table 1 Association between the *MLH1* 415C, *MSH6* 1613C and *MSH6* 2633C variant alleles and risk of CRC**

Set	Allele	Number of CRC cases	Number of controls	OR (95% c.i.)
<i>MLH1</i> 415C				
Test set	Mutated	5	1	
	Wild-type	445	443	5.0 (0.6–42.8)
Validation	Mutated	16	4	
	Wild-type	786	910	4.6 (1.5–13.9)
Combined	Mutated	21	5	
	Wild-type	1,231	1,353	4.6 (1.7–12.3)
<i>MSH6</i> 1613C				
Test set	Mutated	22	12	
	Wild-type	424	432	1.9 (0.9–3.8)
Validation	Mutated	19	21	
	Wild-type	779	899	1.0 (0.6–2.0)
Combined	Mutated	41	33	
	Wild-type	1,176	1,184	1.1 (0.8–1.8)
<i>MSH6</i> 2633C				
Test set	Mutated	6	17	
	Wild-type	444	430	0.30 (0.1–0.9)
Validation	Mutated	19	29	
	Wild-type	748	885	0.80 (0.4–1.4)
Combined	Mutated	25	46	
	Wild-type	1,192	1,315	0.60 (0.4–0.99)

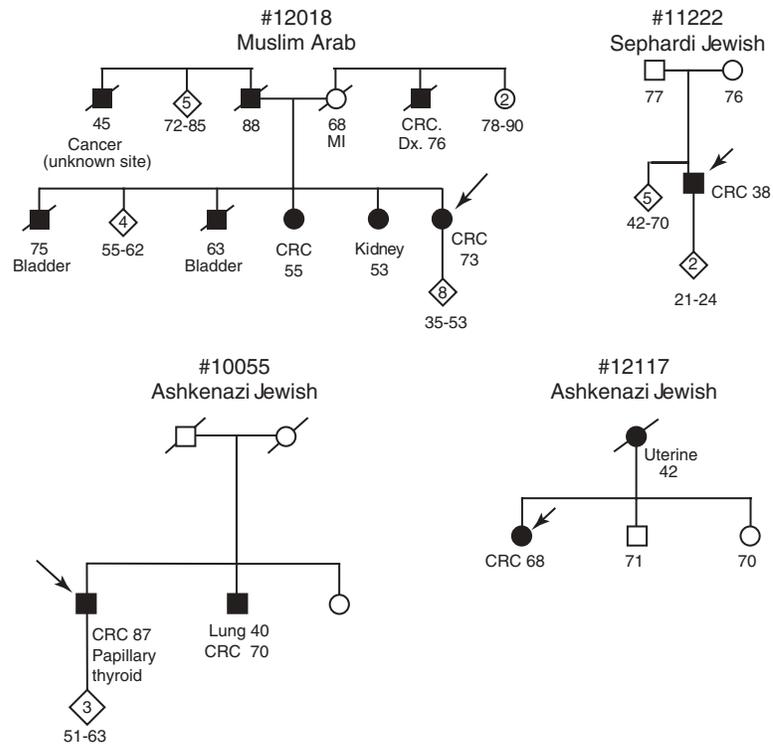
(Fig. 3), which in turn destabilizes the ATP lid and reduces the ATPase activity. The importance of the  $\beta$ -hairpin is shown by the MLH1 A128P mutation associated with HNPCC, which disrupts the  $\beta$ -hairpin formation completely rather than destabilizing a specific interaction. The effects of mutations that impair a protein function are often proportional to the spatial distance from the center of action. MLH1 D132H is more distant from the ATPase active site ( $\sim 16$  Å) than MLH1 A128P and other mutations that cause HNPCC, such as MLH1 M35R and MLH1 S44F (7–9 Å). Structural analysis predicts that MLH1 D132H will attenuate, but not eliminate, MLH1 ATPase activity.

Perhaps the most notable feature of MLH1 D132H is that it does not cause a strong MSI phenotype like classic *MLH1* mutations. Previously, *MLH1* K618A and *MLH1* E578G were described in probands with HNPCC and CRCs without MSI<sup>11,12</sup>. But these mutations occur in a region of the protein that is not conserved across multiple species, and thus their interpretation has been controversial<sup>13–15</sup>. An analogous situation is seen with *MSH6* alleles. Classic *MSH6* mutations cause CRC without MSI<sup>16</sup>, but a variant in the *MSH6* ATP binding site was identified that confers CRC with MSI<sup>17</sup>.

