

## MLH1 Founder Mutations with Moderate Penetrance in Spanish Lynch Syndrome Families

Ester Borràs<sup>1</sup>, Marta Pineda<sup>1</sup>, Ignacio Blanco<sup>8</sup>, Ethan M. Jewett<sup>10</sup>, Fei Wang<sup>12</sup>, Àlex Teulé<sup>2</sup>, Trinidad Caldés<sup>15</sup>, Miguel Urioste<sup>16,17</sup>, Cristina Martínez-Bouzas<sup>18</sup>, Joan Brunet<sup>9</sup>, Judith Balmaña<sup>3</sup>, Asunción Torres<sup>19</sup>, Teresa Ramón y Cajal<sup>4</sup>, Judit Sanz<sup>5</sup>, Lucía Pérez-Cabornero<sup>20</sup>, Sergi Castellví-Bel<sup>6</sup>, Àngel Alonso<sup>21</sup>, Àngel Lanas<sup>22</sup>, Sara González<sup>1</sup>, Victor Moreno<sup>7</sup>, Stephen B. Gruber<sup>13,14</sup>, Noah A. Rosenberg<sup>10,11</sup>, Bhramar Mukherjee<sup>12</sup>, Conxi Lázaro<sup>1</sup>, and Gabriel Capellá<sup>1</sup>

### Abstract

The variants c.306+5G>A and c.1865T>A (p.Leu622His) of the DNA repair gene *MLH1* occur frequently in Spanish Lynch syndrome families. To understand their ancestral history and clinical effect, we performed functional assays and a penetrance analysis and studied their genetic and geographic origins. Detailed family histories were taken from 29 carrier families. Functional analysis included *in silico* and *in vitro* assays at the RNA and protein levels. Penetrance was calculated using a modified segregation analysis adjusted for ascertainment. Founder effects were evaluated by haplotype analysis. The identified *MLH1* c.306+5G>A and c.1865T>A (p.Leu622His) variants are absent in control populations and segregate with the disease. Tumors from carriers of both variants show microsatellite instability and loss of expression of the *MLH1* protein. The c.306+5G>A variant is a pathogenic mutation affecting mRNA processing. The c.1865T>A (p.Leu622His) variant causes defects in *MLH1* expression and stability. For both mutations, the estimated penetrance is moderate (age-cumulative colorectal cancer risk by age 70 of 20.1% and 14.1% for c.306+5G>A and of 6.8% and 7.3% for c.1865T>A in men and women carriers, respectively) in the lower range of variability estimated for other pathogenic Spanish *MLH1* mutations. A common haplotype was associated with each of the identified mutations, confirming their founder origin. The ages of c.306+5G>A and c.1865T>A mutations were estimated to be 53 to 122 and 12 to 22 generations, respectively. Our results confirm the pathogenicity, moderate penetrance, and founder origin of the *MLH1* c.306+5G>A and c.1865T>A mutations. These findings have important implications for genetic counseling and molecular diagnosis of Lynch syndrome. *Cancer Res*; 70(19); 7379–91. ©2010 AACR.

### Introduction

Lynch syndrome (MIM 120435) is an autosomal-dominant condition caused by germline mutations in mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, and *PMS2* (1). It is characterized by early-onset colorectal cancer (CRC) and an increased risk of other cancers (1). At age 70, cumulative CRC risk has been estimated to be 27% to 66% for men

and 22% to 47% for women; risk for endometrial cancer (EC) varies between 14% and 40% (2–5).

Several founder mutations have been described in MMR genes (6). In *MLH1*, founder mutations were identified in several populations (6–8), where they can explain a substantial fraction of Lynch syndrome occurrences (9, 10), facilitating genetic diagnosis. In Spain, founder mutations in MMR genes have only been identified in the *MSH2* gene (11, 12).

**Authors' Affiliations:** <sup>1</sup>Laboratori de Recerca Translacional, Institut Català d'Oncologia, IDIBELL, Hospitalet de Llobregat; <sup>2</sup>Programa de Consell Genètic en Càncer, Institut Català d'Oncologia, Hospital Duran i Reynals, Hospitalet de Llobregat; <sup>3</sup>Servei d'Oncologia Mèdica, Hospital Vall d'Hebron; <sup>4</sup>Servei d'Oncologia Mèdica, Hospital de la Santa Creu i Sant Pau; <sup>5</sup>Institut d'Oncologia Corachan-Centre Mèdic; <sup>6</sup>Servei de Gastroenterologia, Institut de Malalties Digestives i Metabòliques, Hospital Clínic; <sup>7</sup>Unitat de Bioinformàtica i Bioestadística, Institut Català d'Oncologia i Universitat de Barcelona, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain; <sup>8</sup>Programa de Consell Genètic en Càncer, Institut Català d'Oncologia, Hospital Germans Trias i Pujol, Badalona, Spain; <sup>9</sup>Programa de Consell Genètic en Càncer, Institut Català d'Oncologia, Hospital Josep Trueta, Girona, Spain; <sup>10</sup>Center for Computational Medicine and Bioinformatics and <sup>11</sup>Department of Human Genetics, and the Life Sciences Institute, University of Michigan; <sup>12</sup>Department of Internal Medicine, University of Michigan Medical School and School of Public Health; <sup>13</sup>Departments of Internal Medicine and Human Genetics, University of Michigan Medical School; <sup>14</sup>Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan; <sup>15</sup>Laboratorio de Oncologia Molecular, Hospital Clínico San Carlos;

<sup>16</sup>Departamento de Genética Humana, Centro Nacional de Investigaciones Oncológicas; <sup>17</sup>Centro de Investigación en Red de Enfermedades Raras, Madrid, Spain; <sup>18</sup>Laboratorio de Genética Molecular, Hospital de Cruces, Bizkaia, Spain; <sup>19</sup>Unitat de Consell Genètic, Hospital Universitari Sant Joan, Reus, Spain; <sup>20</sup>Instituto de Biología y Genética Molecular, Valladolid, Spain; <sup>21</sup>Servicio de Genética, Hospital Virgen del Camino, Pamplona, Spain; and <sup>22</sup>Servicio de Enfermedades Digestivas Hospital Universitario, Universidad de Zaragoza, Zaragoza, Spain

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

E. Borràs and M. Pineda contributed equally to this work.

**Corresponding Author:** Gabriel Capellá, Gran Via 199-203, 08907 L'Hospitalet de Llobregat, Barcelona, Spain. Phone: 34-932607952; Fax: 34-942607466; E-mail: gcapella@iconcologia.net.

doi: 10.1158/0008-5472.CAN-10-0570

©2010 American Association for Cancer Research.

MMR founder mutations reported thus far are heterogeneous in function and frequency (6). Most of them encode truncated proteins and are readily categorized as pathogenic mutations. However, founder MMR variants with uncertain pathogenicity have also been identified (11, 13, 14). To investigate the functional effect of MMR variants, numerous assays at the RNA and protein level have been developed (15, 16). The classification of a variant as a disease-causing mutation requires integration of data from different sources, including its frequency in control populations, its cosegregation with the disease in families, its clinicopathologic features, and its properties in functional studies (17).

We have identified and characterized two frequently occurring *MLH1* variants in the Spanish population. To address their clinical significance, we examined their penetrance and performed functional studies. The founder effect hypothesis was further explored by haplotype analysis.

## Materials and Methods

### Patients and samples

A total of 57 families were included in this study. Twenty-nine families were carriers of c.306+5G>A or c.1865T>A *MLH1* variants (17 and 12 families, respectively) and were assessed through Spanish genetic counseling and molecular diagnostic laboratories. Twenty-eight families were carriers of other pathogenic *MLH1* germline mutations assessed at the Catalan Institute of Oncology (ICO). Detailed family histories from at least three generations and geographic origins were obtained. Clinical data collection included tumor location, age at diagnosis, microsatellite instability (MSI) testing, and immunohistochemistry of MMR proteins in tumors. Genealogic investigations based on interviews failed to identify relationships between individuals from different families. The study protocol was approved by the Human Research Ethics Committees of participating centers, and informed consent was obtained from all subjects evaluated.

