IJC
International Journal of Cancer

An American founder mutation in *MLH1*

Jerneja Tomsic[1], Sandya Liyanarachchi[1], Heather Hampel[1], Monika Morak[2,3], Brittany C. Thomas[4], Victoria M. Raymond[5,6,7], Anu Chittenden[8], Hans K. Schackert[9], Stephen B. Gruber[5,6,7], Sapna Syngal[8,10], Alessandra Viel[11], Elke Holinski-Feder[2,3], Stephen N. Thibodeau[4] and Albert de la Chapelle[1]

[1] Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, Columbus, OH
[2] Medical Department Campus Innenstadt, University Clinic, Munich, Germany
[3] MGZ, Center of Medical Genetics, Munich, Germany
[4] Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, MN
[5] Department of Epidemiology, University of Michigan, Ann Arbor, MI
[6] Department of Internal Medicine, University of Michigan, Ann Arbor, MI
[7] Department of Human Genetics, University of Michigan, Ann Arbor, MI
[8] Population Sciences Division, Dana-Farber Cancer Institute, Boston, MA
[9] Department of Surgical Research, Dresden University of Technology, Dresden, Germany
[10] Gastroenterology Division, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA
[11] Experimental Oncology 1, CRO IRCCS, National Cancer Institute, Aviano, Italy

Mutations in the mismatch repair genes cause Lynch syndrome (LS), conferring high risk of colorectal, endometrial and some other cancers. After the same splice site mutation in the *MLH1* gene (c.589-2A>G) had been observed in four ostensibly unrelated American families with typical LS cancers, its occurrence in comprehensive series of LS cases (Mayo Clinic, Germany and Italy) was determined. It occurred in 10 out of 995 LS mutation carriers (1.0%) diagnosed in the Mayo Clinic diagnostic laboratory. It did not occur among 1,803 cases tested for *MLH1* mutations by the German HNPCC consortium, while it occurred in three probands and an additional five family members diagnosed in Italy. In the U.S., the splice site mutation occurs on a large (~4.8 Mb) shared haplotype that also harbors the variant c.2146G>A, which predicts a missense change in codon 716 referred to here as V716M. In Italy, it occurs on a different, shorter shared haplotype (~2.2 Mb) that does not carry V716M. The V716M variant was found to be present by itself in the U.S., German and Italian populations with individuals sharing a common haplotype of 280 kb, allowing us to calculate that the variant arose around 5,600 years ago (225 generations; 95% confidence interval 183–272). The splice site mutation in America arose or was introduced some 450 years ago (18 generations; 95% confidence interval 14–23); it accounts for 1.0% all LS in the Unites States and can be readily screened for.

Germline mutations that occur in more than one or just a few individuals are of two distinct types. The first type, here referred to as "recurrent" arises repeatedly *de novo*, usually because of a sequence peculiarity that predisposes to an abnormal event at meiosis. The second type, referred to here as a "founder" mutation, arose in a single ancestor and is subsequently inherited by numerous descendants.

Lynch syndrome (LS), previously known as hereditary nonpolyposis colorectal cancer (HNPCC), is caused by germline mutations in the mismatch repair genes *MSH2, MLH1, MSH6* and *PMS2.* LS is characterized by an extremely high risk of colorectal and endometrial cancer and to a lesser degree, several other cancers. Both recurrent and founder mutations in the mismatch repair genes have been amply described. As an example, a prototype recurrent mutation in *MSH2* is an A to T transversion in the third nucleotide of the donor splice site of intron 5 (abbreviated c.942+3A>T) leading to the skipping of exon 5 resulting in loss of *MSH2* function. The sequence peculiarity predisposing to this recurrent mutation is apparently the fact that the mutated A is the first of a tract of 26 adenines, which predisposes to misalignment during meiotic pairing.[1,2] This recurrent mutation occurs worldwide and accounts for 5–10% of all *MSH2* mutations. A prototype LS founder mutation is the c.1906G>C transversion in *MSH2* that leads to a missense amino acid

Key words: *MLH1*, colon cancer, founder mutation, genetic predisposition
Additional Supporting Information may be found in the online version of this article.
Conflict of interest: Ms. Hampel has received honoraria for advisory group work from Myriad Genetic Laboratories.
Grant sponsor: National Cancer Institute; **Grant numbers:** CA16058, CA67941
DOI: 10.1002/ijc.26233
History: Received 15 Mar 2011; Accepted 19 May 2011; Online 13 Jun 2011

Correspondence to: Albert de la Chapelle, M.D., Ph.D., Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, 804 Biomedical Research Tower, 460 W. 12th Avenue, Columbus, OH 43210, USA, Tel.: +614-688-4781, Fax: +614-688-4772, E-mail: albert.delachapelle@osumc.edu

substitution of alanine to proline, A632P.[3] As is the case with many founder mutations, this mutation occurs only in one ethnic population, in this case the Ashkenazi Jews, where it may account for approximately one-third of all LS cases.[3]

Recurrent and founder mutations are medically important because they can facilitate diagnostic approaches. In populations where they occur with appreciable frequency, mutation testing can begin by a simple test for the main recurrent or founder mutations. If a mutation is found, further testing can be avoided thus resulting in cost savings. Founder mutations typically occur in isolated, ethnically distinct populations. The U.S. population is ethnically heterogeneous; therefore widespread founder mutations are not common. A remarkable exception is the deletion of exons 1–6 of the *MSH2* gene that is referred to as the American Founder Mutation.[4–6] In this communication, we describe a splice site mutation in *MLH1*. We show that this mutation is relatively common in the U.S. where it either arose in or was brought in by an early immigrant. Moreover, we show that the mutation occurred in an allele that already carried another more common founder sequence change in *MLH1* (missense variant V716M) so that the mutated allele contains both changes. We discuss the implications of these findings.

Material and Methods

Study samples

Our study included cases from Italy, Germany and the United States carrying either the deleterious *MLH1* splice site mutation (c.589-2A>G) or the innocent *MLH1* V716M variant or both. Informed consent, approved by our respective Institutional Review Boards to conduct genetic experiments, was obtained from the study subjects.

