ESTUDI DE VARIANTS DE SIGNIFICAT DESCONEGUT EN LA SÍNDROME DE LYNCH

Ester Borràs Flores

Gener, 2013

ESTUDI DE VARIANTS DE SIGNIFICAT DESCONEGUT EN LA SÍNDROME DE LYNCH

Memòria presentada per

Ester Borràs Flores

Per optar al Grau de

Doctora per la Universitat de Barcelona

Tesi realitzada sota la direcció del

Dr. **Gabriel Capellá Munar** i la Dra. **Marta Pineda Riu**

al Institut Català d'Oncologia del

Institut d'Investigació Biomèdica de Bellvitge

(**ICO-IDIBELL**)

Tesi adscrita a la Facultat de Medicina, Universitat de Barcelona (**UB**)

Programa de Doctorat en Biomedicina.

Tutora: **Virginia Nunes Martínez**

Gabriel Capellá Munar **Marta Pineda Riu** **Virgina Nunes Martínez**

Ester Borràs Flores

Barcelona, 2012

A la meva família

AGRAÏMENTS

Doncs per fi sóc aquí, escrivint la part de la tesi que més gent llegirà, la que més em costarà, la més emotiva, la que significa que la tesi s'ha acabat i una nova etapa s'obre davant meu, amb la incertesa que això comporta. Aquests agraïments són un recordatori de la gent amb qui he conviscut aquest temps durant el que hem compartit moments irrepetibles. A tots, moltes gràcies.

En primer lloc vull donar les gràcies als meus dos directors, en Gabi i la Marta. Gabi, gràcies pel teu recolzament, per confiar en les meves possibilitats, per ensenyar-me a ser crítica, pels bons moments que han estat molts, i pels consells que encara n'han estat més. Marta, gràcies per tot, per donar-me aquesta oportunitat i per estar al meu costat des del principi. Recordo perfectament la primera tarda al laboratori i com em vas ensenyar a posar la meva primera PCR!! Com diria en Gabi, per fer-me de "mama Marta", amb permís d'en Pau i la Clara, per ensenyar-me tantes coses, per tots els moments que hem passat juntes i per la paciència i comprensió que m'has dedicat.

També vull donar les gràcies a la Conxi, qui junt amb la Marta em van fer la primera entrevista a l'ICO i em van oferir la possibilitat de fer el màster. Sense aquesta oportunitat aquesta tesi no hauria estat possible, gràcies també pels teus consells i ànims en tot moment.

Al grup de diagnòstic, amb qui vaig començar els meus primers passos en aquesta aventura i que des del principi m'hi vaig sentir molt a gust. Sara, vaig aprendre molt treballant amb tu en el paper d'APC. Nònia, gràcies per deixar-me la teva cadira i envair cada dos per tres el teu despatx. Jesús, els teus coneixement i consells m'han estat de molta ajuda. I com no a la Mir, com oblidar el nostre primer WB juntes i lo tard que se'ns va fer, gràcies per tot, les xerrades, sortides, la complicitat, sé que sempre puc comptar amb tu.

A la Raquel, l'Eva, l'Anna, l'Anabel, la Susana i la Mònica, per ensenyar-me tant, des de ser organitzada al lab a treballar en equip. Gràcies per la vostra amistat i companyerisme, ha estat un plaer compartir aquest temps amb vosaltres... ja us trobo a faltar!! Carol, no tinc paraules per agrair-te lo bona amiga a part de companya que has estat per mi durant tot aquest temps, per recolzar-me en els moments difícils, pels nostres cafès, sopars, sortides i compres varies. I que dir de l'Olguinsssss? Doncs que està feta una artista! Moltes gràcies per tot!

Als predocs! A la Vanessines per sempre fer-me somriure i pel seu entusiasme. A la Elena per estar sempre allà i pels favors varis. A en Fernando i la Núria amb qui he compartint el "zulito", i on vaig estar-hi molt a gust. A en Paco i en Joan, que sempre han estat organitzant coses per fer més amena la vida del predoc. I a la més recent incorporació, Gardenia, *pero que buena eres*!

A la gent del grup de recerca del LRT2, en Dani, la Mireia Morell, la Nadia, i en especial a la Juani, per estar allà quan la necessitava, i la Laura per l'entusiasme i energia que transmet. Recordar també a l'Edu i en Carles, que ens van acompanyar durant un temps a les reunions. De tots vosaltres n'he après molt, tan personal com professionalment, sou uns *cracks*!. Els vostres consells i suggerències m'han estat de gran ajuda per tirar endavant aquesta tesi.

A l'Ester, pels seus consells quan era al lab i pels seus consells a distància. A la Pili, perquè encara que els últims anys ha estat una mica lluny sempre hem estat molt a prop i m'ha fet costat en tot! Als estudiants de màster, de pràctiques.. i a tots els que han anat passant pel lab durant aquest temps, en especial a la Thais, en Xavi i a la Flipi! A la Iolanda, la Laura i la Natàlia, per ajudar-me en tràmits varis, comandes, papers.... no sé que faríem sense elles! A la gent del LRT1, sempre que m'ha fet falta alguna cosa han estat allà. A la Virginia Nunes, per acceptar de ser la meva tutora.
A les dues "Badalonines" més il·lustres, Torner i Gau, pensàveu que me'n havia oblidat? Moltes gràcies per ser com sou, per tenir sempre un somriure, una broma, però també per tenir un minut per escoltar i compartir. Gràcies de tot cor.

Als membres de la Unitat de Consell Genètic, per la part clínica que hi ha darrere d'aquesta tesi (sou vitals pels pacients!). Al departament de bioestadística, en especial a l'Eli Guinó, qui em va ajudar a resoldre molts dubtes i qui sempre trobava un moment o altre per a mi. A en Guido i l'Angela, per acollir-me com una més en el seu grup...*Danke*. Al grup "españoles en Frankfurt" en especial a l'Albert, Laia, Vicky, Alícia i Pau, amb qui vaig passar molt bones estones i em van ajudar a sentir-me com a casa.

A la gent de Balagué, sobretot al equip "guantanamo matí" i en especial a la Laura i la Meri, que sempre han estat molt bones amigues. A la Magi, qui em va facilitar molt les coses per iniciar aquest camí. A la gent del màster, sobretot a l'Anna i a la Clara, amb qui vaig compartir infinitat de moments i em van recolzar en aquest nou camí. A l'Alba, gràcies per estar sempre al meu costat, per alegrar-te per mi i per la teva amistat.

A la meva nova família, i mai millor dit! Pilar, Joan, moltes gràcies per tot, des del principi m'heu fet sentir com una més de la família. Joan, Marta, Gemma i Eduard, gràcies pel vostre suport i per la confiança dipositada en mi des del principi. Als meus nebots, Carla, Alèxia, Martí, Adrià i Maria, sou uns Sols, gràcies per la vostra alegria, entusiasme i espontaneïtat.

Als meus pares per la seva estima, perquè sempre han estat al meu costat i m'han donat suport en tot el que he fet, independentment de si estaven o no d'acord. Sense ells això no hagués estat possible de cap de les maneres. A la meva germana, Cris, ets la millor germana del món... quan vols! I no podia oblidar-me dels meus avis, sé que no entenen molt bé el que faig però sempre m'han fet costat. Avi et trobo molt a faltar.

I per últim, gràcies a tu Raül, per ser com ets, per estar sempre al meu costat, per aguantar-me aquest últims mesos....sé que no han estat fàcils, pels teus consells, per donar-me ànims en els moments de desesperació, per fer-me somriure quan ho necessitava, per confiar sempre en mi, per fer-me adonar que res és impossible, per fer-me tan feliç i per estimar-me tant.

Per tot això i molt més, només et vull dir...

...vull donar-te el meu món...

...que tot sigui nou...

...mirar de sorprendre't. Gràcies vida!

ABREVIATURES I ACRÒNIMS

AA	aminoàcid
ACVR2A	de l'anglès, *activin A receptor, type IIA*
ADP	adenosina difosfat; de l'anglès *adenosine diphosphate*
AFAP	poliposi adenomatosa familiar atenuada; de l'anglès *Attenuated familial adenomatous polyposis*
APC	de l'anglès, *adenomatous polyposis coli*
ATP	adenosina trifosfat; de l'anglès, *adenosine-5'-triphosphate*
ATPasa	adenosina trifosfatasa; de l'anglès, *adenosine-5'-triphosphatase*
ASE	expressió específica d'al·lel; de l'anglès, *allelic specific expression*
BRAF	de l'anglès, *v-raf murine sarcoma viral oncogene homolog B1*
CCR	càncer colorectal
cDNA	DNA complementari; de l'anglès, *complementary DNA*
CE	càncer d'endometri
CIN	inestabilitat cromosòmica; de l'anglès chromosomal inestability
CIMP	via del fenotip metilador d'illes CpG; de l'anglès *CpG island methylator phenotype*
CMMR-D	síndrome de dèficit constitucional de reparació MMR; de l'anglès *Constitutional mismatch repair-deficiency syndrome*
CSCE	electroforesis capilar sensible a conformació; de l'anglès *conformation sensitive capillary electrophoresis*
DGGE	electroforesi en gel de gradient desnaturalitzant; de l'anglès, *denaturing gradient gel electrophoresis*
DHPLC	cromatografia líquida desnaturalitzant d'alt rendiment; de l'anglès, *denaturing high performance liquid chromatography*
DNA	àcid desoxirribonucleic; de l'anglès, *deoxyribonucleic acid*
ESE	element exònic activador de l'*splicing*; de l'anglès, *exonic splicing enhancer*
EXO I	de l'anglès, exonuclease 1
EXO VII	de l'anglès, exonuclease 7
EXO V	de l'anglès, exonuclease 5
FAP	poliposi adenomatosa familiar; de l'anglès *Familial adenomatous polyposis*
FBXW7	de l'anglès, *F-box and WD repeat domain containing 7*
FOBT	test de sang oculta en femetes; de l'anglès *faecal occult blood test*
HRM	alta resolució de curves de fusió; de l'anglès *high resolution melting*
HNPCC	càncer colorectal hereditari no polipós; de l'anglès Hereditary non-polyposis colorectal cancer
ICO	Institut Català d'Oncologia
IHQ	inmunohistoquímica
kDa	quilodaltons
KRAS	de l'anglès, *v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*
LOH	pèrdua d'heterozigositat; de l'anglès, *loss of heterozigosity*
LOVD	de l'anglès, *Leiden Open Variation Database*
LR-PCR	PCR de llarg abast; de l'anglès, *Long range-PCR*
MAP	poliposis associada al gen *MYH*; de l'anglès *MYH associated polyposis*
MCC	de l'anglès, *mutated in colorectal cancers*