Genomic DNA was extracted from whole blood using the FlexiGene DNA kit (Qiagen) or from formalin-fixed, paraffin-embedded tissues using the QIAamp DNA Mini kit (Qiagen). DNA samples from 325 cases and 309 controls in a hospital-based CRC case-control study conducted to assess gene-environment interactions in relation to CRC risk (18) were used.

### Screening for the c.306+5G>A and c.1865T>A *MLH1* mutations and loss of heterozygosity analysis

Screening for the c.306+5G>A and c.1865T>A *MLH1* mutations was performed by conformation-sensitive capillary electrophoresis (CSCE) in DNA samples from the same CRC case-control study (Supplementary Table S1; conditions available on request; ref. 18). DNAs with patterns differing from those of controls were amplified using nonfluorescent primer and sequenced with BigDye Terminator v.3.1 Cycle Sequencing kit (Applied Biosystems). The methods for analysis of loss of heterozygosity (LOH) of the variant c.1865T>A of *MLH1* in tumors are included in Supplementary Methods.

### Computational methods

The following computational methods were used to predict modifications in splice-sites or exonic splicing enhancer sites: Spliceport (19), NNSplice (20), Rescue\_ESE (21), and ESEfinder (22). The Polyphen (23) algorithm was used to predict the pathogenicity of p.Leu622His variant using default settings. The Clustalw program was used to align the amino acid sequence of *MLH1* in 13 phylogenetically diverse species.

### Effect of the c.306+5G>A and c.1865T>A *MLH1* mutations on mRNA processing and stability

Total RNA was extracted from peripheral blood lymphocytes using Trizol (Invitrogen), and cDNA was synthesized using random primers (Invitrogen) and SuperScript II reverse transcriptase (Invitrogen). Specific primers were used to amplify the appropriate *MLH1* coding region (Supplementary Table S1). The methods for allele-specific expression (ASE) analysis of the variant c.1865T>A of *MLH1* are described in Supplementary Methods.

### Site-directed mutagenesis and construction of p.Leu622His-*MLH1*

The pcDNA3.1+ vector containing wild-type (WT) *MLH1* cDNA (Genbank accession no. NM\_000249.2) was kindly provided by Dr. R. Kolodner (Ludwig Institute for Cancer Research, UC San Diego School of Medicine, La Jolla, USA). The mutant p.Leu622His-*MLH1* cDNA was constructed using the QuikChange Site-Directed Mutagenesis kit (Stratagene) with primer 5'-AAGAAGAAGGCTGAGATGC(A)TG CAGAC-TATTTCTCTTTG-3' (sense strand; mutagenesis site is in parentheses). Sequencing was used to verify the presence of the mutation. The cDNA insert between *Xho*I and *Bam*HI sites was subcloned into WT *MLH1*-pcDNA3.

### Cell culture and transfection

HCT116 [deficient for endogenous *MLH1* (24)] and 293 cells were obtained from the American Type Culture Collection, resuscitated from stocks frozen at low passage within 6 months of purchase, and cultured as described (25, 26). Cell lines were routinely tested by *Mycoplasma* presence and verified by morphology, growth curve analysis, and expression of proteins (e.g., *MLH1* and *PMS2*). HCT116 cells were transfected with pGFP (green fluorescent protein) and WT *MLH1*-pcDNA3.1 or p.Leu622His-*MLH1*-pcDNA3.1 vector using Lipofectamine 2000 (Invitrogen) and Plus (Invitrogen). Transfection efficiency was measured by cytometry 24 hours after transfection.

### *MLH1* and *PMS2* protein expression and stability experiments

Protein extraction from peripheral blood lymphocytes and HCT116 and 293 cells was performed as described elsewhere (27). *MLH1* and *PMS2* expression levels were examined by SDS-PAGE, followed by Western blotting analysis with anti-*MLH1* and anti-*PMS2* antibodies (clones G168-15 and A16-4; BD Biosciences).  $\beta$ -Tubulin antibody (Sigma) was used to assess equal loading in all lanes. Band intensities were quantified using Quantity One software (Bio-Rad), and expression

of *MLH1* and *PMS2* was normalized to  $\beta$ -tubulin expression. The decrease of protein expression was calculated by dividing the normalized protein expression in p.Leu622His-transfected cells by the expression in WT *MLH1*-transfected cells. The stability of WT and variant *MLH1* was assessed by treating cells with cycloheximide, a global inhibitor of *de novo* protein synthesis, as described (25).

### Estimation of age-specific cumulative risk

We used information on the occurrence of CRC and EC in relatives of MMR mutation-positive index cases to estimate age- and gender-specific incidences of CRC and EC (females) in MMR mutation carriers by maximum likelihood using modified segregation analysis (2). The method was implemented in MENDEL (v3.3.5; refs. 28, 29).

Relatives were assumed to be followed from 20 years of age and to be censored at the age of CRC diagnosis, at the age of death, at the age of last follow-up, or at age 70 years, whichever occurred earliest. We did not ignore the cases beyond age 70 but assigned them the same risk at age 70 to avoid the larger variances observed when only sparse data are available. In estimating the risk of CRC and EC, female relatives were censored at age of CRC or EC diagnosis, whichever was diagnosed first. Information on MMR mutation status in relatives was included whenever available. For individuals with missing age information, the age was imputed based on relationship with the proband, age of the proband, and deceased status at last follow-up (dead or alive).

To correct for ascertainment, we maximized the conditional likelihood of observing the phenotypes (CRC and/or EC) and genotypes (c.306+5G>A or c.1865T>A *MLH1* variants) of the entire pedigree given the phenotypic and genotypic information of the index case (proband) and phenotypic information of the pedigree to account for multiplex ascertainment bias. Age-specific cumulative risks (penetrance) and hazard ratio (HR) estimates of CRC and EC risks in families with MMR gene variants were calculated assuming a proportional hazards model, with  $\lambda(t) = \lambda_0(t)\exp[g(t)]$ , where  $\lambda_0(t)$  is the background incidence. They were compared with the risks in the general population, which were assumed to follow the population incidence from the Spanish Cancer Registries (30). The age-specific relative risks in carriers compared with the population rates are modeled through the function  $\exp[g(t)]$ . We estimated the age-specific log HR parameters for the two age intervals <50 and  $\geq 50$ , assuming that  $g(t) = \sum_{k=1}^n \exp[\beta_k]$ , a piecewise constant HR in the  $k^{\text{th}}$  age band  $k = 1, 2, \dots, n$ . In all analysis, cancer incidences in noncarriers were assumed to follow the population cohort-specific rates as obtained through the Spanish Cancer Registries.

To construct confidence intervals (CI) for the log(HR) estimates, we assumed that the maximum likelihood estimates of the parameters were asymptotically normally distributed with covariance matrix given by the inverse of the Fisher information matrix. Cumulative risk (e.g., penetrance) and its 95% CI were calculated from the cumulative incidence  $\Lambda(t)$  given by  $\Lambda(t) = \sum_{k=1}^n i_k t_k \exp[\beta_k]$ , where  $i_k$  is the population incidence obtained from the Spanish Cancer

Registries,  $t_k$  is the length of the  $k^{\text{th}}$  age interval, and  $\beta_k$  is the log(HR) in the  $k^{\text{th}}$  age interval  $k = 1, 2$ . The cumulative risk is given by  $F(t) = 1 - \exp[-\Lambda(t)]$  and a 95% CI for  $F(t)$  is  $1 - \exp[-\Lambda(t) \pm 1.96\sqrt{\text{Var}(\Lambda(t))}]$ , where

$$\begin{aligned} \text{Var}(\Lambda(t)) = & \sum_{k=1}^n i_k^2 t_k^2 \text{Var}(\beta_k) \exp[2\beta_k] \\ & + 2 \sum_{j < k, k=1}^n i_k i_j t_k t_j [\text{Var}(\beta_k) \text{Var}(\beta_j)]^{1/2} \exp(\beta_k \\ & + \beta_j) \text{corr}(\beta_k, \beta_j). \end{aligned}$$

Wald-type tests and CIs were then used based on the point estimate and the estimated SEs (28, 29).