The Italian samples belonged to families A-AV23 and A-AV44[7] and A-AV24.[8] The newly acquired probands for families A-TN5 and A-TN6 were selected in Presidio Ospedaliero Santa Chiara in Trento (Italy) and sent to Centro Riferimento Oncologico in Aviano for genetic testing.

The German cases were counseled and underwent genetic testing at the facilities of the German HNPCC-Consortium as described elsewhere.[9] They were then entered into the German mutation database.

The first description of the splice site mutation was in a colorectal cancer patient.[10] We have studied North American samples from Boston,[11] Ohio,[12] Michigan and the Mayo Clinic Molecular Diagnostic Laboratory. Since the Mayo Clinic is a major reference laboratory in the U.S., the samples obtained from them were from patients undergoing clinical genetic testing for LS from around the U.S. These samples were anonymized to retain confidentiality. The Michigan family was identified in a clinical series tested at the University of Michigan Cancer Genetics Clinic.

Controls genotyped for haplotype analysis were Caucasian ($n = 78$) and African American ($n = 6$) from the Ohio State University Medical Center's Human Genetics Sample Bank. They are derived from the Columbus-area population. This collection of control samples is approved by the Biomedical Sciences Institutional Review Board at Ohio State University Medical Center.

Haplotype analysis

To characterize the haplotypes associated with the *MLH1* splice site and V716M changes we used 14 out of 15 microsatellite markers previously reported[13] that span the *MLH1* locus. We added three additional single nucleotide polymorphisms (SNPs) present in the *MLH1* gene. All available probands and family members and 84 controls were typed for these *MLH1* markers.

Microsatellite markers were typed utilizing a labeled M13 primer in conjunction with an M13-tailed, amplicon specific, primer in a three primer PCR. Each 15-μl PCR reaction contained 7.5 μl of AmpliTaq Gold master mix (PE Applied Biosystems, Foster City, CA), 25 ng of genomic DNA, 10 pmol of untailed primer, 5 pmol of M13-tailed primer and 10 pmol of the FAM-labeled M13 primer. Reactions were cycled using the following profile: 96°C for 10 min, 36 cycles of 96°C for 30 sec, 58°C for 30 sec, 72°C for 30 sec, and final extension at 72°C for 5 min. The PCR product was sized using an ABI 3730 DNA Analyzer.

For the genotyping of SNPs, we used the same PCR conditions as above in the presence of 10 pmol forward and reverse primers. The PCR product containing the SNP was subjected to the SNaPshot reaction (PE Applied Biosystems, Foster City, CA). The sequences for the primers are listed in Supporting Information Table 1.

Estimating the age of the variants

The DMLE+2.3 software developed by Reeve and Rannala[14] was used to estimate the age of the two variants. The program, which is freely available from www.dmle.org, uses a Bayesian approach to compare differences in linkage disequilibrium between the mutation and flanking markers, among DNA samples from mutation carriers and unrelated controls. This software uses genotype data for cases and controls, marker locations, population growth rates and an estimate of the proportion of the mutation bearing chromosomes being analyzed.

Marker locations were obtained from the human genome reference sequence (Human build 37.2). The deCODE genetic positions were known for most of the microsatellite markers. For the markers and SNPs not in the deCODE map, genetic distances were obtained with spline interpolation using known genetic positions of adjacent deCODE markers.

The age of the V716M variant in Europeans was calculated using German ($n = 13$) and Italian ($n = 2$) probands. For the population growth rate we used 1.05 fold per generation.[15,16] We have estimated the number of the V716M variant bearing chromosomes for the Italian and German populations using the frequency of the variant found in the samples tested, assuming that an innocent variant should have a similar frequency in controls as in cases. Based on the population sizes (Italy ~ 60 million, Germany ~ 82 million) the numbers of chromosomes were 402,791 and 907,399, respectively.

The age of the splice site mutation in the American population was calculated using a growth rate of 1.65 fold.[6] To estimate the proportion of disease bearing chromosomes studied, we used the following data: 5.5% lifetime risk for CRC; 2.8% of CRCs are LS and 1.0% of LS cases are due to this splice site mutation.

Additionally, we applied the Estiage program[17] to estimate the age of the most recent common ancestor carrying the V716M variant, since the shared haplotype is relatively small (280 kb), and it has been densely genotyped. We were unable to use the same method for the estimation of the most recent common ancestor carrying the splice site mutation. The shared haplotype is ~17-fold longer (~4.8 Mb) and the number of markers used increases only from 5 to 15, resulting in significant distances between the markers that is not suitable for the program.[17]

Results and Discussion

The MLH1 intron 7 splice site mutation

The c.589-2A>G mutation, first published by Luce *et al.*,[10] disrupts the acceptor splice site of intron 7 and results in the skipping of exon 8. This predicts loss of *MLH1* function and LS, both of which are amply confirmed by the data presented here. This mutation is one of several hundred different *MLH1* mutations that are known today based on two major databases (www.insight-group.org/mutation; http://www.med.mun.ca/mmrvariants/). Subsequently Syngal *et al.*[11] found the same mutation in a LS proband of a family that met the Amsterdam I criteria.

We found the splice site mutation in the proband of a large LS family identified in Ohio. The proband was enrolled in a population-based study of LS in 563 unselected newly diagnosed patients with endometrial cancer of whom 14 had LS (Refs. [12,18] and unpublished data). To assess the proportion of cases carrying the splice site mutation in the Ohio series, we can use additional data from a study of LS in 1,566 consecutive colorectal cancer patients, of whom 44 had LS.[19] Thus, the prevalence of the splice site mutation among probands with LS is 1/58 or 1.7%. The mutation occurs in six further members of the Ohio family, four of whom have been diagnosed with LS cancer. Of note, the splice site mutation is not mentioned in the original reports[12,19] because it was not called by the sequencing software and missed when the testing was first performed and was only recognized later in a collaboration with and thanks to our colleagues in Germany. Instead, the V716M missense variant was originally found in the proband and those family members in whom the splice site mutation was later detected (Pedigree in the Supporting Information Fig. 3). To conclusively show that the splice site mutation and the V716M variant were on the same chromosome, we used the method of conversion to haploidy[20] confirming that the two changes are in cis (see Supporting Information for description). We thus had evidence of the splice site mutation in three US families; a fourth family was communicated to us by Dr. Gruber, and it appeared that the proportion of all LS having the splice site mutation might be high enough to warrant further investigation. In our study, it was of interest to answer two questions. First, how common is the mutation in the US and elsewhere? Second, is it recurrent or of founder type?