MGMT	de l'anglès, *O-6-methylguanine-DNA methyltransferase*
MLH1	de l'anglès, *mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)*
MLPA	amplificació depenent de lligasa de múltiples sondes; de l'anglès, *multiplex ligation-dependent probe amplification*
MMR	de l'anglès, mismatch repair
mRNA	àcid ribonucleic missatger; de l'anglès *messenger RNA*
MSH2	de l'anglès, *mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)*
MSH3	de l'anglès, *mutS homolog 3 (E.Coli)*
MSH6	de l'anglès, *mutS homolog 6 (E.Coli)*
MSI	inestabilitat de microsatèl·lits; de l'anglès *microsatellite instability*
MSI-H	inestabilitat de microsatèl·lits elevada; de l'anglès *microsatellite instability high*
MSI-L	inestabilitat de microsatèl·lits baixa; de l'anglès *microsatellite instability low*
MS-MLPA	MLPA per metilació específica; de l'anglès *methylation-specific MLPA*
MSS	sense inestabilitat de microsatèl·lits; de l'anglès *microsatellite stable*
MYH	de l'anglès, *mutY homolog (E. coli)*
NGS	seqüenciació de nova generació; de l'anglès *Next-generation sequencing*
NMD	de l'anglès, *nonsense mediated mRNA decay*
NRAS	de l'anglès, *neuroblastoma RAS viral (v-ras) oncogene homolog*
PCNA	de l'anglès, *proliferating cell nuclear antigen*
PCR	de l'anglès, *polymerase chain reaction*
PI3KCA	de l'anglès, *phosphatidyl inositol 3-kinase catalytic subunit*
PMS1	de l'anglès, *PMS1* postmeiotic segregation increased 1 (S. cerevisiae)
PMS2	de l'anglès, *PMS2 postmeiotic segregation increased 2 (S. cerevisiae)*
PMS2CL	de l'anglès, *PMS2 C-terminal like pseudogene*
RNA	àcid ribonucleic, de l'anglès, *ribonucleic acid*
RPA	de l'anglès, *replication protein A*
RT-PCR	de l'anglès, *reverse transcription polymerase chain reaction*
SMAD4	de l'anglès, *SMAD family member 4*
SLC9A9	de l'anglès, *solute carrier family 9, subfamily A, member 9*
SSB	de l'anglès, single-stranded DNA-binding protein
SSCP	polimorfisme de conformació de cadena senzilla; de l'anglès, *single strand conformation polymorphism*
STRs	de l'anglés *short tandem repeats*
TGFβR2	de l'anglès, *transforming growth factor, beta receptor II*
TCF7L2	de l'anglès, *transcription factor 7-like 2*
TP53	de l'anglès, *tumor protein p53*
VSD	variant de significat desconegut

ÍNDEX

INTRODUCCIÓ

1. EL CÀNCER COLORECTAL

1.1 Epidemiologia del càncer colorectal

El càncer colorectal (CCR), és el tercer tipus de càncer més freqüent en homes i el segon en dones a nivell mundial. L'any 2008, es van comptabilitzar més de 1,2 milions de nous casos al món, i més de mig milió de persones van morir a causa d'aquesta malaltia (Ferlay, et al., 2010). Les taxes d'incidència més elevades es donen a Austràlia, Nova Zelanda, Europa i Amèrica del Nord, mentre que les taxes més baixes es donen a l'Àfrica i al centre i sud de l'Àsia. Pel que fa les taxes de mortalitat, aquestes estan disminuint en alguns països occidentals, degut en gran part com a resultat d'un millor tractament, una major sensibilització i detecció primerenca, mentre que segueixen augmentant en molts països amb recursos i infraestructures sanitàries limitades, particularment a l'Europa de l'Est i l'Amèrica del Sud i Central (Jemal, et al., 2011). Alguns factors de risc a tenir en compte pel desenvolupament del CCR són el tabaquisme, la manca d'activitat física, el sobrepès, l'obesitat, el consum de carn vermella o processada, i el consum excessiu d'alcohol (Boyle and Levin, 2008; Ferrari, et al., 2007; Giovannucci, 2006).

A Espanya, el CCR és la neoplàsia més freqüent si tenim en compte els dos sexes alhora (15,51% de freqüència relativa), amb 28000 casos nous diagnosticats cada any. A més, és la tercera neoplàsia més freqüent en homes (14,9%), després de les de pròstata (22,56%) i pulmó (17,96%), i la segona neoplàsia en incidència en dones (16,45%), després del càncer de mama (30,49%) (Bray, et al., 2012; Ferlay, et al., 2010) (Figura 1). Per altra banda, el CCR representa la segona causa de mort per càncer a Espanya tant en homes com en dones, després del càncer de pulmó i de mama, respectivament, així com si considerem ambdós sexes alhora (Bray, et al., 2012; Ferlay, et al., 2010) (Figura 1).

En els països desenvolupats, més del 90% dels casos diagnosticats de CCR es donen en individus majors de 50 anys. La presentació de variants hereditàries mostra un avançament de 10-15 anys respecte a les mateixes localitzacions esporàdiques. El CCR s'inicia amb l'aparició de pòlips adenomatosos benignes a l'epiteli colorectal, que amb el temps es transformaran en tumors malignes. En el desenvolupament del CCR hi intervenen factors genètics i ambientals. Així, el CCR és el resultat d'un procés gradual d'adquisició de noves característiques per part de les cèl·lules epitelials, resultant de la continua acumulació de múltiples canvis genètics i epigenètics.

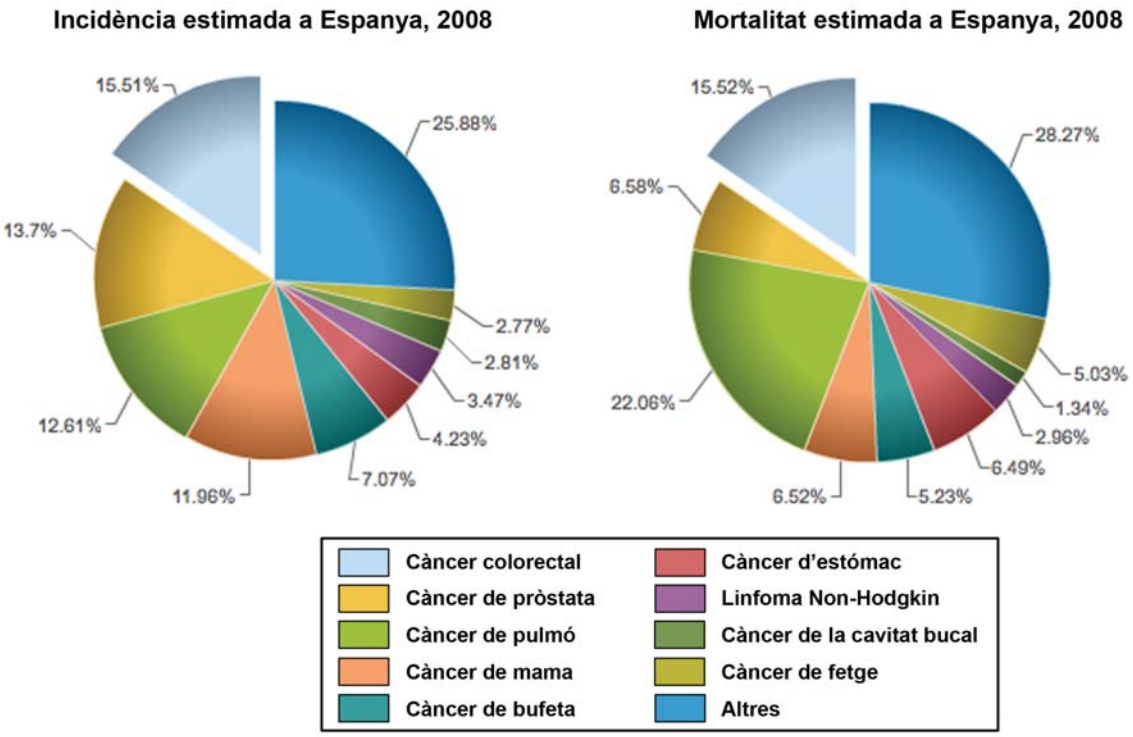


Figura 1. Incidència i mortalitat estimada dels diferents tipus de càncer en ambdós sexes a Espanya l'any 2008. Adaptat de Ferlay, et al., 2010.

1.2 Anatomia i histologia del còlon i recte

El còlon i el recte són la part terminal del tracte digestiu. Tots dos formen el budell gruixut que té una longitud aproximada de 1,5 metres. La funció del còlon és la d'absorbir nutrients, electròlits i aigua dels aliments ingerits per incorporar-los al torrent sanguini. El recte fa de reservori del material abans de l'expulsió. Anatòmicament el còlon es subdivideix en còlon ascendent, transvers, descendent i sigmoide, tot i que a nivell clinico-patològic sol dividir-se en còlon proximal o dret (que comprèn l'ascendent i el transvers fins l'angle esplènic) i distal o esquerra (que comprèn el descendent, el sigmoide i el recte) (Figura 2A). Histològicament, el gruix del tub està format per diferents capes, des de la llum fins a l'exterior: mucosa (formada de cèl·lules epitelials, làmina pròpia i muscularis mucosa), submucosa, muscularis pròpia, subserosa i serosa (Figura 2B).