The analysis accounted for both genotyped and ungenotyped relatives. Missing genotype information was handled by including the allele frequency as a parameter in the likelihood and then maximizing the marginal likelihood of the phenotype and genotype data of the entire pedigree, summing over all possible configurations of the unobserved genotype matrix, given the observed genotypes. Cumulative risks and HR for CRC and EC were estimated separately for males and females; risk of EC was estimated only for females.

### Haplotype analysis

Haplotype analysis was performed using three *MLH1* single-nucleotide polymorphisms and seven microsatellite markers (Supplementary Table S1; conditions available on request). One hundred and twenty-two DNA samples from members of the studied families and 50 control individuals randomly selected from the CRC case-control study (18) were analyzed, including individuals that come from the areas of origin of the founders. To deduce the mutation-associated haplotype, intrafamilial segregation analysis was performed under the assumption that the number of crossovers between adjacent markers was minimal. The frequency of disease haplotypes in the control population was estimated using PHASE2 (default settings; ref. 31).

### Estimation of founder mutation age

We used a modification of the method of Schroeder and colleagues (32) to estimate the time to the most recent common ancestor (TMRCA) separately for the c.306+5G>A and c.1865T>A alleles. This method uses a count of the number of recombination events that have occurred on copies of the ancestral mutant haplotype, together with an estimate of the recombination map length of the haplotype, to estimate the total time length of the genealogy of sampled copies of a mutant allele in units of generations. In our application of the approach (32), to estimate TMRCA from the genealogy length, we relied on a multiplier  $c(n)$ , a ratio of TMRCA to tree length that was estimated from coalescent simulations with sample size  $n$ . The TMRCA estimate was then taken as an estimate of the age of the mutation, and CIs were obtained by considering the uncertainty in converting the length estimate for the genealogy into an estimate of TMRCA. Because analysis of family

**Table 1.** Clinical features of affected carriers

Family	Referral center	Criteria met	Affected carrier	Gender	Tumor type	Age of onset	MSI	IHC			CRC location	CRC stage: TNM/Dukes
								MLH1	MSH2	MSH6		
A. c.306+5G>A families												
A1	ICO, Barcelona	AC	A1.1	F	CRC	38	+	-	+	+	R	T4N0M0
			A1.2	F	CRC	41					R	T2N0M0
			A1.3	M	CRC	69	+	-	+	+	R	T3N0M0
A2	ICO, Barcelona	BC	A2.1	F	CRC	28	+	-	+	+	Rc	T2N0M0
			A2.2	M	CRC	51	+	-	+	+	R	T3N0M0
A3	ICO, Barcelona	BC	A3.1	M	CRC	38						
			A3.8	M	CRC	52					R	T2N0M0
			A3.10	M	CRC	52						
					GC	70						
A4	ICO, Barcelona	AC	A4.1	F	CRC	38	+	NV	+	+	R	T3N1Mx
			A4.2	F	CRC	60	+	NV	+	+	R	T4NxMx
A5	ICO, Barcelona	BC	A5.1	F	CRC	50	+	-	+	+	R	T2NxM0
					DC	68						
					CRC	80					Rc	T3N0M0
A6	ICO, Barcelona	BC	A6.1	F	CRC	38	+	NV	+	NV	L	T4N1Mx
			A6.2	F	CRC	49					L	T2N0Mx
A7	ICO, Barcelona	BC	A7.1	F	CRC	52	+	-	+	+	R	T1N0M0
A8	HCSC, Madrid/ HULB, Zaragoza	BC	A8.1	F	BC*	48						
					3CRC	50	+	NV	+	+	R,R,L	B1,B1,B2
A9	HCSC, Madrid	AC	A9.1	F	CRC	20	+	-			L	T3N2Mx
A10	CNIO, Madrid	BC	A10.1	F	CRC	30	+				R	B1
A11	H. Cruces, Bizkaia	BC	A11.1	M	CRC	59					R	T2N0M0
			A11.2	M	CRC	48					L	T3N0M0
A12	H. Cruces, Bizkaia	BC	A12.1	F	EC	51						
					TC*	65						
A13	ICO, Barcelona	AC	A13.1	F	EC	42						
					2CRC	42		-	+	+	R,R	T3N0M0, T3N0M0
A14	ICO, Barcelona	AC	A14.2	M	CRC	45						
					CRC	49					L	
A15	IBGM, Valladolid	AC	A15.1	F	CRC	27	+	-	+	+	Rc	T4N0M0
			A15.2	F	CRC	50					R	
			A15.8	M	CRC	39					R	T3N0Mx
A16	IBGM, Valladolid	AC	A16.1	M	CRC	59	+	-	+	+	R	
A17	H. Clinic, Barcelona/HVC Pamplona	AC	A17.10	M	CRC	58	+	-	+	+	R	B2
			A17.11	M	CRC	35		-	+	+	R	C
B. c.1865T>A families												
B1	ICO, Barcelona	BC	B1.1	M	CRC	27	+	-	+	+	R	T3N0M0
					CRC	39					L	T1N0M0
B2	H. St Joan, Reus	AC	B2.1	M	2CRC	56					R,L	T1N0M0
			B2.2	M	CRC	24	+	-	+	NV	L	T3N0M0
B3	H. St Pau, Barcelona	AC	B3.1	F	CRC	43	+	-	+	NV	R	T3N0M0
			B3.2	F	CRC	28					L	T3M1N0
B4	ICO, Barcelona	AC	B4.1	F	CRC	33	+	-	+	NV	R	T3N2Mx
			B4.2	M	CRC	42					R	T3N0M0
					OC*	58						
B5	ICO, Girona	AC	B5.1	M	CRC	42	+	-	+	+	R	T3N0M0

(Continued on the following page)

**Table 1.** Clinical features of affected carriers (Cont'd)

Family	Referral center	Criteria met	Affected carrier	Gender	Tumor type	Age of onset	MSI	IHC			CRC location	CRC stage: TNM/Dukes	
								MLH1	MSH2	MSH6			
B6	HCSC, Madrid	AC	B6.1	F	CRC	38	+	-	+	+	R	D	
			B6.2	F	EC	53	+	-	+	+	L	B	
			B6.3	F	CRC	65							
			B6.4	M	EC	55	+	-	+	+	R	T3N2Mx	
B7	HCSC, Madrid	BC	B7.1	M	CRC	59	+	-	+	+			
B8	IOC, Barcelona	BC	B8.2	F	CRC	43		-	+	+			
B9	HVH, Barcelona	BC	B9.1	M	CRC	55	+	-	+	+	R	B	
B10	ICO, Barcelona	AC	B10.1	F	EC	54	+	NV	+	+			
B11	CNIO, Madrid	AC	B11.1	M	CRC	56	+				R	A	
B12	CNIO, Madrid	AC	B12.5	F	EC	38							
					BC*	79							
			B12.7	F	CRC	48	+				R	A	
			B12.8	F	2CRC	51					R	A	

NOTE: A. Clinical features of affected carriers of variant c.306+5G>A of *MLH1*. B. Clinical features of affected carriers of variant c.1865T>A of *MLH1*.

Abbreviations: ICO, Institut Català d'Oncologia; HCSC, Hospital Clínico San Carlos; CNIO, Centro Nacional de Investigaciones Oncológicas; H. Cruces, Hospital de Cruces; HVH, Hospital Vall d'Hebron; H. St Joan, Hospital Universitari Sant Joan; H. St Pau, Hospital de la Santa Creu i Sant Pau; IDOC, Institut d'Oncologia Corachan; IBGM, Instituto de Biología y Genética Molecular; H. Clinic, Hospital Clinic; HVC, Hospital Virgen del Camino; HULB, Hospital Universitario Lozano Blesa; AC, Amsterdam criteria; BC, Bethesda criteria; M, male; F, female; BC, breast cancer; CRC, colorectal cancer; DC, duodenal cancer; EC, endometrial cancer; GC, gastric cancer; OC, otorhinolaryngological cancer; TC, thyroid cancer; IHC, immunohistochemical analysis of MMR proteins in tumor tissue; NV, not evaluable result; R, right; L, left; Rc, rectum.