Disease causing mutations are usually too rare to search for in the general population. Instead it is meaningful to ask how common the mutation is among LS probands. The data from individual laboratories are almost never extensive enough for this purpose, so we turned to three larger sources, the molecular diagnostic laboratory of the Mayo Clinic, the German HNPCC consortium and a collaborative group in Italy. Italy was studied because reports of the mutation had been published.[8,21]

The archives of a large molecular diagnostic laboratory (the Mayo Clinic) were searched for the splice site mutation. It had been seen in totally 10 individuals, 8 of whom were ostensibly unrelated, likely representing different families. During the same time period, a total of 995 cases with LS had been diagnosed in this laboratory. Thus, the proportion of carriers of the splice site mutation in the US was 1.0% (10/995), but none was found in Germany. While it is clearly desirable to study further cohorts from various geographic regions, we hypothesize that the mutation is indeed rare in most other populations than the US and Italy because it has not been reported to the databases quoted above. In total, we are presently aware of 22 individuals carrying the splice site mutation in the US (Table 1).

The MLH1 exon 19 missense variant

In the splice site mutation carrier, Syngal *et al.*[11] noted the presence of the V716M variant in exon 19, termed missense mutation, caused by the c.2146G>A sequence change. This change has been reported numerous times in smaller and larger series of patients tested for LS. For instance, the INSiGHT database (www.insight-group.org/mutations) states that it has been "cited 58 times" but due to double reporting the real number may be smaller. The amino acid in position 716 is only weakly conserved among species but resides in the MLH1 domain that binds PMS2. The wild type amino acid, valine is neutral and hydrophobic while methionine is also hydrophobic, but contains sulfur.

Typically, missense mutations can be either deleterious or nondeleterious, and it is almost impossible to predict the pathogenicity of missense changes based solely on the functional domain and the nature of the amino acids involved. Several groups of investigators have used a variety of strategies to investigate the variant's effect on mismatch repair. Raevaara *et al.*[22] examined protein expression, stability, subcellular localization, PMS2 interaction and mismatch repair efficiency and concluded that V716M is functionally normal. The same result emerged from functional assays in yeast,[23,24]