L'epiteli del còlon té invaginacions que formen unes glàndules intestinals anomenades criptes de Lieberkühn, que incrementen la superfície del budell. És precisament a les criptes, on es regula l'homeòstasi de l'epiteli colònic gràcies a un equilibri dinàmic entre producció i mort cel·lular, necessari degut a l'elevada toxicitat química i estrès mecànic que pateix aquest epiteli (Nathke, 2004). Al fons de les criptes es troba el reservori de les cèl·lules mare, que van renovant les cèl·lules de l'epiteli en

un procés de proliferació, diferenciació i migració cap al lumen del budell. Les cèl·lules mare són cèl·lules pluripotents o progenitores no diferenciades que ocupen la tercera part més basal de la cripta i es divideixen cada 12 hores aproximadament (van de Wetering, et al., 2002). A la regió mitjana de la cripta, les cèl·lules progenitores es diferencien en: enteròcits (absorbeixen aigua i nutrients), cèl·lules caliciformes (productores de mucina) i cèl·lules neuroendoncrines (productores d'hormones). Quan les cèl·lules arriben a la superfície epitelial pateixen un procés d'apoptosi i/o extrusió cap al lumen. El trajecte de les cèl·lules des de la base de la cripta fins a la seva extrusió dura aproximadament entre 3 i 5 dies (Nathke, 2004; Radtke and Clevers, 2005).

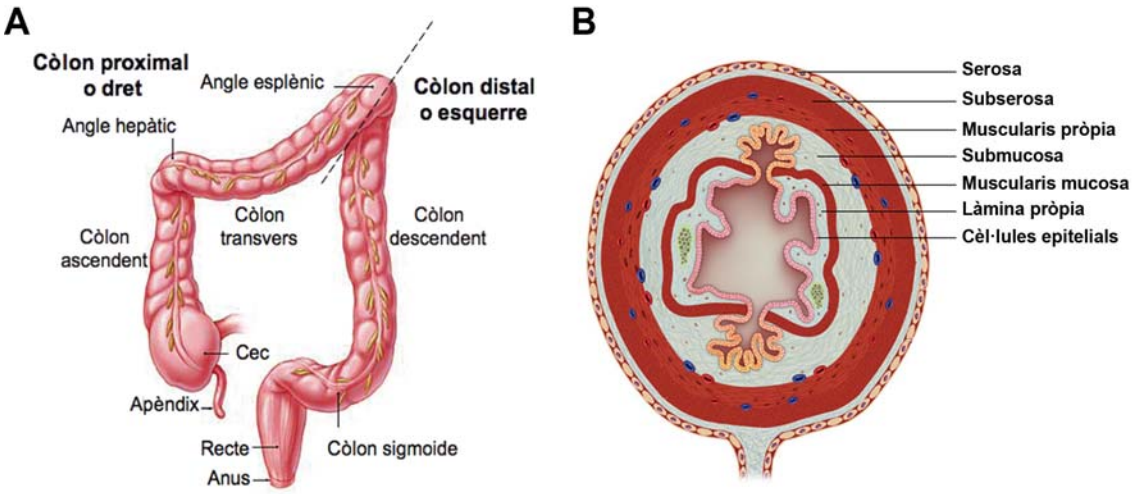


Figura 2. Representació esquemàtica del còlon. A. Parts del còlon. **B.** Capes de la paret del còlon.

1.3 Classificació i estadiatge del càncer de còlon i recte

En el cas del CCR, l'estadi tumoral és el factor pronòstic més important i el principal determinant del tractament dels pacients. Tant és així, que en estadis inicials la cirurgia acostuma a ser l'únic tractament necessari, mentre que en estadis més avançats són necessàries quimioteràpia i/o radioteràpia adjuvant i/o neoadjuvant. L'evolució clínica i els estadis del tumor es defineixen en funció del seu grau d'invasió i infiltració a través de la paret de l'intestí gros, del seu grau de disseminació ganglionar i de la presència de metàstasis a distància. La classificació de Dukes és el sistema de classificació original dels càncers de còlon i recte (Dukes, 1932). Es basa en el grau d'infiltració del tumor primari a través de les diverses capes de l'intestí gros i l'afectació a òrgans veïns i ganglis. Posteriorment es va desenvolupar la classificació d'Astler-Coller (Astler and Coller, 1954), la qual diferencia la carcinogènesis del CCR en 6 grups (Taula 1).

Taula 1. Classificació de Dukes modificada per Astler-Coller.

Estadis	Extensió del tumor
A	El tumor està limitat a la mucosa
B1	El tumor envaeix fins la muscularis pròpia
B2	El tumor travessa la muscularis pròpia fins la capa serosa
C1	Igual que l'estadi B1 més metàstasi ganglionar
C2	Igual que l'estadi B2 més metàstasi ganglionar
D	Metàstasi a distància o invasió parietal d'òrgans adjacents

Més recentment, es va definir una tercera classificació per tal d'evitar confusions en les dues classificacions anteriors, és el sistema TNM de la *International Union Against Cancer* (Fleming, et al., 1997) (Taula 2). Aquest sistema ens permet classificar els pacients en diversos estadis amb pronòstics associats diferents.

Taula 2. Classificació TNM per al càncer colorectal.

Categories T	Extensió que ocupa el tumor primari
Tx	La informació és incompleta i no es pot definir l'extensió del tumor
Tis	Carcinoma *in si*tu. És l'estadi més inicial, el tumor no ha travessat la mucosa
T1	El tumor ha travessat la mucosa, la muscularis mucosa, i envaeix la submucosa
T2	El tumor ha travessat la mucosa, la muscularis mucosa, la submucosa i envaeix la muscularis pròpia
T3	El tumor ha travessat la mucosa, la muscularis mucosa, la submucosa, la muscularis pròpia, i envaeix la subserosa sense arribar a cap teixit veí
T4	El tumor envaeix altres òrgans o estructures
Categories N	**Absència o presència de metàstasi als ganglis limfàtics**
Nx	La informació és incompleta i no es coneix l'afectació de ganglis limfàtics
N0	No hi ha cap gangli afectat
N1	1-3 ganglis afectats
N2	4 o més ganglis afectats
Categories M	**Absència o presència de metàstasi a distància**
Mx	La informació és incompleta i no hi ha descripció de metàstasi a distància
M0	No hi ha metàstasi a distància
M1	Presència de metàstasi a distància

El CCR en les primeres etapes no sol produir símptomes específics, i no són diferents dels originats per altres malalties benignes comunes en la nostra societat. És per aquest motiu que són molt importants les proves de detecció precoç, com el test de

sang oculta en femtes (FOBT) o la colonoscòpia, entre d'altres, ja que la detecció precoç millora notablement la supervivència dels pacients. Tot i així, el CCR pot presentar símptomes variats i inespecífics, como ara: canvi en el ritme de les deposicions de diverses setmanes de duració (diarrea, restrenyiment i sensació d'evacuació incompleta després de les deposicions), anèmia ferropènica no detectada prèviament i sense una causa que la pugui justificar, presència de sang a les femtes, molèsties abdominals repetides i d'intensitat progressiva sense causa coneguda (dolor, inflamació i acumulació de gasos que poden augmentar amb la ingesta) o pèrdua de pes sense causa coneguda.

1.4 La carcinogènesis colorectal

El primer model molecular de carcinogènesis colorectal va ser descrit per Fearon i Vogelstein l'any 1990 (Fearon and Vogelstein, 1990). Aquest model oferia la primera aproximació a la base molecular de la seqüència adenoma-carcinoma i postula que un número reduït d'alteracions genètiques (entre 4 i 6 alteracions) que activen oncògens (com pot ser *KRAS*) o inactiven gens supressors tumorals (ST) (per exemple: *APC, SMAD4, TP53*) són necessaris perquè una cèl·lula epitelial colònica mostri totes les característiques pròpies de la neoplàsia (revisat a Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011).

En un estudi recent, s'ha analitzat una sèrie de tumors colorectals, i s'han identificat diferents perfils de mutació. Hi ha tumors que presenten hipermutacions somàtiques mentre que d'altres no en presenten. Tot i així, aquests dos tipus de tumors comparteixen alguns gens mutats. Els tumors sense hipermutació presenten mutacions en *APC, TP53, KRAS, PIK3CA, FBXW7, SMAD4, TCF7L2* i *NRAS*, entre d'altres. En canvi, els tumors amb hipermutació presenten altres gens mutats com *ACVR2A, APC, TGFβR2, MSH3, MSH6, SLC9A9* i *TCF7L2*, així com mutacions en *BRAF*. Cal destacar que dos dels gens típicament mutats en els tumors sense hipermutacions tenen una freqüència molt menor en els tumors amb hipermutacions: *TP53* (60% versus 20%) i *APC* (81% versus 51%). Això indica que aquests dos tipus de tumors presenten un procés de tumorgènesi amb diferents esdeveniments genètics. A més, els tumors amb hipermutacions però que presenten *MLH1* silenciat i inestabilitat de microsatèl·lits mostren unes diferències addicionals en el perfil de mutació (Cancer Genome Atlas, 2012).

Es postula que la inestabilitat genètica augmentada en les cèl·lules tumorals permet l'acumulació de múltiples alteracions genètiques (revisat a Hanahan and Weinberg,

2011). Actualment, es coneix que el procés de carcinogènesis és altament heterogeni, i s'ha postulat que aquesta heterogeneïtat ve condicionada, en part, pel tipus d'inestabilitat genètica present en una determinada cèl·lula tumoral. Segons aquesta inestabilitat s'accepta l'existència de tres vies de carcinogènesis colorectal: la via de la inestabilitat cromosòmica (CIN), la via de la inestabilitat de microsatèl·lits (MSI) i la via del fenotip metilador d'illes CpG (CIMP) (Figura 3).