\*Tumor type not included in Lynch syndrome tumor spectrum.

history revealed no relationships among sampled families in the most recent three to five generations, our age estimation was performed from a reference point four generations in the past, so that at the end of the estimation process, the estimate was increased by four generations. See Supplementary Materials and Methods for further details.

## Results

### Identification of frequently occurring germline *MLH1* variants

Mutational screening of MMR genes performed in suspected Lynch syndrome patients at ICO led to the identification of two frequently occurring germline *MLH1* variants: c.306+5G>A and c.1865T>A. No other alterations in *MLH1* coding sequence or at exon-intron boundaries were detected in carriers. Of the 83 families with germline MMR alterations identified at ICO, 45 were carriers of an *MLH1* alteration (28 pathogenic mutations and 17 variants of unknown significance). The c.306+5G>A and c.1865T>A variants were identified in nine and four families, constituting 20% and 9% of all *MLH1* carrier families, respectively. An extended collaborative study with other referral centers (see legend of Table 1) enabled the identification of 16 additional carrier families. In all, we identified 17 families with c.306+5G>A and

12 with c.1865T>A *MLH1* variants (Table 1). Neither of these two variants was detected by CSCE in a panel of paired CRC cases and controls, reducing the probability of being polymorphisms or CRC risk alleles.

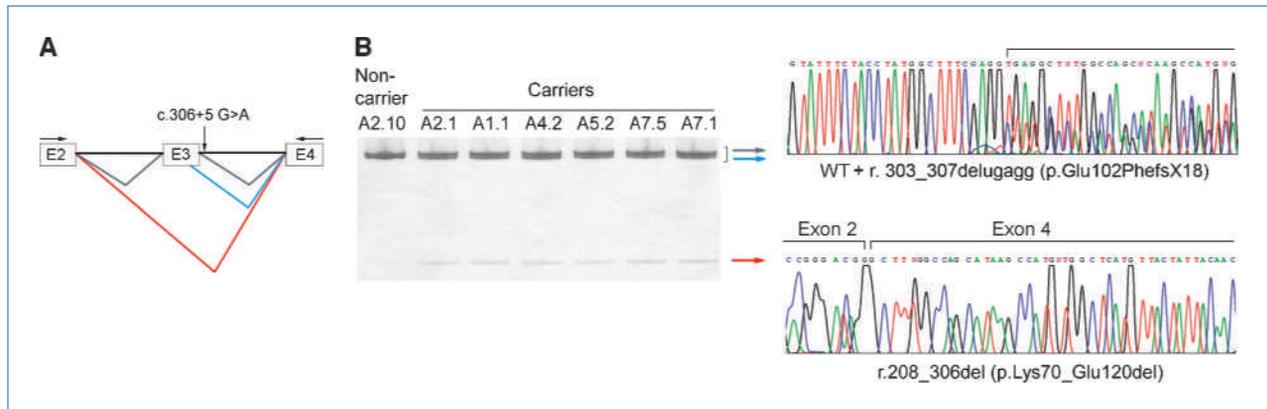
### Familial clinical features

Most of the included families fulfilled the modified Amsterdam criteria (Table 1). The majority of tumors diagnosed in carriers belonged to the Lynch syndrome spectrum. The median age at diagnosis was 49.5 years (range, 20–80) for c.306+5G>A carriers and 45.5 years (range, 24–79) for c.1865T>A carriers. CRC and EC tested were microsatellite unstable (MSI+), and informative cases showed loss of expression of the MLH1 protein (Table 1).

Both c.306+5G>A and c.1865T>A *MLH1* variants cosegregated with cancer in 29 and 21 affected members, respectively, and were absent in 37 and 18 unaffected members, respectively. The c.306+5G>A variant was also identified in 21 unaffected members (median age, 34.0; range, 22–63), and c.1865T>A in 21 unaffecteds (median age, 29.0; range, 18–41).

### Pathogenicity assessment of the c.306+5G>A and c.1865T>A *MLH1* variants

For the c.306+5G>A variant, splicing prediction programs predicted the creation of a new donor site 5 bp upstream.

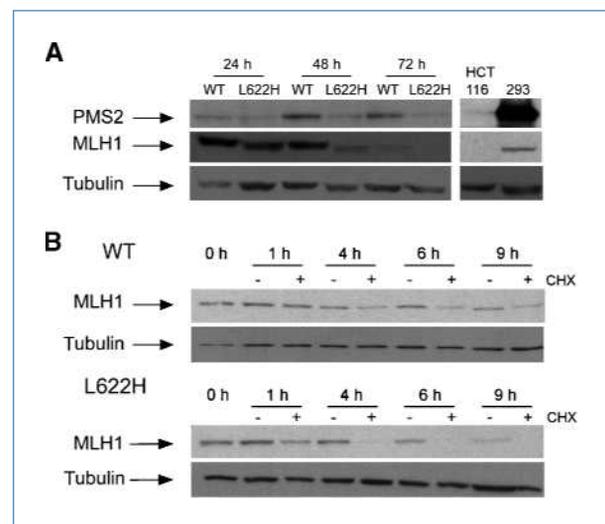


**Figure 1.** Characterization of the c.306+5G>A *MLH1* variant. A, schematic overview of *MLH1* exons 2 to 4 with a representation of the normal and aberrant transcripts caused by the c.306+5G>A mutation. Black arrows represent primers used for RT-PCR amplification. B, acrylamide gel showing RT-PCR products obtained from c.306+5G>A carriers and a noncarrier. On the right, direct sequencing of the RT-PCR products is shown.

Reverse transcription-PCR (RT-PCR) analysis on RNA from lymphocytes of carriers confirmed the generation of this aberrant mRNA transcript, r.303\_307del, expected to generate a truncated protein. This transcript was associated with an increased amount of the r.208\_306del alternative constitutional transcript, corresponding to the in-frame skipping of *MLH1* exon 3 (Fig. 1). Although the variant c.306+5G>A is pathogenic at the RNA level, neither abnormal bands nor differences in *MLH1* protein expression were observed in lymphocytes from c.306+5G>A carriers, as assessed by Western blotting (data not shown).

The *MLH1* c.1865T>A variant is predicted to generate the missense change p.Leu622His, which was classified as pathogenic by the Polyphen algorithm and MAPP-MMR and SIFT predictions (33, 34). The affected amino acid residue is highly conserved and is located at the interaction domain for MutL (the homologue of *MLH1* in *Escherichia coli*; ref. 35). Spliceport and NNSplice did not predict any effect on mRNA processing, whereas ESEfinder and Rescue\_ESE predicted the creation of a new potential binding site for SRp proteins. In lymphocytes of c.1865T>A carriers, differences neither at the mRNA level [as assessed by RT-PCR (data not shown) and allele-specific expression analysis (Supplementary Fig. S1)] nor at the protein expression level (data not shown) were observed. The transient transfection of the p.Leu622His variant into the HCT116 cell line (24), which lacks endogenous *MLH1*, resulted in diminished *MLH1* expression when compared with WT *MLH1*-transfected cells (Fig. 2A). *MLH1* heterodimerizes with PMS2, stabilizing its expression (25). As expected, the diminished p.Leu622His expression was associated with lower PMS2 expression compared with WT *MLH1*-transfected cells (Fig. 2A). After cycloheximide treatment, *MLH1* expression was lower in both p.Leu622His-transfected cells and WT *MLH1*-transfected cells. However, the effect was greater in p.Leu622His-transfected cells, indicating that p.Leu622His protein was less stable than WT *MLH1* (Fig. 2B). Five of the six tumors analyzed from c.1865T>A carriers lost the *MLH1* WT allele (Supplementary Fig. S1), suggesting a growth advantage associated

with loss of the WT protein. Together, these lines of evidence support the notion that the variant c.1865T>A is a pathogenic mutation via a mechanism in which *MLH1* protein expression is decreased.