Table 1. Clinical features of c.589–2A>G mutation carriers

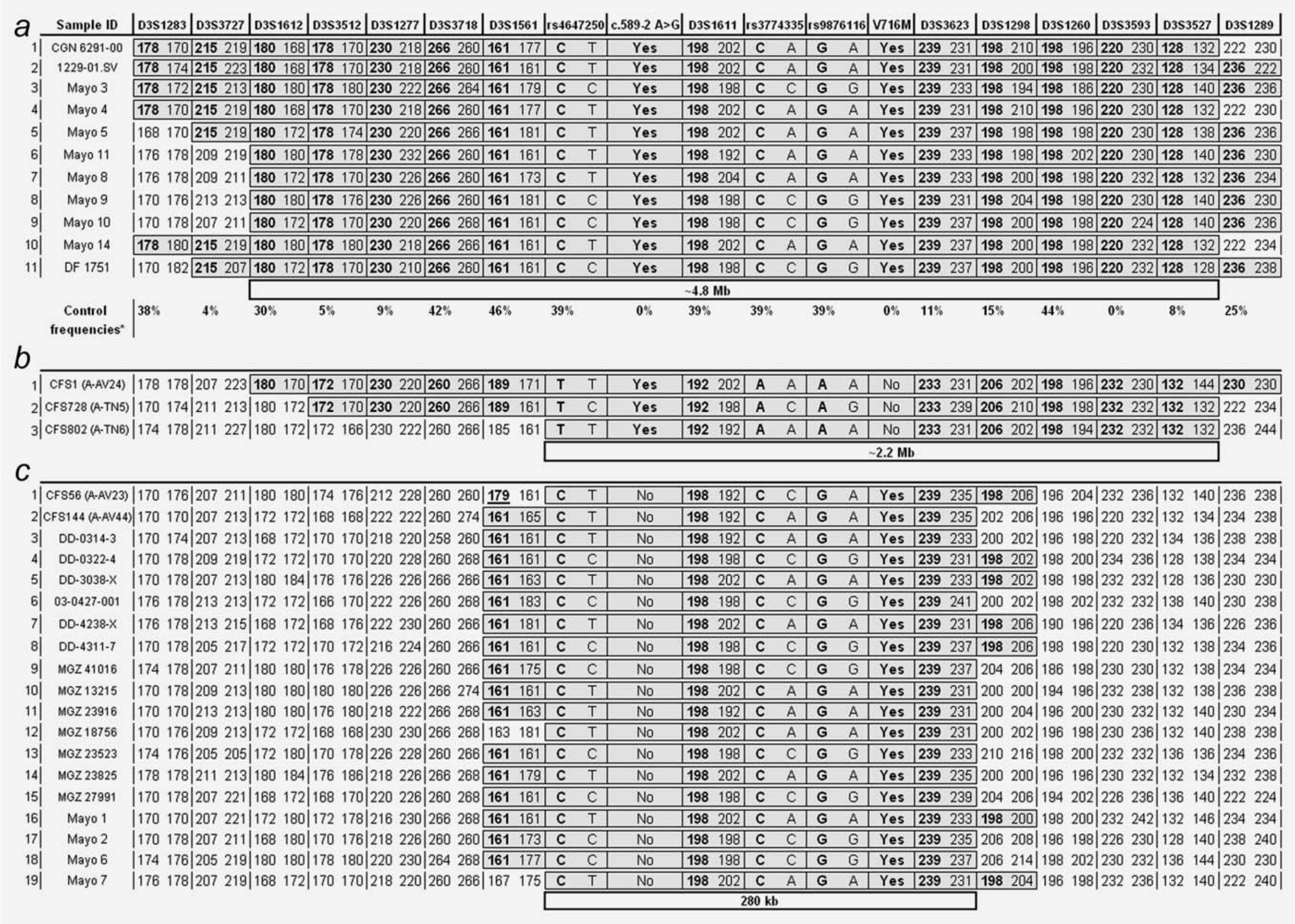
Case	Gender	Age[1]	Cancer History[2]	Relationship
U.S. Cases				
CGN 6291–00	Female	50	EC;48, Ovarian;48	Proband
1229–01.SV	Female	50	EC;39	Proband
1229–02.BV	Female	75	CRC;41, EC;63	Mother
1229–04.NV	Female	48	Unaffected	Sister
1229–07.2JV	Male	31	Unaffected	Son
1229–11.2PS	Male	72	Urothelial;60	Maternal Uncle
1229–25.1MS	Male	d.43	Small bowel;41	Maternal first cousin (son of 1229–11.2PS)
1229–11.1MS	Male	d.51	8–10 Colon polyps	Maternal Uncle
1229–24.1LH	Female	40	CRC;28	Maternal first cousin (daughter of 1229–11.3MS)
Mayo 3	Female	42	EC; by 42, CRC by 42; Ovarian by 42	Proband
Mayo 4	Female	48	EC; 48	Proband
Mayo 5	Male	37	CRC; by 37	Proband
Mayo 8	Male	50	Synchronous CRC; 45	Proband
Mayo 9	Male	33	Unaffected	Son of proband (not tested at Mayo)
Mayo 10	Female	65	CRC; age unknown	Relative of proband (not tested at Mayo)
Mayo 11	Male	78	CRC; 49, 51, 61	Proband
Mayo 12	Female	53	Endocervical; 45, Lung; 46, CRC; 50	Daughter of Mayo 11
Mayo 13	Female	23	CRC; 19	Granddaughter of Mayo 11
Mayo 14	Male	28	Unaffected	Son of proband (not tested at Mayo)
DF 1751	Male	d.75	CRC;33, CRC;39, Melanoma;65	Proband
	Female	d.61	Breast;39, CRC;47, CRC;59	Daughter of DF 1751
	Male	54	CRC;53	Son of DF 1751
Italian Cases				
CFS1 (A-AV24)	Female	70	EC;52	Proband
CFS87 (A-AV24)	Male	51	CRC;30	Cousin of CFS1
CFS88 (A-AV24)	Female	59	CRC;52, EC;57	Cousin of CFS1
CFS507 (A-AV24)	Female	46	EC;44, Ovarian;44	Daughter of CFS1
CFS629 (A-AV24)	Female	28	CRC;26	Granddaughter of CFS1 (daughter of CFS507)
CFS728 (A-TN5)	Female	63	EC;56, CRC;60	Proband
CFS802 (A-TN6)	Female	46	CRC;27, Stomach;43	Proband
CFS803 (A-TN6)	Male	64	CRC;56	Brother of CFS802

Abbreviations: CRC, colorectal cancer; EC, endometrial cancer; d, died.
[1] Current age/Age at death. [2] Numbers refer to age at diagnosis.

from a cell free complementation assay[25] and from a recent study combining several methods.[26] Thus, ample functional evidence strongly suggests that V716M does not affect mismatch repair.

Existing evidence from the case-report type publications points in the same direction. V716M has been reported numerous times, but it is not well established how many cases totally were studied for *MLH1*. In the Mayo Clinic series, V716M (without the splice site mutation) occurred in at least 7 individuals out of 1,385 in whom *MLH1* was studied for mutation (~0.5%). In the German series, 20 V716M carriers were found among 1,803 patients in whom *MLH1* was studied (1.1%). Several authors report having searched for V716M in controls and no V716M mutation carriers were found in a total of some 400 controls reported in these studies.[12,26–28] In a study of primary antibody deficiency syndromes in Sweden,[29] 991 control individuals were tested for V716M. The allele frequency was 3/1904 (*i.e.*, a carrier frequency of approximately 3/952, 0.3%). In a comprehensive study by Barnetson *et al.*,[30] the V716M variant was considered to be benign based on a variety of functional criteria. It was seen in 1/932 (0.1%) colorectal cancer patients and in 6/1,000 (0.6%) controls.

Thus, in summary, the proportion of V716M carriers in patients studied for LS (~0.8%) appears to be similar to the