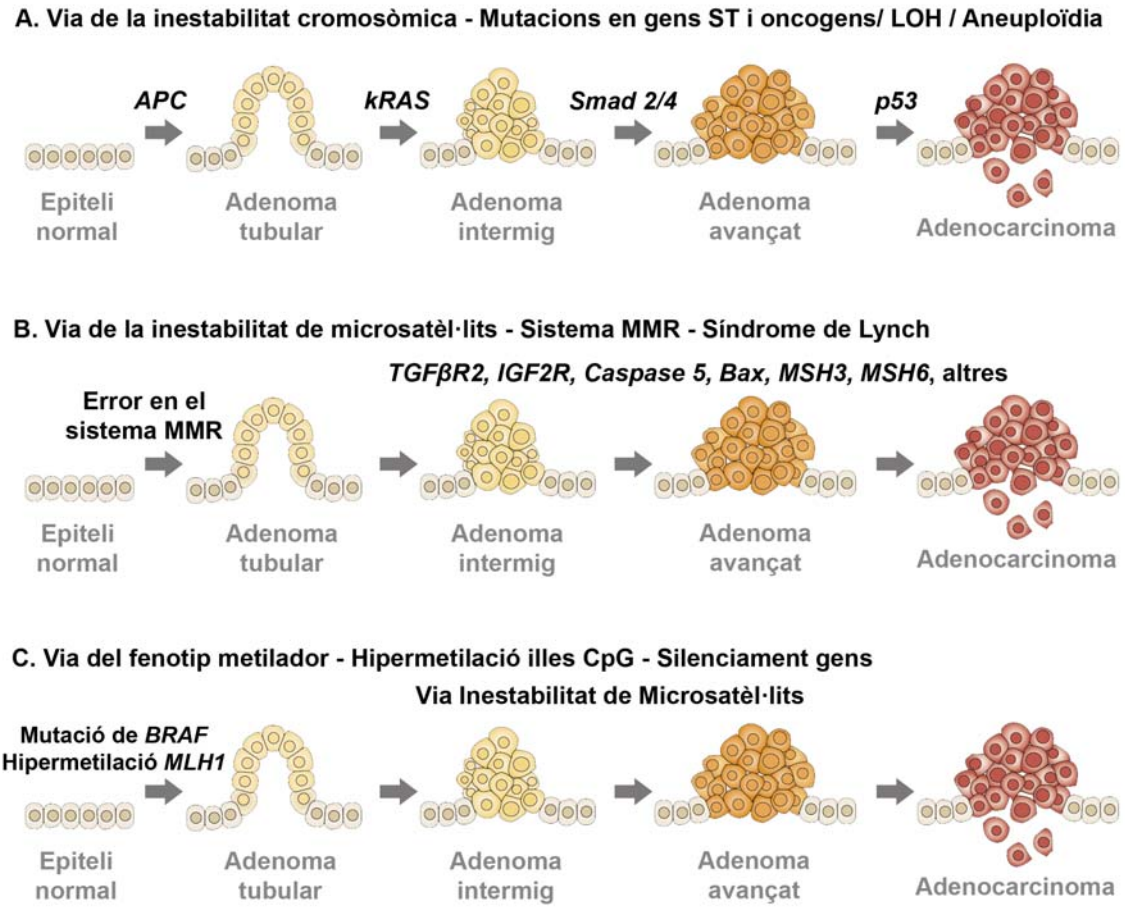


Figura 3. Vies de carcinogènesis colorectal. Modificat de Ahnen, 2011.

Via de la inestabilitat cromosòmica

Aproximadament entre un 80-85% de tots els casos de CCR segueixen la via "tradicional", anomenada també via de la inestabilitat cromosòmica (CIN) o fenotip supressor (Grady and Carethers, 2008). Aquesta via segueix una carcinogènesis seqüencial en que l'adquisició i acumulació de diferents alteracions a nivell molecular genera l'aparició d'estadis més avançats del procés cancerós (Figura 3). Els canvis genòmics inclouen l'activació de proto-oncògens com *KRAS*, i la inactivació de com a

mínim tres gens supressors tumorals, com ara la pèrdua de *APC*, pèrdua de *p53* i la pèrdua de heterozigositat del braç llarg del cromosoma 18 (Weinberg, 2006). Recentment, s'han descrit mutacions que impliquen altres gens, com per exemple *TGFβ2R* i *PIK3CA*, i que s'han inclòs en el model de seqüència adenoma-carcinoma (Armaghany, et al., 2012; Cancer Genome Atlas, 2012). A més, els tumors CIN mostren freqüentment un cariotip anòmal, amb pèrdues i guanys cromosòmics.

Via de la inestabilitat de microsatèl·lits

Els microsatèl·lits, també anomenats *short tandem repeats* (STRs), són seqüències repetitives del DNA, composades per repeticions de 1 a 5 nucleòtids, distribuïdes per tot el genoma (Abeloff, 1995). La inestabilitat de microsatèl·lits (MSI) és defineix com l'augment o disminució de la longitud d'un microsatèl·lit determinat en una cèl·lula en comparació amb la seva longitud en línia germinal. Aquesta variació és, per tant, un canvi somàtic.

Entre el 7 i el 15% de tots els casos de CCR -depenent de l'origen geogràfic dels pacients- mostren MSI, com també el 95% dels tumors de pacients amb síndrome de Lynch (Aaltonen, et al., 1993; Gonzalez-Garcia, et al., 2000; Ionov, et al., 1993; Pedroni, et al., 1999). Aquesta inestabilitat és produïda per la inactivació dels gens del sistema de reparació de desaparellaments erronis (gens *Mismatch repair*, MMR), donant lloc a la no correcció dels errors produïts durant la replicació del DNA. Els microsatèl·lits són particularment susceptibles a errors durant aquest procés, ja que la seva estructura repetitiva propicia que la DNA polimerasa "s'equivoqui" al copiar la cadena motlle (Ionov, et al., 1993). Quan els microsatèl·lits es troben en una regió codificant, la no reparació d'aquests errors pot provocar la generació de mutacions de canvi de pauta de lectura (*frameshift*), que com a resultat donaran lloc a la pèrdua de funció de la proteïna i que condiciona l'espectre mutacional característic d'aquests tumors. A la Figura 4, es mostren gens relacionats amb la tumorogènesi que són diana de la MSI, alguns d'aquests gens, com per exemple *TGFβR2* (Markowitz, et al., 1995), o *ACVR2A* (Jung, et al., 2004) són gens supressors tumorals. Es pensa que aquest fet provoca l'acceleració en el procés de la carcinogènesi, escurçat-lo en el temps.

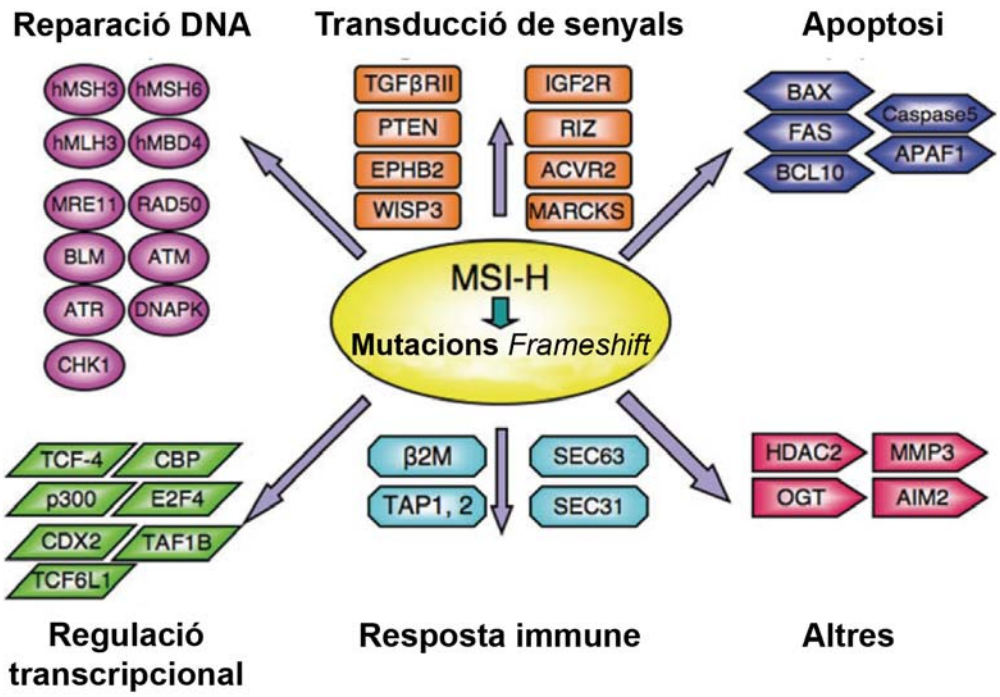


Figura 4. Representació dels gens diana en els tumors amb inestabilitat de microsatèl·lits. Adaptat d'Imai and Yamamoto, 2008.

Via del fenotip metilador d'illes CpG

La metilació d'illes CpG del promotor d'alguns gens caracteritzen l'anomenat fenotip CIMP, descrit per primer cop per Toyota (Toyota, et al., 1999). Els dinucleòtids CG estan localitzats en *clusters* (illes CpG) en la regió del promotor i normalment es troben no metilats. Quan aquestes illes es troben metilades de forma aberrant, típicament en gens supressors tumorals, el resultat és un silenciament transcripcional del gen. Això proporciona un mecanisme alternatiu de pèrdua de funció dels gens supressors tumorals en càncer (Wong, et al., 2007). Alguns gens implicats en la carcinogènesi colorectal es troben silenciat per la hipermetilació del DNA, com poden ser *APC, MCC, MLH1, MGMT*, i d'altres. S'ha descrit, que alguns factors ambientals com el tabaquisme i l'edat avançada correlacionen amb un augment de l'estat de metilació (Samowitz, et al., 2006; Toyota, et al., 1999; Toyota and Issa, 1999).