**Figure 2.** Expression and stability of the p.Leu622His variant of *MLH1* in HCT116 cells. A, time course of expression of WT *MLH1* and p.Leu622His (L622H) variants. Top and middle, cell lysates probed with anti-PMS2 and anti-*MLH1* antibodies; bottom, to verify equal protein loading, cell lysates were probed with variant anti-tubulin antibody. Untransfected HCT116 and 293 cells were used as controls. One representative experiment among three is shown. The ranges of decrease in *MLH1* expression at 24, 48, and 72 h were 35% to 55%, 42% to 75%, and 81% to 100%, respectively. The ranges of decrease in PMS2 expression at 24, 48, and 72 h were 57% to 89%, 48% to 55%, and 43% to 64%, respectively. B, stability of WT *MLH1* and p.Leu622His variant as assessed by cycloheximide (CHX) treatment. "+" and "-" represent treatment with cycloheximide or DMSO as vehicle, respectively. The decrease in expression at 1, 4, 6, and 9 h after cycloheximide treatment was 39%, 64%, 55%, and 61% in WT *MLH1* and 47%, 97%, 99%, and 100% in p.Leu622His, respectively. One representative experiment among three is shown.

**Table 2.** Characteristics of the study population by mutation type

	c.306+5G>A	c.1865T>A	Other <i>MLH1</i> mutations
No. probands/pedigrees	17	12	28
Total no. individuals	514	447	975
No. females	257	204	444
No. males	240	221	486
No. individuals with missing gender information	17	22	45
No. individuals with missing age information	226	208	442
No. first-degree relatives	89	106	193
No. subjects genotyped	88	60	174
No. tested mutation-positive subjects	50	42	88
Percentage of mutation-positive subjects among subjects genotyped	56.8	70.0	50.5
Median age at diagnosis of CRC in men (range)	52 (38–80)	43 (20–65)	41 (25–80)
Median age at diagnosis of CRC in women (range)	50 (20–80)	48 (23–87)	44 (18–77)
Percentage of CRC diagnosed before age 50 among affected subjects	38.3	57.1	73.7
Percentage of CRC or EC affected subjects among carriers	56.0	47.6	64.8
Percentage of CRC or EC affected subjects among noncarriers	2.6	0	0
Percentage of CRC or EC affected subjects among nontested individuals	10.1	12.1	7.5

NOTE: Blood- and nonblood-related relatives are included.

### Cumulative risks and HRs derived from the penetrance analysis

A total of 1,936 individuals were included in the penetrance analysis, 57 of which were probands and 388 of which were first-degree relatives of probands (Table 2). Age-specific cumulative risks of CRC by decade compared with risks in Spanish Cancer Registries (30) are shown in Table 3A. For c.306+5G>A carriers, the risk of CRC by age 50 significantly exceeds the cumulative risk in the general population. Risk of CRC continues to increase, and by age 70, lifetime risk for CRC is estimated at 20.1% (95% CI, 1.4–35.9%) for men and 14.1% for women (95% CI, 1.2–25.2%). The overall HR for CRC is 7.7 (95% CI, 2.9–20.6) for males and 9.0 (95% CI, 3.6–22.6) for females (Table 3B). The age-specific HR estimates in the age intervals 20 to 49 and 50 to 69 clearly indicate much higher relative risk compared with the general population, declining in the age range from 50 to 69 years. Cumulative risk for EC also exceeds that of the general population at age 70 years, equaling 7.2% (95% CI, 0–16.9%) with HR of 6.3 (95% CI, 1.4–27.9; Table 3).

Although the cumulative risk of CRC for c.1865T>A carriers exceeds the risk in the general population by age 60 years, it is lower than for c.306+5G>A. Risk of CRC continues to increase and by age 70 is 6.8% (95% CI, 0–15.8%) for males and 7.3% (95% CI, 0–16.5%) for females. The overall HR for CRC is 2.4 (95% CI, 0.6–10.2) for males and 4.5 (95% CI, 1.1–18.0) for females. Risk for EC tends to exceed that of the general population at age 70 years, equaling 3.4% (95% CI, 0–10.0%) with HR of 2.9 (95% CI, 0.4–23.6; Table 3).

In carriers of other Spanish *MLH1* pathogenic mutations, estimated lifetime risk for CRC is 26.3% (95% CI, 7.8–41.1%) for men and 10.7% for women (95% CI, 0–16.8%), with HR of 10.5 (95% CI, 5.0–21.8) for males and 5.4 (95% CI, 2.0–14.9) for

females, and cumulative risk for EC is 6.5% (95% CI, 0–13.1%) with HR of 5.7 (95% CI, 1.9–17.0; Table 3). Therefore, both c.306+5G>A and c.1865T>A mutations show a penetrance within the range of variability estimated in families with other *MLH1* mutations, with a nonsignificant trend to lower penetrance for the c.1865T>A mutation.

### Haplotype analysis and estimation of mutation age

Ancestors of families carrying the c.306+5G>A mutation came from the Ebro river valley in northern Spain, whereas ancestors of families with the c.1865T>A mutation originated from the region of Jaén in southern Spain (Fig. 3). Common geographic origins of carrier families suggested that each of the two mutations could have occurred as a unique event in a single founder individual. This hypothesis was confirmed by haplotype analysis (Table 4).

The existence of a clear shared haplotype in carrier individuals permitted estimation of mutation age. In this estimation for the c.306+5T>A mutation, we considered disease haplotypes in the region extending from D3S2369 to D3S1298 (D3S1612 to D3S1298 for c.1865T>A; Table 4), which has a recombination map length of 0.9796 cM (2.2578 cM for c.1865T>A). In the control population, the estimated frequency of the identified minimum common haplotype, excluding the disease mutation, was 0.069 for c.306+5T>A and 0.048 for c.1865T>A (Supplementary Table S2).

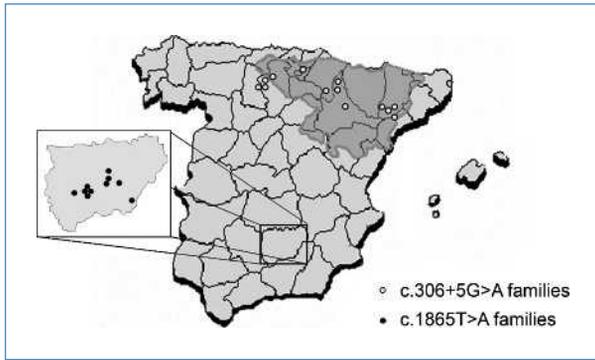
To estimate the age of the mutation, we relied on an estimate of the number of recombination events that occurred on the ancestral disease haplotype for each disease mutation. At the 5' end of the common haplotype of c.306+5G>A, a likely recombination event at marker D3S2369 was detected in family A10; at the 3' end (D3S1298), the common haplotype was lost in 6 of the 17 families, likely in two separate

**Table 3.** Age-specific cumulative risk and HR estimates for CRC and EC in carriers of c.306+5G>A, c.1865T>A, and other pathogenic *MLH1* mutations