Our studies provide strong molecular epidemiologic, structural, bioinformatic and functional evidence that MLH1 D132H underlies a new phenotype relevant to ‘sporadic’ CRC without MSI and suggest that such alleles may account for a higher proportion of population-based susceptibility to CRC than previously realized. Similarly, promoter polymorphisms that cause a partial reduction in MMR gene transcription may contribute to sporadic CRC<sup>18</sup>.

In individuals carrying *MLH1* 415G $\rightarrow$ C mutations, loss of heterozygosity rarely occurs (Supplementary Table 4 online). One explanation is that, as in sporadic CRC, *MLH1* promoter hypermethylation of the wild-type allele<sup>19</sup> occurs instead of loss of heterozygosity. Alternatively, MLH1 D132H ATP-binding activity is notably increased relative to that of wild-type MLH1 at the same time that ATPase activity is decreased (data not shown). Together, these results suggest that MLH1 D132H bound to ATP may be ‘trapped’ in a transition state and unable to catalyze ATP hydrolysis. This trapped MLH1 D132H protein complexed to MSH2, MSH6 and PMS2-MLH3 may increase susceptibility to CRC through a dominant-negative mechanism. Future mechanistic studies will be necessary to define more precisely the mechanism of MLH1 D132H action.

Future studies analyzing knock-in mice expressing Mlh1 D132H will be important for understanding precisely the different MMR-mediated mechanisms by which the attenuated allele causes cancer susceptibility. Furthermore, analyzing susceptibility to



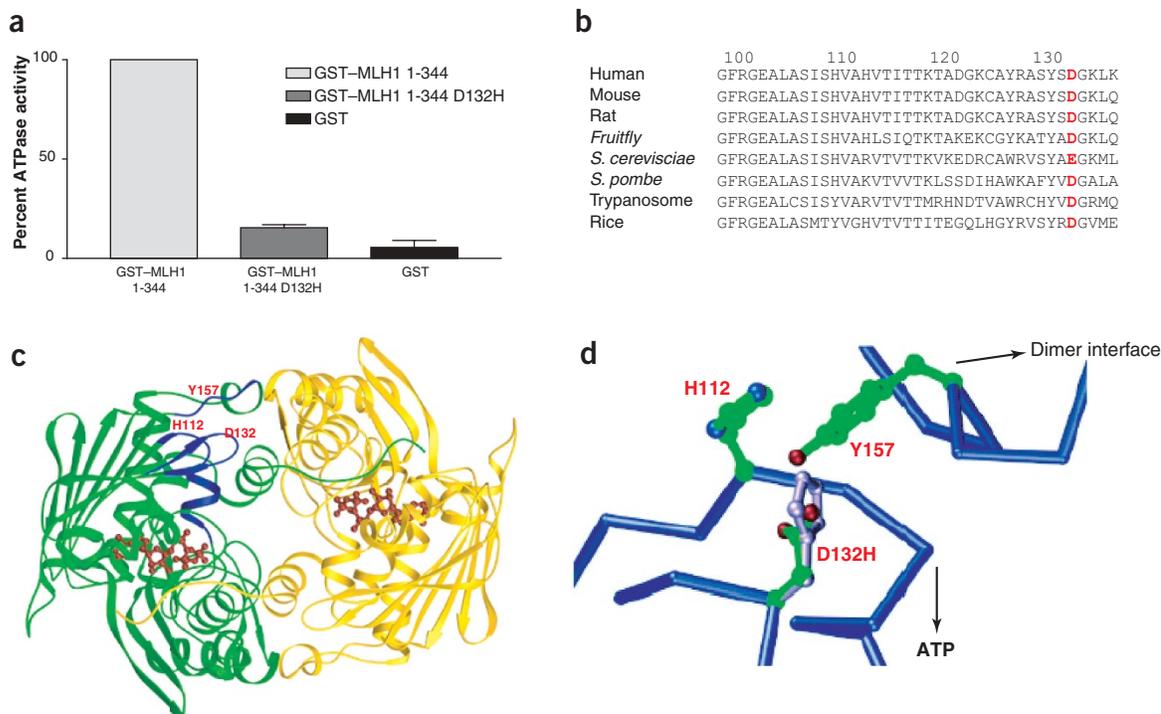
**Figure 2** Pedigrees of selected carriers of *MLH1* 415C. Proband is indicated by the arrows. The self-reported ethnicity of each family is indicated. Current age, or age at death and cause of death (generally cancer of the indicated organ), is indicated. MI, myocardial infarction.

gastrointestinal tumors in mice expressing Mlh1 D132H versus *Mlh1*-null mice models will help to quantify more precisely the cancer risks attributable to inactivation of the non-MSI mismatch repair functions of MLH1.