Encara que el CIMP es va establir com a un fenotip epigenètic únic en el CCR (Ogino, et al., 2007; Samowitz, et al., 2005; Weisenberger, et al., 2006), hi ha una gran controvèrsia sobre els marcadors que el defineixen, l'existència de diferents tipus de CIMP (CIMP 2, CIMP-alt, CIMP-baix) (Grady and Carethers, 2008) i fins i tot la relació d'aquests subtipus amb l'estat de MSI (Kang, 2011) o amb els diferents subtipus epigenètics (alt, intermedi i baix) (Kaneda and Yagi, 2011).

Recentment, s'ha detectat que la majoria (63%) dels tumors colorectals que presenten un genotip hipermetilat mostren silenciament del gen *MLH1*, una elevada freqüència de mutacions en *BRAF* i una baixa freqüència de mutacions en *APC* i *KRAS.* Aquest perfil mutacional és significativament diferent dels tumors que no presenten hipermetilació. Per altra banda, el 17% de les tumors amb hipermetilació no presenten silenciament de *MLH1*, i presenten una elevada taxa de mutacions, així com una incidència elevada de mutacions en els gens *KRAS* i *APC,* però no en el gen *BRAF*. Així doncs, els CCR que presenten hipermetilació inclouen dos subgrups amb unes vies de malignitat diferent. Els autors suggereixen que l'estudi d'expressió del gen *MLH1* i l'anàlisi de mutacions en els gens *APC, KRAS, BRAF* en CCR poden ser d'utilitat pel diagnòstic i podrien influir en les decisions terapèutiques d'aquest tipus de tumors (Cancer Genome Atlas, 2012; Donehower, et al., 2012)**.**

1.5 Història familiar i càncer colorectal

En funció del patró d'herència, el CCR pot ser classificat com a esporàdic, hereditari o familiar. La major part dels casos de CCR corresponen a formes esporàdiques (70-80%), desenvolupant-se per acumulació d'alteracions genètiques al llarg de la vida i on l'estil de vida juga un paper molt important. S'estima que en un 20-25% dels casos de CCR pot existir un component hereditari associat, encara no ben establert, que es coneix com CCR familiar. En aquestes famílies, el CCR es desenvolupa amb una freqüència superior a la del CCR esporàdic però no segueix el patró característic d'una síndrome hereditària (Burt, 2000; Winawer, et al., 1997).

Una petita proporció dels CCR corresponen a formes hereditàries (<5%). Aquestes es poden dividir en dos tipus: les poliposes i les no poliposes. Entre les síndromes poliposes hi trobem la Poliposi Adenomatosa Familiar clàssica (FAP) i la Poliposi Adenomatosa Familiar Atenuada (AFAP) (<1%), la poliposi associada al gen *MYH* (MAP, *MYH Associated Polyposis*) (<1%) i les poliposis hamartomatoses, un grup poc freqüent de poliposis. La síndrome no poliposa més important és la síndrome de Lynch (de la Chapelle, 2004) (Figura 5).

L'estudi de les dues síndromes hereditàries més freqüents –FAP i la síndrome de Lynch– ha resultat de gran importància per la comprensió i maneig del CCR. Així, el seu estudi ha permès conèixer els mecanismes implicats en el desenvolupament del CCR esporàdic, ja que alguns dels gens que es troben mutats a nivell germinal a la FAP i a la síndrome de Lynch, també tenen un paper clau en els casos de CCR esporàdic (Fearon and Vogelstein, 1990). Per altra banda, l'estudi de les síndromes hereditàries també ha

permès establir el diagnòstic presimptomàtic d'individus portadors de mutacions en aquests gens, i per tant, en situació de risc de desenvolupar la malaltia, als quals se'ls recomana el cribratge periòdic del còlon i recte (Giardiello, et al., 2001). Així, el diagnòstic molecular de les formes hereditàries possibilita l'adopció de mesures terapèutiques més radicals que les emprades per les formes esporàdiques, tenint un impacte favorable en el pronòstic d'aquests malalts.

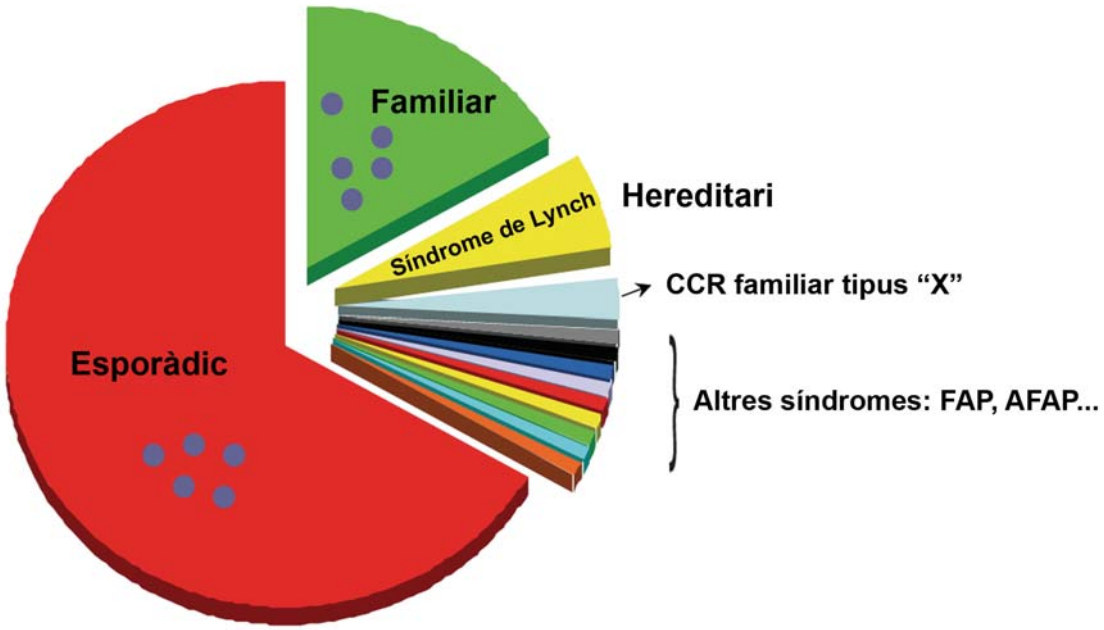


Figura 5. Classificació dels tipus de CCR en funció del patró d'herència. Adaptat de Lynch, et al., 2009.

2. LA SÍNDROME DE LYNCH

La síndrome de Lynch, prèviament anomenada com a càncer colorectal hereditari no associat a poliposis (HNPCC), és la forma de càncer colorectal hereditari més freqüent, representant entre el 2% i el 5% dels CCR depenent de la població estudiada (de la Chapelle, 2004). En un estudi poblacional a Espanya es va detectar que la síndrome de Lynch representava el 0,9 % dels casos de CCR (Pinol, et al., 2005). Es tracta d'una malaltia hereditària autosòmica dominant que té penetrància incomplerta, associada a mutacions o epimutacions germinals en els gens implicats en les vies de reparació d'aparellaments erronis del DNA (MMR o *Mismatch Repair*), concretament en els gens *MLH1, MSH2, MSH6* i *PMS2.* També s'han descrit associacions amb altres gens reparadors (*MLH3, PMS1, EXO I*) (Peltomaki, 2005).

Les cèl·lules portadores d'una mutació germinal en un gen reparador tenen capacitat de reparació dels desaparellaments (Parsons, et al., 1993), però això es veu modificat quan hi ha la inactivació somàtica de l'al·lel salvatge. Aquest inactivació es pot donar, per exemple, per una mutació puntual, per la hipermetilació del promotor, o per una deleció del gen, donant lloc a pèrdua d'heterozigositat (LOH) (Hemminki, et al., 1994). Aquesta inactivació, promourà la tumorogènesis i l'establiment d'un fenotip mutador de caràcter recessiu. Aquest procés va ser proposat per primera vegada de forma teòrica per Knudson com el model dels "*two hits*", per explicar com una alteració genètica en línia germinal podia predisposar a desenvolupar tumors (Knudson, 1971) (Figura 6).

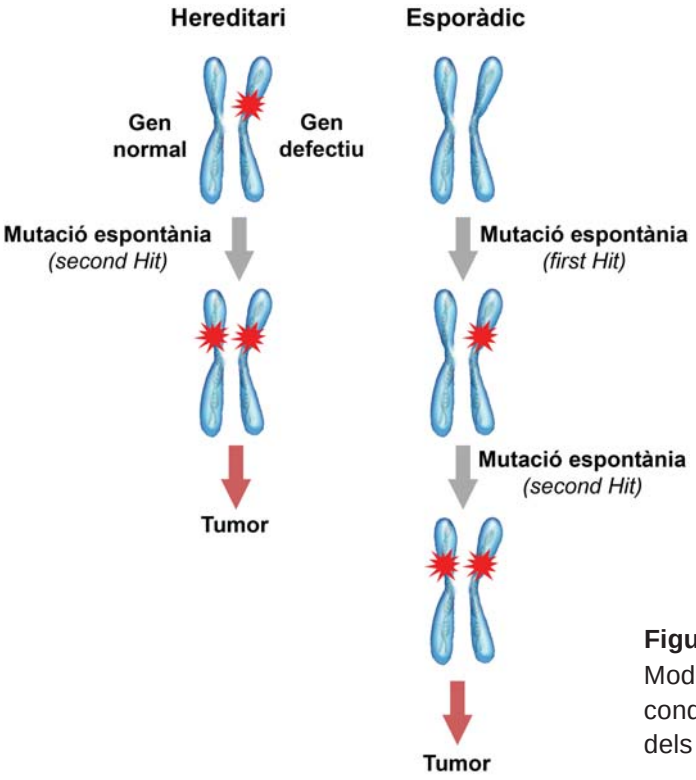


Figura 6. Model dels *"two hits"*. Model proposat per Knudson, que condueix a la inactivació bial·lèlica dels gens reparadors.