<b>A</b>								
Age (y)	CRC (males)				CRC (females)			
	% Cumulative risk population	% Cumulative risk <i>MLH1</i> carriers (95% CI)			% Cumulative risk population	% Cumulative risk <i>MLH1</i> carriers (95% CI)		
		c.306+5G>A	c.1865T>A	Other <i>MLH1</i> mutations		c.306+5G>A	c.1865T>A	Other <i>MLH1</i> mutations
20	0	0	0	0	0	0	0	0
30	0.01	0.08	0.02	0.10	0.01	0.09	0.04	0.05
40	0.06	0.46	0.14	0.63	0.06	0.54	0.27	0.33
50	0.28	2.14 (0.04–4.20)	0.68 (0–1.64)	2.90 (0.78–4.96)	0.24	2.14 (0.17–4.06)	1.07 (0–2.54)	1.3 (0–2.59)
60	1.00	7.46	2.40	10.00	0.72	6.29	3.18	3.85
70	2.87	20.13 (1.42–35.95)	6.80 (0–15.80)	26.32 (7.85–41.07)	1.67	14.05 (1.20–25.23)	7.26 (0–16.50)	10.74 (0–16.80)
<b>EC (females)</b>								
20	0	0	0	0	0	0	0	0
30	0	0	0	0	0.01	0.12	0.04	0.06
40	0.02	0.13	0.06	0.11	0.08	0.97	0.35	0.46
50	0.13	0.82 (0–2.02)	0.38 (0–1.16)	0.74 (0–1.54)	0.37	4.41 (1.25–7.47)	1.63 (0–3.50)	2.10 (0.43–3.74)
60	0.55	3.42	1.60	3.1	1.27	14.41	5.50	7.07
70	1.17	7.16 (0–16.88)	3.37 (0–10.05)	6.5 (0–13.12)	2.84	29.58 (9.34–45.29)	11.98 (0–24.24)	16.24 (3.32–25.67)
<b>CRC and EC (females)</b>								
20	0	0	0	0	0	0	0	0
30	0	0	0	0	0.01	0.12	0.04	0.06
40	0.02	0.13	0.06	0.11	0.08	0.97	0.35	0.46
50	0.13	0.82 (0–2.02)	0.38 (0–1.16)	0.74 (0–1.54)	0.37	4.41 (1.25–7.47)	1.63 (0–3.50)	2.10 (0.43–3.74)
60	0.55	3.42	1.60	3.1	1.27	14.41	5.50	7.07
70	1.17	7.16 (0–16.88)	3.37 (0–10.05)	6.5 (0–13.12)	2.84	29.58 (9.34–45.29)	11.98 (0–24.24)	16.24 (3.32–25.67)
<b>B</b>								
Age (y)	HR (95% CI)							
	CRC (males)			CRC (females)				
	c.306+5G>A	c.1865T>A	Other <i>MLH1</i> mutations	c.306+5G>A	c.1865T>A	Other <i>MLH1</i> mutations		
20–49	14.89 (3.03–73.23)	4.10 (0.18–95.72)	32.42 (12.57–83.57)	13.36 (2.93–60.94)	9.56 (1.33–68.44)	12.63 (3.49–45.62)		
50–69	5.37 (1.47–19.72)	1.89 (0.325–10.95)	3.21 (0.88–11.70)	6.88 (1.95–24.28)	1 (NA)	2.30 (0.48–11.05)		
Overall HR	7.72 (2.89–20.59)	2.42 (0.57–10.23)	10.48 (5.04–21.81)	8.99 (3.58–22.56)	4.48 (1.11–18.01)	5.42 (1.97–14.93)		
<b>EC (females)</b>								
20–49	20.00 (2.77–144.19)	5.22 (0.15–185.08)	16.02 (3.19–80.42)	22.52 (7.49–67.70)	9.13 (1.62–51.61)	20.30 (7.84–52.27)		
50–69	3.67 (0.55–24.51)	2.17 (0.15–32.45)	2.98 (0.55–15.92)	8.41 (3.10–22.78)	1.73 (0.16–18.49)	1.99 (0.54–7.34)		
Overall HR	6.32 (1.43–27.94)	2.92 (0.36–23.57)	5.71 (1.91–17.02)	12.17 (5.92–25.01)	4.43 (1.37–14.36)	5.73 (2.59–12.71)		

NOTE: A. Age-specific cumulative risk of CRC and EC for male and female carriers of *MLH1* mutations compared with corresponding values for the population incidences as reported in the Spanish Cancer Registries (95% CIs are provided for cumulative risk at ages 50 and 70). B. Age-specific and overall HR for CRC and EC for male and female carriers of *MLH1* mutations (95% Wald CI is provided for the HR parameters).

Abbreviation: NA, not available.



**Figure 3.** Map of Spain showing locations where Lynch syndrome families with the c.306+5G>A mutation (white circles) and the c.1865T>A mutation (black circles) originated. For six families (four c.306+5G>A and two c.1865T>A), the original location is not known. The Ebro valley is highlighted in dark gray. The insert shows a larger-scale map of the Jaén province.

recombination events (Table 4A). Therefore, for c.306+5T>A, we counted a minimum of three recombination events necessary to explain all disease haplotypes in the region. For c.1865T>A, we counted a minimum of one event, which occurred at the 3' end (D3S1298) and which explained the loss of the common haplotype by 5 of the 12 families (Table 4B).

From the map lengths ( $r$ ) and recombination counts ( $k$ ), we estimated genealogy lengths ( $L$ ) using the formula  $\hat{L} = k/r$ . We obtained  $\hat{L}_{\text{Ebro}} = 3/0.009796 = 306$  generations and  $\hat{L}_{\text{Jaén}} = 1/0.022578 = 44$  generations. To estimate TMRCA from these lengths, we required an estimate of the ratio  $c$  (see Supplementary Materials and Methods). By simulating  $5 \times 10^7$  trees under a coalescent model of constant population size with  $n = 17$  lineages representing 17 families, we estimated  $\hat{c}_{\text{Ebro}} = 0.2324$ , which gives an estimate of  $\hat{H}_{\text{Ebro}} = \hat{c}_{\text{Ebro}} \hat{L}_{\text{Ebro}} = 3/0.009796 = 75$  generations (1,879 years assuming 25 years per generation) for the c.306+5G>A mutation. Similarly, using  $n = 12$  lineages, we estimated  $\hat{c}_{\text{Jaén}} = 0.2569$ , producing an age estimate of  $\hat{H}_{\text{Jaén}} = \hat{c}_{\text{Jaén}} \hat{L}_{\text{Jaén}} + 4 = 15$  generations (384 years assuming 25 years per generation) for c.1865T>A. The 95% CI for  $\hat{H}_{\text{Ebro}}$  is 53 to 122 generations, and the corresponding interval for  $\hat{H}_{\text{Jaén}}$  is 12 to 22 generations (Supplementary Table S2).

## Discussion

We have identified and characterized two frequent *MLH1* variants, c.306+5G>A and c.1865T>A (p.Leu622His), among Spanish Lynch syndrome families. They represent the first founder *MLH1* mutations identified in the Spanish population.

The predicted effect of the previously reported *MLH1* variant c.306+5G>A (36) on splicing was confirmed, consistent with observations on the nearby mutations c.306+2dupT and c.306+4A>G (37, 38). At the protein level, no differences in expression were observed in mutation carriers when compared with controls, presumably because of a lack of stability

of mutated *MLH1* proteins or due to the limitations of the Western blotting assay.

The c.1865T>A variant (p.Leu622His) was previously described as a germline *MLH1* mutation in two Spanish families (39, 40). Prior functional assays for p.Leu622His were inconclusive to assess its pathogenicity: In yeast-based assays, it was deficient in MMR activity (33, 41), whereas HCT116 cells cotransfected with *MLH1* and *PMS2* retained partial MMR activity (69.2%). The effect on protein expression levels was inconclusive (33). *In vitro* expression studies in *MLH1*-transfected HCT116 cells indicated that this mutation may substantially reduce *MLH1* expression by affecting its stability, which in turn leads to a reduction in the expression of its counterpart *PMS2*. A pathogenic role for the p.Leu622His mutation was further supported by the frequent loss of the *MLH1* WT allele observed in tumors (42). An additional effect of the mutation on the capacity to bind *PMS2* or on the intracellular trafficking of *MLH1/PMS2* could not be ruled out.