a

	Sample ID	D3S1283	D3S3727	D3S1612	D3S3512	D3S1277	D3S3718	D3S1561	rs4647250	c.589-2 A>G	D3S1611	rs3774335	rs9876116	V716M	D3S3623	D3S1298	D3S1260	D3S3593	D3S3527	D3S1289
1	CGN 6291-00	**178** 170	**215** 219	**180** 168	**178** 170	**230** 218	**266** 260	**161** 177	**C** T	**Yes**	**198** 202	**C** A	**G** A	**Yes**	**239** 231	**198** 210	**198** 196	**220** 230	**128** 132	222 230
2	1229-01.SV	**178** 174	**215** 223	**180** 168	**178** 170	**230** 218	**266** 260	**161** 161	**C** T	**Yes**	**198** 202	**C** A	**G** A	**Yes**	**239** 231	**198** 200	**198** 198	**220** 232	**128** 134	**236** 222
3	Mayo 3	**178** 172	**215** 213	**180** 180	**178** 180	**230** 222	**266** 264	**161** 179	**C** C	**Yes**	**198** 198	**C** C	**G** G	**Yes**	**239** 233	**198** 194	**198** 186	**220** 230	**128** 140	**236** 236
4	Mayo 4	**178** 170	**215** 219	**180** 168	**178** 170	**230** 218	**266** 260	**161** 177	**C** T	**Yes**	**198** 202	**C** A	**G** A	**Yes**	**239** 231	**198** 210	**198** 196	**220** 230	**128** 132	222 230
5	Mayo 5	168 170	**215** 219	**180** 172	**178** 174	**230** 220	**266** 266	**161** 181	**C** T	**Yes**	**198** 202	**C** A	**G** A	**Yes**	**239** 237	**198** 198	**198** 198	**220** 230	**128** 138	**236** 236
6	Mayo 11	176 178	209 219	**180** 180	**178** 178	**230** 232	**266** 260	**161** 161	**C** T	**Yes**	**198** 192	**C** A	**G** A	**Yes**	**239** 233	**198** 198	**198** 202	**220** 230	**128** 140	**236** 230
7	Mayo 8	176 178	209 211	**180** 172	**178** 170	**230** 226	**266** 260	**161** 173	**C** T	**Yes**	**198** 204	**C** A	**G** A	**Yes**	**239** 233	**198** 200	**198** 198	**220** 232	**128** 132	**236** 234
8	Mayo 9	170 176	213 213	**180** 180	**178** 176	**230** 226	**266** 260	**161** 181	**C** C	**Yes**	**198** 198	**C** C	**G** G	**Yes**	**239** 231	**198** 204	**198** 198	**220** 230	**128** 140	**236** 230
9	Mayo 10	170 178	207 211	**180** 172	**178** 170	**230** 220	**266** 268	**161** 161	**C** C	**Yes**	**198** 198	**C** C	**G** G	**Yes**	**239** 237	**198** 200	**198** 198	**220** 224	**128** 140	**236** 236
10	Mayo 14	**178** 180	**215** 219	**180** 180	**178** 180	**230** 218	**266** 266	**161** 161	**C** T	**Yes**	**198** 202	**C** A	**G** A	**Yes**	**239** 237	**198** 200	**198** 198	**220** 232	**128** 132	222 234
11	DF 1751	170 182	**215** 207	**180** 172	**178** 170	**230** 210	**266** 260	**161** 161	**C** C	**Yes**	**198** 198	**C** C	**G** G	**Yes**	**239** 237	**198** 200	**198** 196	**220** 232	**128** 128	**236** 238
				~4.8 Mb																
	Control frequencies*	38%	4%	30%	5%	9%	42%	46%	39%	0%	39%	39%	39%	0%	11%	15%	44%	0%	8%	25%

b

	Sample ID	D3S1283	D3S3727	D3S1612	D3S3512	D3S1277	D3S3718	D3S1561	rs4647250	c.589-2 A>G	D3S1611	rs3774335	rs9876116	V716M	D3S3623	D3S1298	D3S1260	D3S3593	D3S3527	D3S1289
1	CFS1 (A-AV24)	178 178	207 223	**180** 170	**172** 170	**230** 220	**260** 266	**189** 171	**T** T	**Yes**	**192** 202	**A** A	**A** A	No	**233** 231	**206** 202	**198** 196	**232** 230	**132** 144	**230** 230
2	CFS728 (A-TN5)	170 174	211 213	180 172	**172** 170	**230** 220	**260** 266	**189** 161	**T** C	**Yes**	**192** 198	**A** C	**A** G	No	**233** 239	**206** 210	**198** 198	**232** 232	**132** 132	222 234
3	CFS802 (A-TN6)	174 178	211 227	180 172	172 166	230 222	260 266	185 161	**T** T	**Yes**	**192** 192	**A** A	**A** A	No	**233** 231	**206** 202	**198** 194	**232** 232	**132** 132	236 244
									~2.2 Mb											

c

	Sample ID	D3S1283	D3S3727	D3S1612	D3S3512	D3S1277	D3S3718	D3S1561	rs4647250	c.589-2 A>G	D3S1611	rs3774335	rs9876116	V716M	D3S3623	D3S1298	D3S1260	D3S3593	D3S3527	D3S1289
1	CFS56 (A-AV23)	170 176	207 211	180 180	174 176	212 228	260 260	**179** 161	**C** T	No	**198** 192	**C** C	**G** A	**Yes**	**239** 235	**198** 206	196 204	232 236	132 140	236 238
2	CFS144 (A-AV44)	170 170	207 213	172 172	168 168	222 222	260 274	**161** 165	**C** T	No	**198** 192	**C** A	**G** A	**Yes**	**239** 235	202 206	196 196	220 232	132 134	234 238
3	DD-0314-3	170 174	207 213	168 172	170 170	218 220	258 260	**161** 161	**C** T	No	**198** 192	**C** A	**G** A	**Yes**	**239** 233	200 202	196 198	220 232	134 136	238 238
4	DD-0322-4	170 178	209 219	172 172	170 170	220 228	260 268	**161** 161	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 231	**198** 202	198 200	234 236	128 138	234 234
5	DD-3038-X	170 178	207 213	180 184	176 176	226 226	266 266	**161** 163	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 233	**198** 202	198 198	232 232	128 136	230 230
6	03-0427-001	176 178	213 213	172 172	166 170	222 226	260 268	**161** 183	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 241	200 202	198 202	232 232	138 140	230 238
7	DD-4238-X	176 178	213 215	168 172	168 176	222 230	260 266	**161** 181	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 231	**198** 206	190 196	220 236	134 136	226 236
8	DD-4311-7	170 178	205 217	172 172	170 172	216 224	260 266	**161** 161	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 237	**198** 206	198 198	220 230	132 138	234 236
9	MGZ 41016	174 178	207 211	180 180	176 178	226 226	260 266	**161** 175	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 237	204 206	186 198	230 230	132 138	234 234
10	MGZ 13215	170 178	209 213	180 180	180 180	226 226	266 274	**161** 161	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 231	200 200	194 196	232 238	132 138	236 238
11	MGZ 23916	170 170	213 213	180 180	176 180	218 222	266 268	**161** 163	**C** T	No	**198** 192	**C** A	**G** A	**Yes**	**239** 231	200 204	196 200	230 232	132 140	230 234
12	MGZ 18756	170 176	209 213	172 172	168 168	230 230	266 268	163 181	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 231	200 202	196 198	230 236	132 140	238 238
13	MGZ 23523	174 176	205 205	172 180	170 178	226 228	260 266	**161** 161	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 233	210 216	198 200	232 232	136 136	234 236
14	MGZ 23825	178 178	211 213	180 184	176 186	218 226	266 268	**161** 179	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 235	200 200	196 196	230 232	132 134	232 238
15	MGZ 27991	170 178	207 221	168 172	168 170	220 226	260 268	**161** 161	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 239	204 206	194 202	226 236	136 140	222 224
16	Mayo 1	170 170	207 221	172 180	172 178	216 230	266 268	**161** 161	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 233	**198** 200	198 200	232 242	132 146	234 234
17	Mayo 2	170 178	207 211	168 180	170 176	218 226	260 260	**161** 173	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 235	206 208	196 198	226 230	128 140	238 240
18	Mayo 6	174 176	205 219	180 180	178 180	220 230	264 268	**161** 177	**C** C	No	**198** 198	**C** C	**G** G	**Yes**	**239** 237	206 214	198 202	230 232	136 144	230 230
19	Mayo 7	176 178	207 219	168 172	170 170	218 220	260 266	167 175	**C** T	No	**198** 202	**C** A	**G** A	**Yes**	**239** 231	**198** 204	196 198	232 236	132 140	222 240
												280 kb								