2.1 Característiques clíniques

La síndrome de Lynch es caracteritza pel desenvolupament del CCR a una edat més jove, amb una mitjana d'edat de 45 anys enfront dels 67,9 del CCR esporàdic (Hampel, et al., 2005b). Els tumors es localitzen de forma preferent al còlon dret i són més freqüents les neoplàsies múltiples sincròniques (diagnosticades alhora) o metacròniques (diagnosticades més de 6 mesos després de la rescissió quirúrgica). Els tumors colorectals en la síndrome de Lynch presenten unes característiques clinico-patològiques específiques que inclouen: poca diferenciació, característiques mucinoses, cèl·lules en anell de segell, limfòcits infiltrats en el tumor, una intensa reacció limfocitària (reacció de tipus Crohn) i un patró de creixement medul·lar (Risio, et al., 1996).

Els individus amb la síndrome de Lynch també presenten un risc incrementat de desenvolupar tumors extracolònics, com els d'endometri (CE), d'estómac, de budell prim, d'ovari, de tracte biliar, d'uroepiteli, de pàncrees, del sistema nerviós central i els tumors sebacis (Umar, et al., 2004). Alguns estudis també han proposat que altres manifestacions com els sarcomes, el càncer de pròstata i mama també podrien ser inclosos dins de l'espectre de la síndrome de Lynch (den Bakker, et al., 2003; Geary, et al., 2008; Hirata, et al., 2006; Soravia, et al., 2003; Westenend, et al., 2005), tot i així, no hi ha consens entre la comunitat científica.

Els tumors de pacients amb síndrome de Lynch tenen una seqüència adenoma-carcinoma accelerada que dura entre 2-3 anys, mentre que en els casos de CCR esporàdic aquest procés pot tardar entre 7-10 anys. Cal destacar però, que els tumors deguts a la síndrome de Lynch presenten un millor pronòstic que la resta de casos de CCR (Lynch, et al., 2007).

També, s'han identificat individus amb mutacions bial·lèliques als gens reparadors. Aquests presenten la síndrome del dèficit constitucional de reparació MMR (CMMR-D, *Constitutional MMR-deficiency*), que es caracteritza pel desenvolupament de càncers en edat infantil o adolescent, principalment hematològics, del sistema nerviós central i també colorectals (Wimmer and Etzler, 2008). Actualment, es suggereix que les característiques clíniques dels tumors en pacients amb mutacions germinals bial·lèliques en *MLH1* o *MSH2*, difereixen d'aquelles amb mutacions bial·lèliques germinals en *MSH6* o *PMS2*. Els primers presenten amb major freqüència càncers hematològics que tendeixen a desenvolupar-se a la infantesa. En canvi, els pacients amb mutacions bial·lèliques en *MSH6* o *PMS2*, mostren una major prevalença a desenvolupar tumors cerebrals, i tenen una major probabilitat de sobreviure al primer tumor i a desenvolupar-

ne un segon. En general, la prevalença dels tumors associats a la síndrome de Lynch és major en pacients amb mutacions bial·lèliques en *MSH6* o *PMS2*. Aquests factors faciliten el diagnòstic clínic de CMMR-D en pacients amb mutacions en *MSH6* o *PMS2*, i pot explicar en part l'elevada prevalença de mutacions bial·lèliques en *PMS2* en els casos publicats (revisat a Wimmer and Etzler, 2008).

2.2 Risc estimat i penetràncIa

La penetràncIa de la síndrome de Lynch ha estat analitzada en varis estudis amb l'objectiu d'establir el risc de càncer associat en aquests pacients i així poder millorar el seu maneig clínic. En els primers estudis, els riscos estimats de desenvolupar CCR i CE en pacients amb la síndrome de Lynch eren de 70-80% i 40-60%, respectivament (Aarnio, et al., 1995; Aarnio, et al., 1999; Dunlop, et al., 1997; Vasen, et al., 1996; Watson and Lynch, 2001). Aquests treballs utilitzaven dades extretes de bases de dades recollides a través dels registres de Càncer Familiar Europeu. Posteriorment, es va suggerir que aquest risc podia estar sobreestimat, degut a la sobrerepresentació de famílies amb àmplies històries familiars de càncer, entre d'altres biaixos (Quehenberger, et al., 2005). Quan s'utilitzen mètodes estadístics per controlar aquests biaixos potencials, els riscos estimats de CCR i CE es veuen reduïts al 22-47% i 14-30%, respectivament (Alarcon, et al., 2007; Bonadona, et al., 2011; Jenkins, et al., 2006; Quehenberger, et al., 2005). No obstant això, les estimacions de risc en famílies nord-americanes són més elevades (66-47% CCR i 39% CE) (Stoffel, et al., 2009).

També s'ha reportat que el risc de desenvolupar càncer en pacients amb síndrome de Lynch difereix en funció del gen mutat (Bonadona, et al., 2011; Choi, et al., 2009; Ramsoekh, et al., 2009; Stoffel, et al., 2009). La majoria d'aquest estudis mostren com mutacions en els gens *MLH1* i *MSH2*, presenten uns riscs associats més elevats que no pas mutacions en els gens *PMS2* i *MSH6* (Bonadona, et al., 2011; Kempers, et al., 2011; Senter, et al., 2008).

Cal destacar que s'ha descrit el fenomen de l'anticipació genètica en famílies amb la síndrome de Lynch (Larsen, et al., 2009; Nilbert, et al., 2009; Rodriguez-Bigas, et al., 1996; Stella, et al., 2007; Vasen, et al., 1994; Westphalen, et al., 2005). L'anticipació genètica es dóna quan, en generacions successives, es redueix l'edat d'aparició d'una malaltia i/o augmenta la severitat del seu fenotip (Trachan, 1999). La disminució progressiva de l'edat del diagnòstic en el CCR en generacions successives s'ha reportat des de la primera descripció de la síndrome de Lynch el 1925 (Warthin, 1925). Tot i això,

els mecanismes moleculars d'aquest fenomen en la síndrome de Lynch encara no han estat establerts (Bozzao, et al., 2011; Gruber and Mukherjee, 2009).

2.3 Assessorament genètic en la síndrome de Lynch

La identificació dels individus en situació de risc de desenvolupar la síndrome de Lynch, així com qualsevol altre síndrome de predisposició hereditària al CCR, és fonamental per a disminuir la morbiditat i mortalitat de la malaltia. Des de la introducció del diagnòstic molecular predictiu, s'ha facilitat la identificació dels individus en situació de risc que permet, a més d'una valoració individualitzada del risc de desenvolupar càncer, recomanar estratègies de prevenció i cribratge adequades. És per això que després de realitzar el diagnòstic de la síndrome de Lynch en qualsevol membre de la família, s'ha de remetre al pacient i als seus familiars a una Unitat de Consell Genètic per tal de rebre el corresponent assessorament genètic.

En el procés de consell genètic s'avalua el risc personal i familiar de susceptibilitat hereditària al càncer mitjançant la realització d'una exhaustiva història clínica sobre els antecedents familiars i personals de càncer del pacient. En aquest procés, un equip de professionals especialitzats i degudament capacitats en consell genètic (Rolnick, et al., 2011) ajuda als individus i les seves famílies amb totes les qüestions associades a la seva condició hereditària de manera personalitzada, assegurant-los el suport psicosocial si és necessari. Així doncs, l'assessorament genètic s'ha de considerar com una part integral del procés de realització de les proves genètiques (Kääriäinen H, 2008).

Aquest procés inclou construir i avaluar un arbre genealògic, obtenir la història mèdica personal i familiar, i proporcionar informació sobre el risc genètic. Si és necessària la realització d'un estudi genètic, el procés incorpora l'assessorament previ, l'estudi en sí, l'assessorament posterior i el seguiment. Després de l'estudi genètic, a les persones amb un resultat de test positiu (identificació d'una mutació germinal patogènica), se'ls informarà del risc de desenvolupar càncer, tot recomanant-los les estratègies de prevenció i cribratge adequades (Winawer, et al., 2003). En algunes famílies però, el resultat del test genètic és no concloent, degut a la identificació d'una variant de significat desconegut (VSD) (veure apartat 5.1 i 6). Aquest resultat suposa un problema en el consell genètic, ja que es desconeix si aquesta variant afecta la funcionalitat del gen o la proteïna i, en conseqüència, si està o no associada a un augment del risc de desenvolupar càncer en la família.

3. BASES GENÈTIQUES DE LA SÍNDROME DE LYNCH

3.1 Sistema de reparació del DNA

Com ja s'ha comentat, la síndrome de Lynch està causada per mutacions germinals en els gens MMR. El sistema de reparació de desaparellaments erronis és un mecanisme encarregat de corregir els errors produïts per la polimerasa durant el procés de la replicació. La polimerasa pot provocar un desaparellament base-base o bé pot "relliscar" (*slippage*) en les seqüències curtes i repetitives, com els microsatèl·lits, donant lloc a bucles d'inserció/deleció de nucleòtids (Chung and Rustgi, 2003). Si aquests errors no són reparats, en la propera replicació de la cadena filla es fixarà la mutació.

Aquest sistema de reparació va ser inicialment identificat en bactèries, on es va observar que la seva inactivació comportava un augment de mutacions espontànies (Fazakerley, et al., 1986; Modrich and Lahue, 1996; Radman and Wagner, 1986). En eucariotes, com els llevats o els mamífers, hi ha varis homòlegs de les proteïnes del sistema MMR en *E.coli*, que reflecteixen la conservació del procés de reparació així com l'especialització i la superposició del funcionament d'algunes proteïnes (Taula 3). Aquest fet fa pensar en l'elevada importància d'aquest tipus de reparació.

Taula 3. Correlació de les proteïnes del sistema reparador del DNA en *E. coli* i en humans.
NI: No identificada. Modificada de Fukui, 2010.