Once the functional effect of the founder mutations was shown, we explored whether penetrance estimates differed from other *MLH1* mutations. The estimated penetrance of Spanish *MLH1* mutations was moderate and consistent with data reported for Dutch families (5). The penetrance of founder mutations was within the range of variability of penetrance observed for Spanish *MLH1* mutations. The penetrance for c.1865T>A missense mutation may be lower, a fact that may be attributed to the observed decrease in *MLH1* expression. The penetrance estimates are sensitive to the type of ascertainment correction, and elevate with use of an unconditional likelihood or likelihood conditioned only on the proband, a phenomenon also noted in (2). Of note, penetrance estimates for Spanish Lynch syndrome families are lower than those reported in North American populations (2, 3). The fact that we have used the same method as Stoffel and colleagues (2) and that similar cumulative risks are observed in the United States and Spain suggests that geographic differences in cancer risk might exist in Lynch syndrome carriers. Several factors including lifestyle, environmental factors, mutational spectrum, or, as recently reported, the existence of distinct weak alleles capable of producing polygenic effects (43) may account for this observation.

Founder mutations in MMR genes have an important effect in traditional "founder populations": Four different mutations in *MLH1* and *MSH2* genes in Finland, Newfoundland, and in Ashkenazi Jews account for between 25% and 50% of all Lynch syndrome cases in these populations (9, 10, 13, 44). Before our report, two founder mutations in the *MSH2* gene had been identified in Spanish Lynch syndrome families, although their frequencies were low in our population (11, 12). In the cohort of Lynch syndrome families from ICO, the identified c.306+5G>A and c.1865T>A mutations account for 28.8% of the 45 families carrying *MLH1* alterations (13 founders and 32 nonfounders) and represent 17.6% (13 of 74) of all families with mutations in MMR genes. In addition, the c.306+5G>A founder accounts for up to 25% of *MLH1* mutations identified in the Ebro basin area (data not shown).

Our detection of founder mutations in *MLH1* supports the view that they can occur not only in groups commonly

**Table 4.** Haplotypes associated with the c.306+5G>A (A) and c.1865T>A (B) mutations in carrier families**A**

Marker (Mb on chromosome 3)	D3S1609 (29.915)	D3S1612 (34.565)	D3S2369 (36.472)	c.306+5G>A (37.018)	rs4234259 (37.024)	rs179997 (37.029)	D3S1611 (37.044)	rs9876116 (37.059)	D3S3623 (37.419)	D3S1298 (38.024)	D3S3564 (42.394)
Families											
A1	254	106	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	209
A2	250//254	106	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	213
A3	250	96	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>205</b>	213
A4	254	92	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>205</b>	201/211
A5	254/256	92	<b>103</b>	<b>Yes</b>	A/G	<b>A</b>	250/256	A/G	<b>224</b>	<b>205</b>	201/213
A6	250	104	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>205</b>	201/213
A7	250	96/104	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>201</b>	211
A8	250/254	96/106	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	218/224	199/201	203/209
A9	254	96	<b>103</b>	<b>Yes</b>	A/G	<b>A</b>	256/260	A/G	216/224	197/199	213/215
A10	250/256	96/104	<b>101</b>	<b>Yes</b>	<b>G</b>	A/G	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	203/215
A11	250	96	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	209/213
A12	250	96/104	<b>103</b>	<b>Yes</b>	<b>G</b>	A/G	<b>256</b>	<b>G</b>	216/224	195/199	201/215
A13	250	96	<b>103</b>	<b>Yes</b>	<b>G</b>	A/G	<b>256</b>	<b>G</b>	220/224	199/209	209/213
A14	252	96	101/103	<b>Yes</b>	<b>G</b>	A/G	<b>256</b>	<b>G</b>	<b>224</b>	<b>205</b>	211
A15	250//254	96	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	213
A16	250//254	96	<b>103</b>	<b>Yes</b>	<b>G</b>	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	213
A17	254	96	<b>103</b>	<b>Yes</b>	A/G	<b>A</b>	<b>256</b>	<b>G</b>	<b>224</b>	<b>199</b>	209//213
Max 1.55 Mb											
Min 0.40 Mb											
Shared alleles			103	Yes	G	A	256	G	224	199	
Control frequencies (%)			<b>67</b>	<b>0</b>	<b>44</b>	<b>66</b>	<b>44</b>	<b>44</b>	<b>12</b>	<b>30</b>	

(Continued on the following page)

**Table 4.** Haplotypes associated with the c.306+5G>A (A) and c.1865T>A (B) mutations in carrier families (Cont'd)**B**

Marker (Mb on chromosome 3)	D3S1609 (29.915)	D3S1612 (34.565)	D3S2369 (36.472)	rs4234259 (37.024)	rs1799977 (37.029)	D3S1611 (37.044)	rs9876116 (37.059)	c.1865T>A (37.064)	D3S3623 (37.419)	D3S1298 (38.024)	D3S3564 (42.394)
Families											
B1	256	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	205	207
B2	256	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	205	207
B3	252//256	92/ <b>104</b>	<b>101</b>	G	G	256	G	Yes	222	205	201//207
B4	256	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	205	201
B5	250	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	197	203
B6	250	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	205	207
B7	250/256	<b>104</b>	101/103	A/G	A/G	256	G	Yes	222	205	211
B8	250/256	<b>104</b>	101/103	G	A/G	256	G	Yes	222	205	201/207
B9	256	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	197	203
B10	250/254	92/ <b>104</b>	<b>101</b>	A/G	A/G	256/260	A/G	Yes	214/222	197/199	201/203
B11	250/256	96/ <b>104</b>	<b>101</b>	G	G	256	G	Yes	222	197/199	203/213
B12	250//256	<b>104</b>	<b>101</b>	G	G	256	G	Yes	222	197	209
Max 3.46 Mb											
Min 0.95 Mb											
Shared alleles		104	101	G	G	256	G	Yes	222	205	
Control frequencies (%)		<b>32</b>	<b>31</b>	<b>44</b>	<b>34</b>	<b>44</b>	<b>44</b>	<b>0</b>	<b>22</b>	<b>10</b>	

NOTE: The “/” symbol indicates that the phase of the disease haplotype cannot be established. The “//” symbol indicates that recombination has likely occurred within a family. Intragenic *MLH1* markers are shown in a box. Dark gray shading indicates nonrecombinant haplotypes. The sizes of the minimum and maximum conserved haplotypes are shown at the bottom. Inferred haplotypes of subjects in the fifth generation in each family are in bold. The frequencies of disease-associated alleles among a panel of control DNAs are shown at the bottom.

viewed as founder populations but also in geographically localized subsets of larger populations. These two founder mutations were not detected in a prior study of 1,222 incident cases of CRC in Spain (EPICOLON; ref. 45). This study identified 11 germline MMR mutations, 4 in *MLH1*. The lack of detection of founder mutations may be due to the low number of identified mutations or the relatively low coverage of the areas of origin.

Families carrying the c.306+5G>A mutation cluster within the Ebro river basin, in northern Spain, and we estimate that the mutation arose ~1,879 years ago. Taking into account the fact that the river valley is geographically isolated by mountain ranges, and the Ebro river was navigable until the 19th century, we hypothesize that the mutation arose somewhere in the valley and was distributed along the river over the years (Fig. 3). The c.1865T>A mutation is younger (384 years) than c.306+5G>A, and its distribution is restricted to the mountainous province of Jaén. It is noteworthy that probands have been identified in Madrid and Barcelona, frequent destinations of internal migratory movements during the period 1960 to 1970. Thus, the origin of the *MLH1* founder mutations can be linked to specific geographic areas, and their current distribution can be explained by migration patterns. Allele age estimation is generally imprecise (46) due to several factors, including uncertainty in the count of recombination events on ancestral haplotypes, the accuracy with which the recombination count predicts the true genealogy length, and the accuracy with which TMRCA reflects allele age. However, the relatively recent values obtained are sensible for rare mutations restricted to relatively isolated geographic areas.