Figure 1. Genotype data spanning the *MLH1* locus in 33 probands. Haplotypes associated with the mutations are highlighted in gray, with the mutation allele bolded. (*a*) US samples with both the splice site mutation and the V716M variant. (*b*) Italian samples carrying only the splice site mutation. (*c*) Italian (1, 2), German (3–15) and US (16–20) samples with the V716M variant. The sizes of the shared haplotypes are indicated with empty bars below each group. Underlined is the allele associated with the V716M variant in Italian proband 1, which was determined based on genotyping of 2 relatives of the proband (see Supporting Information Fig. 2). *Frequencies in controls of the bolded alleles (in *a*).

proportion in controls (0.6%), supporting the notion that the missense change should be viewed as a harmless polymorphism. Our haplotype data (see below) suggest that V716M is an ancestral founder variant. Such changes are known to vary in frequency between populations; hence, minor biases in the choice of cases and controls may underlie the differences reported.

As quoted above, several extensive studies have now concluded that V716M is clinically innocent. However, at the time we detected V716M in a young proband with endometrial cancer[12] and a family history suggesting LS, the benign nature of V716M had not been definitively established. After the patient and her family had been counseled and tested, it became clear that V716M segregated with the disease in the family as shown in the pedigree (Supporting Information Fig. 3). Later a more penetrant splice site mutation was found in one member of the family that had simply been missed when the proband's DNA had been sequenced. Given the frequent association of V716M with the splice site mutation in the US, we suggest that the sequencing results for intron 7 be carefully evaluated in individuals displaying V716M.

Haplotype analyses and mutation age estimation

We genotyped 3 SNPs and 14 microsatellite markers in carriers of both the splice site mutation and the V716M variant, and constructed haplotypes by the PHASE method[31,32] in mutation carriers (Fig. 1 and Supporting Information Fig. 2). Based on data from one individual each from 11 unrelated American families carrying both sequence changes (Fig. 1*a*) a shared haplotype was found which covers a genomic region of some 4.8 Mb around *MLH1*. Based on these data the DMLE+2.3 program predicts an age of the splice site mutation of some 18 generations, which is the equivalent of around 450 years (confidence interval 340–585 years) since founding (Fig. 2*a*).