In the Ashkenazim, founder mutations in *BLM*, *BRCA1*, *BRCA2*, *FANCC*, *MSH2* and *APC* that are associated with susceptibility to cancer have been described<sup>4,6,20–22</sup>. Together with the *APC* allele encoding I1307K (10% carrier frequency in Ashkenazim with CRC) and *BLM*<sup>Ash</sup> (1.3% carrier frequency), which are associated with susceptibility to

**Table 2** Characteristics of population-based validation set and carriers of *MLH1* 415C with CRC

	Number of cases (%)	Number of controls (%)	Number of carriers of <i>MLH1</i> 415C with CRC (%)
Self-reported religion:			
Jew	704 (85.2)	816 (87.0)	17 (81)
Christian Arab	34 (4.0)	37 (4.0)	2 (10)
Muslim Arab	43 (5.4)	53 (5.6)	1 (5)
Druze Christian	14 (1.7)	10 (1.1)	1 (5)
Bedouin	1 (0.1)	1 (0.1)	0
Christian non-Arab	30 (3.6)	21 (2.2)	0
Self-reported Jewish ethnicity:			
Ashkenazi	454 (65.5)	480 (59.0)	12 (70)
Sephardi	224 (32.3)	315 (39.0)	4 (25)
Both	6 (0.9)	10 (1.23)	1 (6.25)
Other	9 (1.3)	6 (0.8)	
Percentage of males	49.1	49.1	60
Mean age (s.d.)	69.9 (11.6)	68.6 (12.3)	69.9



**Figure 3** Structural and functional analyses of MLH1 D132H. **(a)** Comparison of ATPase activity in MLH1 D132H and wild-type MLH1 *in vitro* using MLH1 amino acids 1–344 expressed as a GST fusion protein. **(b)** ClustalW sequence alignment of MLH1 N terminus in different species. MLH1 Asp132 is highlighted in red. **(c)** Modeling of the MLH1 D132H substitution based on the crystal structures of *E. coli* MutL and human PMS2. The dimeric structure of the ATPase domain of *E. coli* MutL in the presence of a nonhydrolyzable ATP analog, ADPnP (PDB: 1B63). The two MutL subunits represented by ribbon diagrams are shown in green and yellow. The ADPnP molecules are represented by ball-and-sticks in red. The secondary structures potentially disturbed by the D132H mutation in MLH1 are highlighted in blue and labeled with three residues (His112, Asp132 and Tyr157) that interact to stabilize the protein structure. **(d)** Asp132 and its interacting partners, His119 and Tyr164, in human MLH1 are modeled onto the human PMS2 structure (PDB: 1H7U). Protein backbone is represented by calcium traces in blue, the three interacting side chains are shown in green ball-and-sticks, oxygen atoms are shown in red and nitrogen atoms are shown in blue. Asp139 forms a hydrogen bond with Tyr164 and charged interactions with His119. Replacing Asp139 with histidine (shown in gray) would result in clash of His132 with Tyr164 and perturb the interaction with His119. In turn, these perturbations could affect ATP binding, ATPase and dimer formation as indicated.

CRC without MSI, and the *MSH2* allele encoding A636P (0.6% carrier frequency), which is associated with susceptibility to CRC with MSI, the identification of *MLH1* 415G→C (1.3% carrier frequency) should help identify Ashkenazim at greater risk for CRC. As *MLH1* 415G→C occurs in various ethnic groups, future population-based studies in other populations will be required to estimate more precisely the risk of CRC associated with this allele.