Proteïnes en *E. coli*	Funció	Proteïnes homòlogues en humans	Funció
MutS	Reconeixement de l'error	MutSα (MSH2, MSH6)	Reconeixment d'error base-base i zones inserció/deleció d'una sola base
		MutSβ (MSH2/MSH3)	Reconeixment d'error en zones d'inserció/deleció >2 bases
MutL	Reclutar i coordinar altres proteïnes	MutLα (MLH1, PMS2)	Reclutar i coordinar diferents proteïnes
		Mut Lβ (MLH1, PMS1)	Reclutar i coordinar diferents proteïnes
		Mut Lγ (MLH1, MLH3)	Reclutar i coordinar diferents proteïnes, també participa en la meiosis
MutH	Creació del *nick*	NI	
SSB	Escisió de la cadena mutada	RPA	Escisió de la cadena mutada
RecJ, EXO I, EXO VII, EXO V	Escisió de la cadena doble de DNA	EXO I	Escisió de la cadena doble de DNA
DNA pol III	Síntesis de la nova cadena	DNA polimerasa δ	Síntesis de la nova cadena

En les cèl·lules eucariotes, s'han identificat tres proteïnes homòlogues al MutS bacterià: MSH2, MSH3 i MSH6. Aquestes formen dos heterodímers que reconeixen el DNA desaparellat: MutSα (MSH2/MSH6) i MutSβ (MSH2/MSH3). MutSα s'uneix

preferentment a bases desaparellades i a zones d'inserció/deleció petites (menys de 2 bases), mentre que MutSβ s'uneix predominantment a les zones d'inserció/deleció de més de 2 bases (Fukui, 2010; Hsieh, 2001; Kolodner and Marsischky, 1999; Kunz, et al., 2009). De la mateixa manera que el MutS bacterià, MSH2 i MSH6 tenen activitat intrínseca d'unió i hidròlisis d'ATP. Aquests proteïnes són imprescindibles per al reconeixement d'errors produïts en la replicació del DNA. Un cop MutS s'ha unit a la zona del desaparellament, es produeix la hidròlisis de l'ATP, que donarà lloc a un canvi conformacional que facilitarà les interaccions entre proteïnes i/o el desplaçament al llarg del DNA (Gradia, et al., 1999). Ara a MutS s'hi uneix un altre heterodímer, normalment MutLα, format per MLH1 i PMS2. També s'han descrit els homòlegs MutLβ, format per MLH1 i PMS1, i MutLγ, format per MLH1 i MSH3, que també està involucrat en la recombinació meiòtica. S'ha observat que MutLγ sembla tenir més afinitat per unir-se a MutSβ (procés de reparació de les insercions/delecions de més de 2 nucleòtids) que no pas a MutLα (Fukui, 2010; Marsischky and Kolodner, 1999). L'heterodimer MutL, que interactuarà amb el complex MutS-DNA, és l'encarregat de reclutar altres proteïnes reparadores i modular la seva activitat (Fukui, 2010; Kunkel and Erie, 2005; Li, 2008). MutLα posseeix una activitat endonucleasa latent (que es troba en la proteïna PMS2), que serà activada per la interacció del MutSα, PCNA *(proliferating cell nuclear antigen)* i RFC (*replication factor C*) en un procés depenent d'ATP. MutL introduirà un tall a la cadena on s'ha produït l'error i això servirà perquè l'exonucleasa I (EXO I) sigui capaç d'escindir nucleòtid a nucleòtid el fragment de DNA on hi ha el desaparellament (Fukui, 2010; Kadyrov, et al., 2006; Kunz, et al., 2009). Mentre EXO I està actuant, la proteïna RPA (*replication protein A*) s'uneix al DNA de cadena senzilla per protegir-lo de les nucleases. Cal remarcar que a vegades el procés de reparació pot donar-se en absència de EXO I (Kadyrov, et al., 2006). La resíntesis d'aquesta cadena la fa la polimerasa δ o polimerasa ε, en presència de PCNA, de la proteïna RPA i de RFC. Finalment, la DNA lligasa I acaba el procés unint els dos extrems del fragment sintetitzat amb la cadena de DNA (Kunkel and Erie, 2005) (Figura 7)**.** La reparació dels errors de replicació es necessària per la transició de la fase S a la fase G2 de la mitosis (Heinen, et al., 2002a).

A part de la seva funció principal en el manteniment de la fidelitat de la replicació, també s'han descrit altres funcions importants associades a aquests gens, com per exemple la modulació de la recombinació del DNA i la facilitació de la senyalització enfront a danys en el DNA, entre d'altres (Altieri, et al., 2008; Jun, et al., 2006; Kolas and Cohen, 2004).

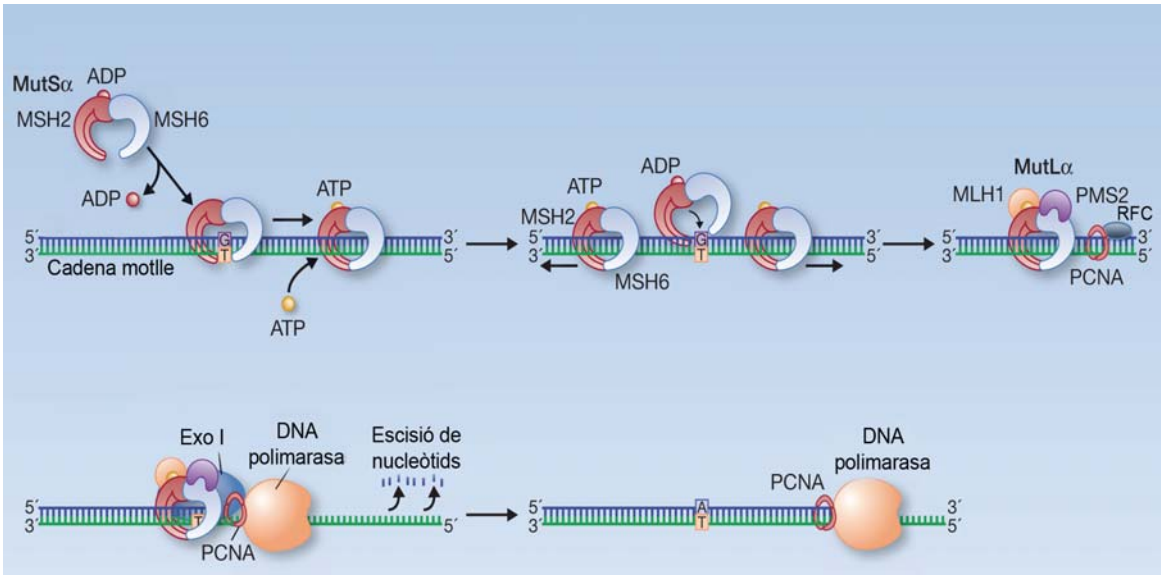
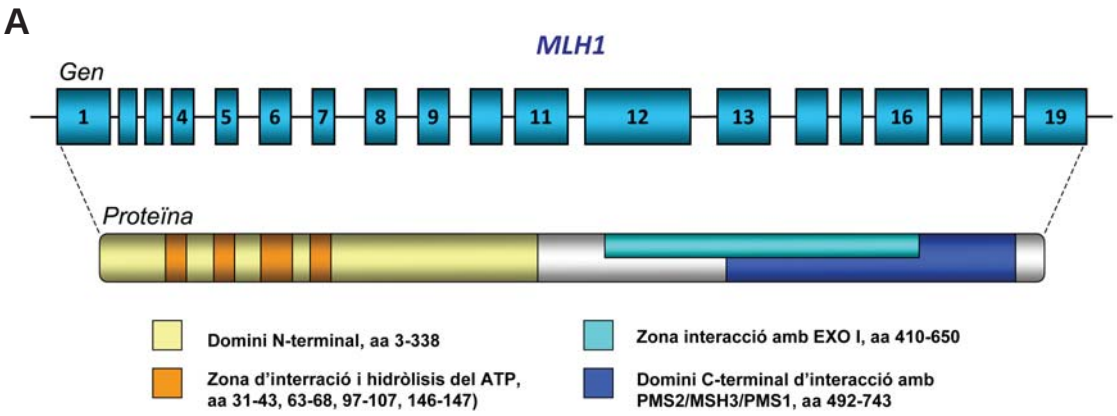


Figura 7. Model de la via de reparació dels errors de desaparellament. Modificat de Sinicrope and Sargent, 2012.

3.2 Els gens reparadors associats a la síndrome de Lynch

MLH1

El gen *MLH1* (NM_00249.3) va ser identificat l'any 1994 per anàlisi de lligament en famílies amb història de càncer colorectal hereditari (Bronner, et al., 1994; Papadopoulos, et al., 1994). Aquest gen es localitza en el braç llarg del cromosoma 3 (3p21.3) i està format per 19 exons i comprèn una regió de 72557 pb. El transcrit de mRNA fa 2752 pb. Aquest gen codifica per una proteïna que consta de 756 aminoàcids amb un pes molecular de 86 kDa (Han, et al., 1995). La proteïna resultant actua com a heterodímer, unint-se normalment a PMS2, i la seva funció principal és reclutar altres proteïnes reparadores i modular la seva activitat. Aquest heterodímer és homòleg a MutL bacterià (Figura 8). La proteïna MLH1 té un domini aminoterminal, que és el responsable de la unió i la hidròlisis de l'ATP, un domini central d'interacció amb MSH2, MSH3 i MSH6 i el domini carboxiterminal, que proporciona un lloc d'interacció necessari per l'heterodimerització amb PMS2, MLH3 i PMS1 (Guerrette, et al., 1999; Kondo, et al., 2001; Raschle, et al., 2002; Tran, et al., 2001) (Figura 8). La proteïna MLH1 presenta senyals de localització nuclears (NLS), ja que la funció de reparació es realitza al nucli (Brieger, et al., 2005; Leong, et al., 2009).