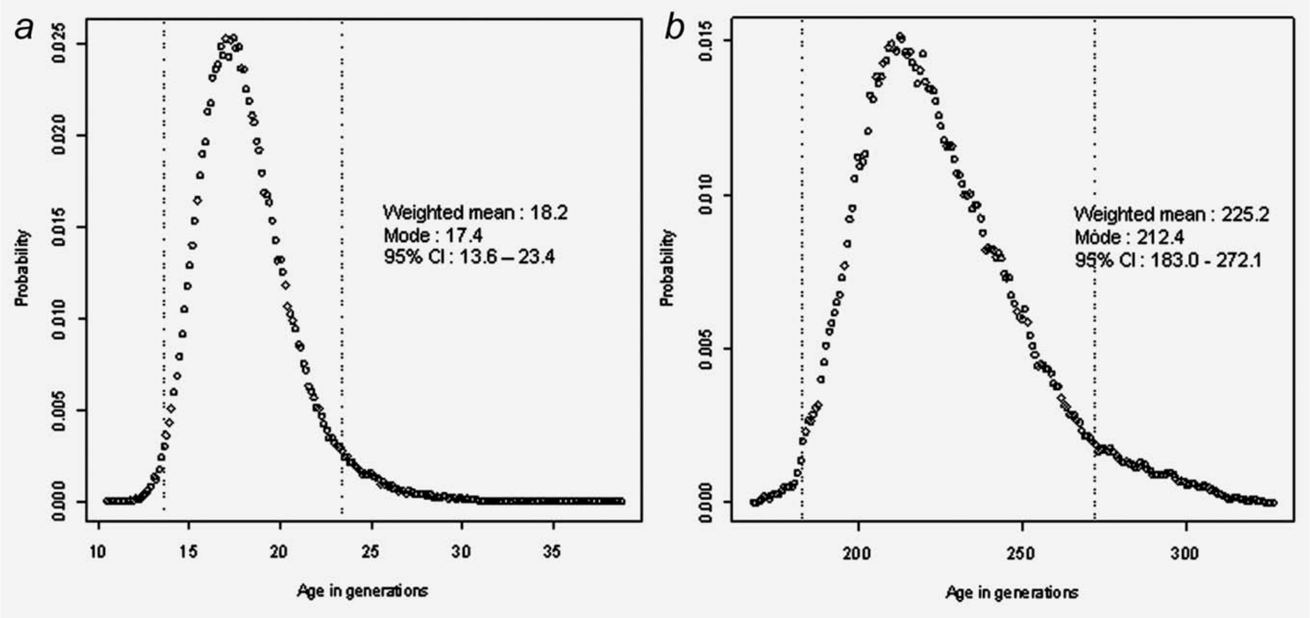


Figure 2. Age estimate for the *MLH1* variants. (*a*) Age estimate for splice site founder mutation in the US population. The posterior probability distribution plot of the mutation age (in generations), as estimated by the software DMLE+2.3 is shown when a population growth rate of 1.65-fold per generation (25 y) is assumed. The dotted lines show the 95% CIs. (*b*) Age estimate for the V716M variant in Europeans. The posterior probability distribution plot of the mutation age (in generations), as estimated by the software DMLE+2.3 is shown when a population growth rate of 1.05-fold per generation (25 y) is assumed. The dotted lines show the 95% CI.

Based on data from one individual in each of the three Italian families with the splice site mutation (without V716M) an unequivocally determined shared haplotype (Fig. 1*b*) suggests a founder mechanism, and the haplotype is entirely different from that of the American cases (Fig. 1*a*). It is also smaller, comprising some 2.2 Mb. These data imply that the splice site mutation arose at least twice, once in an early American immigrant (or in an ancestor of an immigrant) and once elsewhere (perhaps Italy) a long time ago.

All of the US, Italian and German carriers of V716M without the splice site change (Fig. 1*c*) share a short haplotype comprising some 280 kb. Importantly, this haplotype is identical with the central part of the haplotype seen in US carriers of both the splice site mutation and V716M (Fig. 1*a*). These haplotypes suggest the possibility that the V716M in all or most cases represents a single, ancient mutational event. Moreover, the splice site mutation seen in the US, but not the one seen in Italy, arose more recently in a chromosome carrying the ancestral V716M. Under this assumption we performed the age calculation for the V716M variant for the group of Italian and German carriers of the V716M and the age was estimated to be some 225 generations (95% CI: 183–272) or around 5,600 years (Fig. 2*b*).

Applying another method, the Estiage program (Ref. 17; see Material and Methods), we estimated the age of the most recent common ancestor carrying the V716M variant to be some 219 generations (95% CI: 152–317). This is similar to the age of the mutation estimate obtained using DMLE+2.3 program but with a wider confidence interval. We were not able to apply the Estiage program to estimate the age of splice site mutation reliably due to low density of genotyped markers in the larger shared region (see Material and Methods).

Examples of other *MLH1* founder mutations are the two described in Finland[13,33] that together account for up to 50% of all LS in Eastern Finland. Another example is the one-base pair insertion in exon 13 of *MLH1* (c.1489_1490insC) which has been seen frequently in Germany, and a few times in surrounding countries, but not elsewhere. The authors suggested a founder mutation.[34] As the founding of the US mutation is relatively recent its proportion of all LS is modest. We nevertheless consider it worthwhile to suggest screening cancer patients with immunohistochemical loss of staining for MLH1 for the splice site mutation as a first or early step in the mutation detection procedure.

In our study, we estimated the age of both changes. These estimates tend to have broad margins of error based on what parameters are used in the calculation (see Materials and Methods). We are confident that the age of the splice site mutation is considerably younger than the age of the V716M variant based on the size of the shared haplotype and as a result of mutation age estimation. However, the age for the innocent variant could be significantly younger if the frequency of the mutation in the population turns out to be smaller (Supporting Information Fig. 4).

Finally, it is worth considering whether the presence of V716M might predispose to the splice site change. Since both changes are rare, the *a priori* likelihood of them being found together is small, but the absence of the V716M variant in Italian cases with the splice site mutation speaks (weakly) against this hypothesis.

Acknowledgements

The authors thank S. Tanner for helpful discussions, W. Li and J. Panescu for technical support, R. Boland for information, S. Sun for help with DMLE software and S. Girlando for providing samples.