## METHODS

**Israeli individuals with CRC and matched controls.** The MECC study is a population-based, matched case-control study that includes more than 3,000 incident CRC cases and controls. The MECC study participants were previously described<sup>22,23</sup>. The sampling frame is defined by all men, women and children at risk of CRC who live in the geographic region of northern Israel that includes the Northern and Haifa Districts. Eligible cases include any person newly diagnosed with CRC between 1 January 1999 and 31 March 2004 in this geographic region. Eligible cases were invited to participate and interviewed within 6 months of diagnosis on average. Potential controls are identified from the same sampling frame by generating a list of individuals from the Clalit Health Services database, matched with respect to exact year of birth, sex and clinic code. Ethnicity is evaluated in detail in the interview for better control of this source of variation in the analyses. Individuals previously diagnosed with CRC are not eligible to participate. The study was approved by the Institutional Review Boards at the University of Michigan and Carmel Medical Center in Haifa. Written, informed consent was required for eligibility.

## High-density oligonucleotide array-based identification of sequence variants.

We designed a pair of oligonucleotide microarrays (Affymetrix), each containing 250,000 25-nucleotide probes, to interrogate sense and antisense sequences for all possible *MLH1*, *MSH2* and *MSH6* sequence variations in the 45 coding exons and >2 kb of total perixonic flanking sequences. These included probes complementary to all possible *MLH1*, *MSH2* and *MSH6* single nucleotide substitutions, one-, two-, three- and four-base-pair deletions and one- and two-base-pair insertions. A series of perfect match probes complementary to every 25-nucleotide segment of the *ATM* coding region are present in quadruplicate for both sense and antisense sequences across the arrays. We prepared genomic DNA from study participants and hybridized the arrays as previously described<sup>24</sup>. We wrote custom sequence analysis software to analyze the HNPCC Chip using DNA chip sequence normalization algorithms originally tested, and then improved, for *BRCA1*, *BRCA2* and *ATM* DNA chips. This analysis software is available on request. We set the signals present after normalization to data generated from separate two-color cohybridization experiments (to correct for reproducible fluctuations in the ratio of test and reference perfect match probe hybridization) to a value of one for the exon containing a probe and to zero for all other exons. Two additional quantities reflect multiplicative intra- and intermolecular probe structure normalization scores. These attempt to normalize for inter- and intramolecular array probe hybridization and secondary structures that may also contribute to fluctuations in perfect match probe signal ratios. Similarly, we incorporated an algorithm to predict the potential for duplex formation between adjacent probes (intermolecular structure) within a feature in the array (intramolecular structure). The inter- and intramolecular probe structure predictions differ in that complementary nucleotides less than five positions apart are considered only in

intramolecular probe structure analysis. This reflects the need for a loop of at least four residues to be present in a stable intramolecular hairpin structure. Finally, we averaged corrected ratios of test and reference perfect match probe hybridization signals against those generated from individual comparisons with data sets generated from multiple HNPCC Chip hybridization experiments. This provides a final set of multiplicative correction factors to minimize systematic fluctuations in hybridization signal ratios. We then plotted the ratios of reference and test perfect match probe signals against their respective nucleotide positions. We scored the loss-of-hybridization peaks according to peak height and width. We confirmed the sequences of all variants identified in study participants by fluorescent dideoxy DNA sequencing (Applied) using PolyPhred sequence analysis software.

**SNP genotyping.** We genotyped SNPs using Masscode Technology (QIAGEN GmbH), which uses highly multiplexed detection of low molecular weight photocleavable Mass Tags in a single quadrupole mass spectrometer. The assay uses a two-step PCR process for signal amplification and SNP discrimination. Allele-specific incorporation of the Mass Tags occurs during the second amplification step, as previously described<sup>25</sup>. We carried out SNP assays using 4 ng of genomic DNA. We included positive and negative (no DNA) controls on every 96-well plate.

**Statistical genetic analyses of variant allele frequencies.** We carried out matched and unmatched univariate analyses on each individual SNP using conditional and unconditional logistic regression as implemented in SAS version 8.2 (SAS Institute). We analyzed each SNP as presence versus absence of the minor allele.

**MSI assays.** We analyzed MSI as previously described<sup>23</sup>. Tumors from 11 case subjects who carried the 415G→C polymorphism were available for MSI analysis. We extracted normal and tumor DNA from microdissected DNA and analyzed it for the consensus panel of five markers<sup>26</sup>. We labeled forward primers for Bat 25, Bat 26, D2S125 and D5S346 and reverse primers for D17S250 with [<sup>32</sup>P]ATP and included them in a 20- l PCR reaction that included 1 l of microdissected DNA. We separated PCR products on 6% polyacrylamide gels for ~3 h at 65 watts and exposed them to film at -80 °C for 12–20 h. We double-scored the films and marked them as stable, unstable or indicating loss of heterozygosity.