## References

- Lynch HT, Lynch JF, Lynch PM, Attard T. Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. *Fam Cancer* 2008;7:27–39.
- Stoffel E, Mukherjee B, Raymond VM, et al. Calculation of risk of colorectal and endometrial cancer among patients with lynch syndrome. *Gastroenterology* 2009;137:1621–7.
- Choi YH, Cotterchio M, McKeown-Eyssen G, et al. Penetrance of colorectal cancer among *MLH1*/*MSH2* carriers participating in the colorectal cancer familial registry in Ontario. *Hered Cancer Clin Pract* 2009;7:14.
- Alarcon F, Lasset C, Carayol J, et al. Estimating cancer risk in HNPCC by the GRL method. *Eur J Hum Genet* 2007;15:831–6.
- Quehenberger F, Vasen HF, van Houtwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the *hMLH1* and *hMSH2* gene: correction for ascertainment. *J Med Genet* 2005;42:491–6.
- Pineda M, Gonzalez S, Lazaro C, Blanco I, Capella G. Detection of genetic alterations in hereditary colorectal cancer screening. *Mutat Res* 2009.
- Chong G, Jarry J, Marcus V, et al. High frequency of exon deletions and putative founder effects in French Canadian Lynch syndrome families. *Hum Mutat* 2009;30:E797–812.
- Tang R, Hsiung C, Wang JY, et al. Germ line *MLH1* and *MSH2* mutations in Taiwanese Lynch syndrome families: characterization of a founder genomic mutation in the *MLH1* gene. *Clin Genet* 2009;75:334–45.
- Nystrom-Lahti M, Kristo P, Nicolaidis NC, et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat Med* 1995;1:1203–6.
- Moisio AL, Sistonen P, Weissenbach J, de la Chapelle A, Peltomaki P. Age and origin of two common *MLH1* mutations predisposing to hereditary colon cancer. *Am J Hum Genet* 1996;59:1243–51.
- Medina-Arana V, Barrios Y, Fernandez-Peralta A, et al. New founding mutation in *MSH2* associated with hereditary nonpolyposis colorectal cancer syndrome on the Island of Tenerife. *Cancer Lett* 2006;244:268–73.
- Menéndez M, Castellví-Bel S, Pineda M, et al. Founder effect of a pathogenic *MSH2* mutation identified in Lynch syndrome Spanish families. *Clin Genet* 2010;78:186–90.
- Foulkes WD, Thiffault I, Gruber SB, 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* 2002;71:1395–412.
- Cederquist K, Emanuelsson M, Wiklund F, Golovleva I, Palmqvist R, Gronberg H. Two Swedish founder *MSH6* mutations, one nonsense and one missense, conferring high cumulative risk of Lynch syndrome. *Clin Genet* 2005;68:533–41.
- Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM. Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat* 2008;29:1304–13.
- Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de Wind N. Assessment of functional effects of unclassified genetic variants. *Hum Mutat* 2008;29:1314–26.
- Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum Mutat* 2008;29:1265–72.
- Moreno V, Gemignani F, Landi S, et al. Polymorphisms in genes of

The high prevalence and moderate penetrance of the c.306+5G>A and c.1865T>A variants have implications for molecular and clinical approaches to founder mutations in many settings. In specific populations such as the Spanish population, initial screening for founder mutations increases the efficiency of testing of molecular diagnostic labs. Even more importantly, careful clinical characterization of founder mutations can lead to mutation-specific counseling and improved clinical care.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

We thank the staff at the Genetic Counseling Units who have participated in this study and Michael DeGiorgio for helpful discussions about the mutation age estimates.

## Grant Support

Fundació La Caixa grant BM 04-107-0; Fundació Gastroenterologia Dr. Francisco Vilardell grant F05-01; Ministerio de Educación y Ciencia grants AGL2004-07579-04, SAF 06-6084, and SAF09-7319; Spanish Networks RTICCC grants RD06/0020/1050, 1051, 0021, and 0028; Acción en Cáncer (Instituto de Salud Carlos III); Fundació Roses Contra el Càncer; Fondo de Investigación Sanitaria (CP 03-0070); and NIH grant CA81488.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 02/17/2010; revised 07/15/2010; accepted 08/03/2010; published OnlineFirst 09/21/2010.

- nucleotide and base excision repair: risk and prognosis of colorectal cancer. *Clin Cancer Res* 2006;12:2101–8.
19. Dogan RI, Getoor L, Wilbur WJ, Mount SM. SplicePort—an interactive splice-site analysis tool. *Nucleic Acids Res* 2007;35:W285–91.
  20. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311–23.
  21. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science* 2002;297:1007–13.
  22. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003;31:3568–71.
  23. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;30:3894–900.
  24. Boyer JC, Umar A, Risinger JI, et al. Microsatellite instability, mismatch repair deficiency, and genetic defects in human cancer cell lines. *Cancer Res* 1995;55:6063–70.
  25. Perera S, Bapat B. The *MLH1* variants p.Arg265Cys and p.Lys618Ala affect protein stability while p.Leu749Gln affects heterodimer formation. *Hum Mutat* 2008;29:332.
  26. Pelletier R, Farrell BT, Miret JJ, Lahue RS. Mechanistic features of CAG\*CTG repeat contractions in cultured cells revealed by a novel genetic assay. *Nucleic Acids Res* 2005;33:5667–76.
  27. Lutzen A, de Wind N, Georgijevic D, Nielsen FC, Rasmussen LJ. Functional analysis of HNPCC-related missense mutations in *MSH2*. *Mutat Res* 2008;645:44–55.
  28. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with *BRCA1* or *BRCA2* mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003;72:1117–30.
  29. Lange K, Weeks D, Boehnke M. Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* 1988;5:471–2.
  30. Curado MP, Edwards B, Shin HR, editors. *Cancer incidence in five continents, vol. IX*. IARC Scientific Publications No. 160. Lyon: IARC; 2007.
  31. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003;73:1162–9.
  32. Schroeder KB, Jakobsson M, Crawford MH, et al. Haplotypic background of a private allele at high frequency in the Americas. *Mol Biol Evol* 2009;26:995–1016.
  33. Takahashi M, Shimodaira H, Andreutti-Zaugc 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.
  34. Chao EC, Velasquez JL, Witherspoon MS, et al. Accurate classification of *MLH1/MSH2* missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). *Hum Mutat* 2008;29:852–60.
  35. Guarne A, Ramon-Maiques S, Wolff EM, et al. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO J* 2004;23:4134–45.
  36. Becouarn Y, Rullier A, Gorry P, et al. Value of microsatellite instability typing in detecting hereditary non-polyposis colorectal cancer. A prospective multicentric study by the Association Aquitaine Gastro. *Gastroenterol Clin Biol* 2005;29:667–75.
  37. Pagenstecher C, Wehner M, Friedl W, et al. Aberrant splicing in *MLH1* and *MSH2* due to exonic and intronic variants. *Hum Genet* 2006;119:9–22.
  38. Tournier I, Vezain M, Martins A, et al. A large fraction of unclassified variants of the mismatch repair genes *MLH1* and *MSH2* is associated with splicing defects. *Hum Mutat* 2008;29:1412–24.
  39. Godino J, de La Hoya M, Diaz-Rubio E, Benito M, Caldes T. Eight novel germline *MLH1* and *MSH2* mutations in hereditary non-polyposis colorectal cancer families from Spain. *Hum Mutat* 2001;18:549.
  40. Palicio M, Balmana J, Gonzalez S, et al. Mismatch repair gene analysis in Catalonian families with colorectal cancer. *J Med Genet* 2002;39:E29.
  41. Wanat JJ, Singh N, Alani E. The effect of genetic background on the function of *Saccharomyces cerevisiae* *mlh1* alleles that correspond to HNPCC missense mutations. *Hum Mol Genet* 2007;16:445–52.
  42. Tuupainen S, Karhu A, Jarvinen H, Mecklin JP, Launonen V, Aaltonen LA. No evidence for dual role of loss of heterozygosity in hereditary non-polyposis colorectal cancer. *Oncogene* 2007;26:2513–7.
  43. 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.
  44. Froggatt NJ, Green J, Brassett C, et al. 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.
  45. Pinol V, Castells A, Andreu M, et al. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005;293:1986–94.
  46. Slatkin M, Rannala B. Estimating allele age. *Annu Rev Genomics Hum Genet* 2000;1:225–49.