References

1. Desai DC, Lockman JC, Chadwick RB, Gao X, Percesepe A, Evans DGR, Miyaki M, Yuen ST, Radice P, Maher ER, Wright FA, de la Chapelle A. Recurrent germline mutation in MSH2 arises frequently de novo. *J Med Genet* 2000;37:646–52.
2. Froggatt NJ, Green J, Brassett C, Evans DG, Bishop DT, Kolodner R, Maher ER. A common MSH2 mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer. *J Med Genet* 1999;36:97–102.
3. Foulkes WD, Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C, Shia J, Markowitz A, Figer A, Friedman E, Farber D, Greenwood CMT, et al. The founder mutation MSH2*1906G→C is an important cause of Hereditary Non-Polyposis Colorectal Cancer in the Ashkenazi Jewish population. *Am J Hum Genet* 2002;71:1395–412.
4. 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, et al. Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and charactrization of an American founder genomic deletion of the MSH2 gene. *Am J Hum Genet* 2003;72: 1088–100.
5. Lynch HT, Coronel SM, Okimoto R, Hampel H, Sweet K, Lynch JF, Barrows A, Wijnen J, van der Klift H, Franken P, Wagner A, Fodde R, et al. A founder mutation of the MSH2 gene and hereditary nonpolyposis colorectal cancer in the United States. *JAMA* 2004;291:718–24.
6. Clendenning M, Baze MD, Sun S, Walsh K, Liyanarachchi S, Fix D, Schunemann V, Comeras I, Deacon M, Lynch JF, Gong G, Thomas BC, et al. Origins and prevalence of the American founder mutation of MSH2. *Cancer Res* 2008;68:2145–53.
7. Genuardi M, Carrara S, Anti M, Ponz de Leòn M, Viel A. Assessment of pathogenicity criteria for constitutional missense mutations of the hereditary nonpolyposis colorectal cancer genes MLH1 and MSH2. *Eur J Hum Genet* 1999; 7:778–82.
8. Viel A, Genuardi M, Lucci-Cordisco E, Capozzi E, Rovella V, Fornasarig M, Ponz de Leòn M, Anti M, Pedroni M, Bellacosa A, Percesepe A, Covino M, et al. Hereditary nonpolyposis colorectal cancer: an approach to the selection of candidates to genetic testing based on clinical and molecular characteristics. *Community Genet* 1998;1:229–36.
9. Goecke T, Schulmann K, Engel C, Holinski-Feder E, Pagenstecher C, Schackert HK, Kloor M, Kunstmann E, Vogelsang H, Keller G, Dietmaier W, Mangold E, et al. Genotype-phenotype comparison of German MLH1 and MSH2 mutation carriers clinically affected with Lynch syndrome: a report by the German HNPCC Consortium. *J Clin Oncol* 2006;24: 4285–92.
10. Luce MC, Marra G, Chauhan DP, Laghi L, Carethers JM, Cherian SP, Hawn M, Binnie CG, Kam-Morgan LN, Cayouette MC. In vitro transcription/translation assay for the screening of hMLH1 and hMSH2 mutations in familial colon cancer. *Gastroenterology* 1995;109:1368–74.
11. Syngal S, Fox EA, Li C, Dovidio M, Eng C, Kolodner RD, Garber JE. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing. *JAMA* 1999;282:247–53.
12. Hampel H, Frankel W, Panescu J, Lockman J, Sotamaa K, Fix D, Comeras I, La Jeunesse J, Nakagawa H, Westman JA, Prior TW, Clendenning M, et al. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res* 2006;66:7810–7.
13. Moisio AL, Sistonen P, Weissenbach J, de la Chapelle A, Peltomäki P. Age and origin of two common MLH1 mutations predisposing to hereditary colon cancer. *Am J Hum Genet* 1996;59:1243–51.
14. Reeve JP, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* 2002;18:894–5.
15. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* 1999;16:1791–8.
16. Clendenning M, Senter L, Hampel H, Robinson KL, Sun S, Buchanan D, Walsh MD, Nilbert M, Green J, Potter J, Lindblom A, de la Chapelle A. A frame-shift mutation of PMS2 is a widespread cause of Lynch syndrome. *J Med Genet* 2008;45:340–5.
17. Genin E, Tullio-Pelet A, Begeot F, Lyonnet S, Abel L. Estimating the age of rare disease mutations: the example of Triple-A syndrome. *J Med Genet* 2004;41:445–9.
18. Hampel H, Panescu J, Lockman J, Sotamaa K, Fix D, Comeras I, LaJeunesse J, Nakagawa H, Westman JA, Prior TW, Clendenning M, de la Chapelle A, et al. Comment on: screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res* 2007;67:9603.
19. Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Clendenning M, Sotamaa K, Prior T, Westman J, Panescu J, Fix D, et al. Feasibility of screening for Lynch syndrome among colorectal cancer patients. *J Clin Oncol* 2008;26:5783–8.
20. 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, et al. Conversion of diploidy to haploidy. *Nature* 2000;403: 723–4.
21. Ponz de Leon M, Benatti P, Di Gregorio C, Pedroni M, Losi L, Genuardi M, Viel A, Fornasarig M, Lucci-Cordisco E, Anti M, Ponti G, Borghi F, et al. Genetic testing among high-risk individuals in families with hereditary nonpolyposis colorectal cancer. *Br J Cancer* 2004;90:882–7.
22. Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lönnqvist KE, Holinski-Feder E, Sutter C, McKinnon W, Duraisamy S, Gerdes A-M, Peltomäki P, et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology* 2005;129:537–49.
23. 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: 4595–604.
24. Martinez SL, Kolodner RD. Functional analysis of human mismatch repair gene mutations identifies weak alleles and polymorphisms capable of polygenic interactions. *Proc Natl Acad Sci U S A* 2010;107:5070–5.
25. Drost M, Zonneveld JB, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind N. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Hum Mutat* 2010;31:247–53.
26. Pastrello C, Pin E, Marroni F, Bedin C, Fornasarig M, Tibiletti MG, Oliani C, Ponz

de Leon M, Urso ED, Della Puppa L, Agostini M, Viel A. Integrated analysis of unclassified variants in mismatch repair genes. *Genet Med* 2011;13:115–24.

27. Kowalski LD, Mutch DG, Herzog TJ, Rader JS, Goodfellow PJ. Mutational analysis of MLH1 and MSH2 in 25 prospectively-acquired RER+ endometrial cancers. *Genes Chromosomes Cancer* 1997; 18:219–27.
28. Dieumegard B, Grandjouan S, Sabourin JC, Le Bihan ML, Lefrère I, Bellefqih, Pignon JP, Rougier P, Lasser P, Bénard J, Couturier D, Bressac-de Paillerets B. Extensive molecular screening for hereditary non-polyposis colorectal cancer. *Br J Cancer* 2000;82:871–80.
29. Offer SM, Pan-Hammarström Q, Hammarström L, Harris RS. Unique DNA repair gene variations and potential associations with the primary antibody deficiency syndromes IgAD and CVID. *PLoS One* 2010;5:e12260.
30. Barnetson RA, Cartwright N, van Vliet A, Haq N, Drew K, Farrington S, Williams N, Warner J, Campbell H, Porteous ME, Dunlop MG. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. *Hum Mutat* 2008;29: 367–74.
31. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68: 978–89.
32. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003;73: 1162–9.
33. Nyström-Lahti M, Kristo P, Nicolaides NC, Chang SY, Aaltonen LA, Moisio AL, Järvinen HJ, Mecklin JP, Kinzler KW, Vogelstein B. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat Med* 1995;1: 1203–6.
34. Mangold E, Pagenstecher C, Friedl W, Mathiak M, Buettner R, Engel C, Loeffler M, Holinski-Feder E, Müller-Koch Y, Keller G, Schackert HK, Krüger S, et al. Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. *Int J Cancer* 2005;116:692–702.