**Microsatellite genotyping for conserved haplotype analysis and loss of heterozygosity studies.** We genotyped four microsatellite markers (5'–3': D3S1298, D3S1277, D3S1611 and D3S3527) in and around MLH1 in 24 carriers and 21 noncarriers of the 415C allele to determine whether this allele was part of a conserved haplotype. Genotyping for D3S1277 was done in the University of Michigan Sequencing Core. Genotyping for D3S1298, D3S1611 and D3S3527 was optimized in the laboratory using <sup>32</sup>P-ATP-labeled forward primers with the annealing temperatures 54 °C (D3S1298), 60 °C (D3S1611) and 56 °C (D3S3527). We separated the PCR products on a 6% polyacrylamide gel for 3–4 h at 65 watts and exposed them at -80 °C for 12–20 h. We analyzed individuals from the Centre d'Etude du Polymorphisme Humain (CEPH) family 1347 with known genotypes alongside the study participants and used their results to double-score the genotypes.

**Mapping of MLH1 415G→C onto MutL and PMS2 N-terminal crystal structures.** First, we superimposed the structures of *E. coli* MutL and human PMS2 ATPase domains, which showed that the ATPase active site, including the  $\beta$ -hairpin that contains MLH1 D132, is conserved. PMS2 His139, equivalent to MLH1 D132, forms a hydrogen bond with Asp119, which is equivalent to MLH1 His112. Therefore, we replaced His139 of PMS2 (using the ONO graphic program<sup>27</sup> on a Silicon Graphic Workstation) with an aspartate (Asp132) and replaced Asp119 of PMS2 with a histidine (His112). We used the amino acid rotamer library to optimize the side chain conformations. We noticed that Tyr157 of MLH1 would be in the vicinity of Asp132 and His112 based on the *E. coli* MutL structure, which contains a tyrosine at the equivalent position rather than a serine as found in PMS2. The resulting model of MLH1 structure (Fig. 3b) is a composite of human PMS2 (all the backbone atoms and side chains of Asp132 and His112) and *E. coli* MutL (Tyr157 position).

**ATPase biochemical studies of protein sequences of wild-type and D132H MLH1.** We amplified amino acids 1–304 of wild-type and D132H MLH1 using Pfx Polymerase (Invitrogen) and ligated them in-frame to the C terminus of GST (pGEX KT; Amersham Pharmacia), with a stop codon immediately following the MLH1 coding sequences. We purified GST fusion proteins to complete homogeneity as assessed by Coomassie-stained acrylamide gels as previously described<sup>28</sup>. We carried out ATPase comparative activity assays using equal amounts of protein for the wild-type and mutated allele and quantified and analyzed them as previously described<sup>29,30</sup>.

Note: Supplementary information is available on the Nature Genetics website.

#### ACKNOWLEDGMENTS

We would like to thank J. Nguyen and L. Paxton for technical assistance. This work was supported by an American Cancer Society Research Scholar Grant (to S.M.L.), Public Health Service grants (to S.B.G. and L.S.R.) and the Ravitz Foundation (S.B.G.).

#### COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the Nature Genetics website for details).

Received 4 February; accepted 22 April 2004

Published online at <http://www.nature.com/naturegenetics/>

- Wijnen, J. *et al.* Familial endometrial cancer in female carriers of MSH6 germline mutations. *Nat. Genet.* **23**, 142–144 (1999).
- Charames, G.S., Millar, A.L., Pal, T., Narod, S. & Bapat, B. Do MSH6 mutations contribute to double primary cancers of the colorectum and endometrium? *Hum. Genet.* **107**, 623–629 (2000).
- Peterlongo, P. *et al.* MSH6 germline mutations are rare in colorectal cancer families. *Int. J. Cancer* **107**, 571–579 (2003).
- Szabo, C.I. *et al.* Are ATM mutations 7271T→G and IVS10-6T→G really high-risk breast cancer-susceptibility alleles? *Cancer Res.* **64**, 840–843 (2004).
- Laken, S.J. *et al.* Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat. Genet.* **17**, 79–83 (1997).
- de Wind, N. *et al.* HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. *Nat. Genet.* **23**, 359–362 (1999).
- Spampinato, C. & Modrich, P. The MutL ATPase is required for mismatch repair. *J. Biol. Chem.* **275**, 9863–9869 (2000).
- Tran, P.T. & Liskay, R.M. Functional studies on the candidate ATPase domains of *Saccharomyces cerevisiae* MutLalpha. *Mol. Cell Biol.* **20**, 6390–6398 (2000).
- Scholz, B., Rechter, S., Drach, J.C., Townsend, L.B. & Bogner, E. Identification of the ATP-binding site in the terminase subunit pUL56 of human cytomegalovirus. *Nucleic Acids Res.* **31**, 1426–1433 (2003).
- Guarne, A., Junop, M.S. & Yang, W. Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *EMBO J.* **20**, 5521–5531 (2001).
- Liu, T. *et al.* Missense mutations in hMLH1 associated with colorectal cancer. *Hum. Genet.* **105**, 437–441 (1999).
- Farrington, S.M. *et al.* Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. *Am. J. Hum. Genet.* **63**, 749–759 (1998).
- Scartozzi, M. *et al.* Mutations of hMLH1 and hMSH2 in patients with suspected hereditary nonpolyposis colorectal cancer: correlation with microsatellite instability and abnormalities of mismatch repair protein expression. *J. Clin. Oncol.* **20**, 1203–1208 (2002).
- Samowitz, W.S. *et al.* Missense mismatch repair gene alterations, microsatellite instability, and hereditary nonpolyposis colorectal cancer. *J. Clin. Oncol.* **20**, 3178–3179 (2002).
- Guerrette, S., Acharya, S. & Fishel, R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J. Biol. Chem.* **274**, 6336–6341 (1999).
- Kolodner, R.D. Germline MSH6 mutation in colorectal cancer families. *Cancer Res.* **59**, 5068–5074 (1999).
- Das Gupta, R. & Kolodner, R.D. Novel dominant mutations in *Saccharomyces cerevisiae* MSH6. *Nat. Genet.* **24**, 53–56 (2000).
- Gazzoli, I. & Kolodner, R.D. Regulation of the human MSH6 gene by the Sp1 transcription factor and alteration of promoter activity and expression by polymorphisms. *Mol. Cell Biol.* **23**, 7992–8007 (2003).
- Lynch, H.T. & de la Chapelle, A. Hereditary colorectal cancer. *N. Engl. J. Med.* **348**, 919–932 (2003).
- Shiri-Sverdlov, R. *et al.* Mutational analyses of BRCA1 and BRCA2 in Ashkenazi and non-Ashkenazi Jewish women with familial breast and ovarian cancer. *Hum. Mutat.* **16**, 491–501 (2000).
- Foulkes, W.D. *et al.* The founder mutation MSH2\*1906G→C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *Am. J. Hum. Genet.* **71**, 1395–1412 (2002).
- Gruber, S.B. *et al.* BLM heterozygosity and the risk of colorectal cancer. *Science* **297**, 2013 (2002).

23. Greenson, J.K. *et al.* 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 (2003).
24. Hacia, J.G. *et al.* Two color hybridization analysis using high density oligonucleotide arrays and energy transfer dyes. *Nucleic Acids Res.* **26**, 3865–3866 (1998).
25. Kokoris, M. *et al.* High-throughput SNP genotyping with the Masscode system. *Mol. Diagn.* **5**, 329–340 (2000).
26. Boland, C.R. *et al.* A National Cancer Institute Workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257 (1998).
27. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47** (Pt 2), 110–119 (1991).
28. Lipkin, S.M., Naar, A.M., Kalla, K.A., Sack, R.A. & Rosenfeld, M.G. Identification of a novel zinc finger protein binding a conserved element critical for Pit-1-dependent growth hormone gene expression. *Genes Dev.* **7**, 1674–1687 (1993).
29. Acharya, S., Foster, P.L., Brooks, P. & Fishel, R. The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *Mol. Cell.* **12**, 233–246 (2003).
30. Quaresima, B. *et al.* Human mismatch-repair protein MutL homologue 1 (MLH1) interacts with Escherichia coli MutL and MutS in vivo and in vitro: a simple genetic system to assay MLH1 function. *Biochem. J.* **371**, 183–189 (2003).