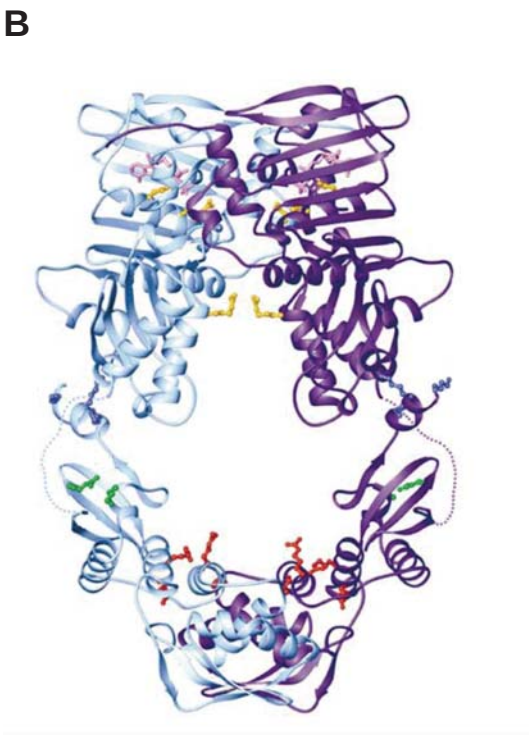


Figura 8. Gen i proteïna MLH1. A. Representació gràfica del gen i de la proteïna MLH1. Modificat d'Andersen, et al., 2012. **B.** Diagrama de l'estructura cristal·litzada de la proteïna MutL de E.Coli. La unió dels dominis N-terminal i C-terminal es representa per una línea de punts. Adaptat de Joseph, et al., 2006.

MSH2

El gen *MSH2* (NM_000251.2) es localitza al cromosoma 2 (2p21), i està composat per 16 exons que inclouen una regió de 80259 pb. El transcrit del mRNA fa 3307 pb, i codifica per una proteïna de 934 aminoàcids. Aquesta proteïna és imprescindible pel correcte reconeixement dels errors produïts en la replicació del DNA, i actua en forma de heterodímer, juntament amb la proteïna MSH6. Aquest heterodímer és homòleg al MutS bacterià (Figura 9). La proteïna MSH2 té varis dominis: el domini d'interacció amb el DNA, localitzat a l'extrem N-terminal; el domini connector, el qual connecta la subunitat d'unió al DNA amb la resta del heterodímer MutS (MSH3 o MSH6) i és responsable de les interaccions intramoleculars i de les senyals al·lostèriques produïdes entre els diferents dominis proteics (Warren, et al., 2007); el domini palanca o *lever*, que connecta el domini d'unió a l'ATP amb el domini abraçadora; el domini abraçadora o

clamp, que és capaç de enviar senyals entre les dues zones de la proteïna (Gammie, et al., 2007); i, finalment, el domini ATPasa, que modula la conformació de la proteïna en funció de la seva unió a l'ATP o ADP (Figura 9). Igual que la proteïna MLH1, pel seu correcte funcionament, la proteïna MSH2 ha de localitzar-se al nucli, encara que actualment hi ha certa controvèrsia sobre el funcionament del seu import nuclear (Christmann, et al., 2002; Gassman, et al., 2011; Hayes, et al., 2009).

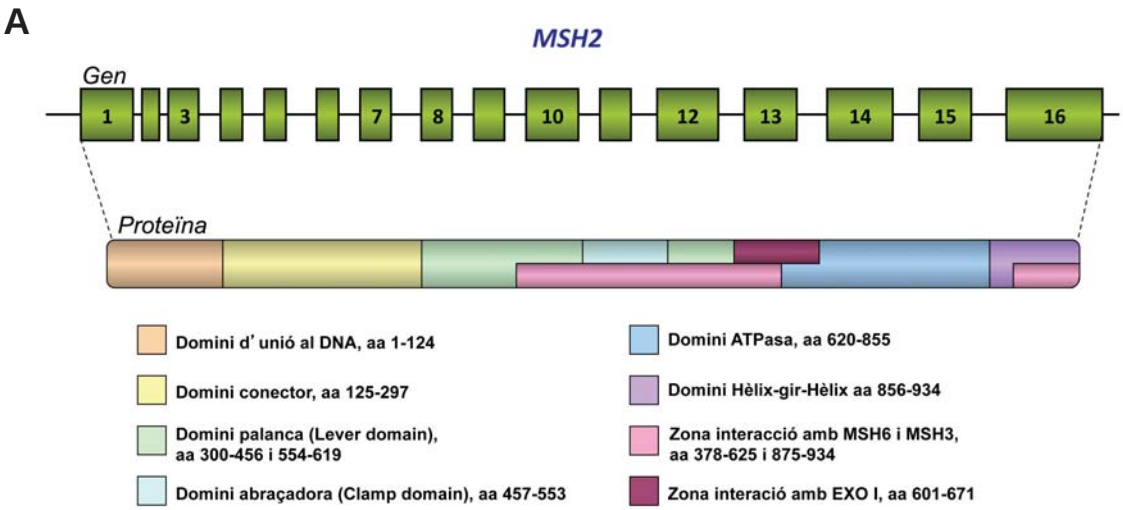


Figura 9. Gen i proteïna MSH2. A. Representació gràfica del gen i la proteïna MSH2. Modificat de Kansikas, et al., 2011. **B.** Diagrama de l'estructura cristal·litzada de la proteïna MutS de *E.Coli*. Es mostren els diferents dominis: domini N-terminal de reconeixement (blau fosc), domini connector (blau clar), domini palanca (en vermell), domini abraçadora (taronja), domini ATPasa (verd, el ADP amb vermell), el domini doble hèlix (groc). L'altre monòmer es mostra en gris, el DNA es mostra en vermell i el desaparellament en groc. Modificat de Joseph, et al., 2006.

MSH6

El gen *MSH6* (NM_000179.2) es localitza en la regió cromosòmica 2p16 i s'estén a llarg de 23871 pb. El seu transcrit de mRNA fa 4328 pb i està constituït per 10 exons, dels quals l'exó 4 ocupa gairebé un 60% de tot el gen (2545 pb). Aquest mRNA codifica per una proteïna de 1360 aminoàcids, amb un pes molecular de 160 kDa. La proteïna

MSH6 actua normalment en forma de heterodímer amb la proteïna MSH2, i la seva funció principal és el reconeixement del desaparellament erroni. Aquesta unió condiciona en part, l'estabilitat de la proteïna MSH6. Aquesta està constituïda de cinc dominis: el domini d'interacció al DNA, el domini connector, el domini palanca o *lever*, el domini abraçadora o *clamp*, i el domini ATPasa (Kansikas, et al., 2011). Aquests dominis tenen les mateixes funcions que els dominis que porten els mateixos noms en la proteïna MSH2. A més, conté una sèrie de residus localitzats al N-terminal, que comprenen la zona d'unió a PCNA, una zona de funció desconeguda i un fragment d'unió no específic al DNA. Recentment, s'ha localitzat un domini globular denominat PWWP, que sembla ser el responsable de la unió a la cadena doble del DNA (Laguri, et al., 2008) (Figura 10). Apart dels diferents dominis, també presenta senyals de localització nuclears (NLS) (Gassman, et al., 2011).

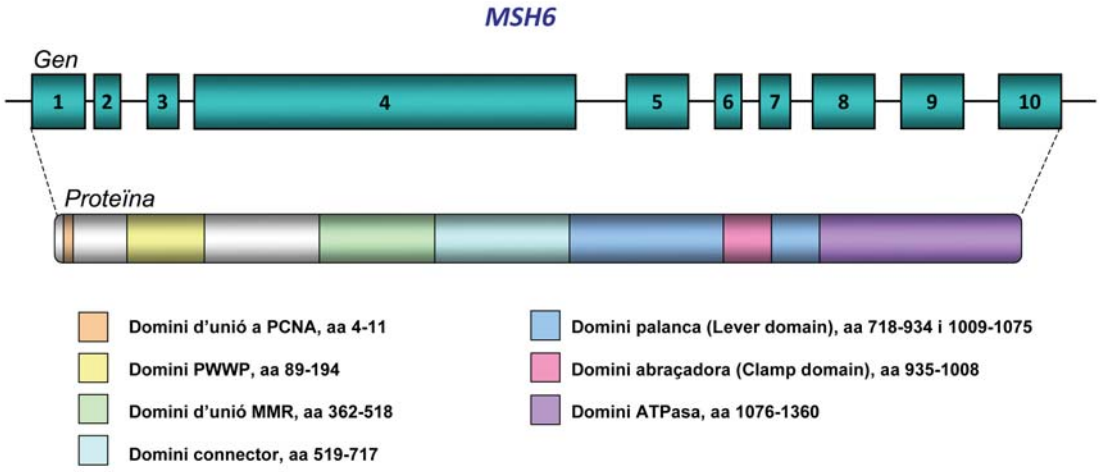


Figura 10. Representació gràfica del gen i la proteïna MSH6. Adaptat de Kansikas, et al., 2011.

PMS2

El gen *PMS2* (NM_000535.5) es localitza en el braç llarg del cromosoma 7 (7p22,2), està format per 15 exons, i comprèn una regió de 35886 pb. El transcrit del mRNA fa 2855 pb. Aquest gen codifica per una proteïna que consta de 862 aminoàcids (96 KDa). La proteïna PMS2 actua normalment unida a MLH1, formant un heterodímer, encarregat de reclutar varies proteïnes per la correcta reparació del error. A més, d'aquesta unió també en depèn en part, la seva localització nuclear (Brieger, et al., 2005) i la seva estabilitat (Chang, et al., 2000). PMS2 té un domini aminoterminal, que és el responsable de la unió i hidròlisis de l'ATP, un domini carboxiterminal, que proporciona un lloc d'interacció per l'heterodimerització amb MLH1 (Raschle, et al., 2002) i un domini amb activitat exonucleasa intrínseca (Kadyrov, et al., 2006) (Figura 11).

L'anàlisi del gen *PMS2*, és especialment complex a causa de la presència en el nostre genoma de múltiples pseudogens. Dotze d'ells comparteixen una elevada homologia en l'extrem 5' de *PMS2* (Clendenning, et al., 2006), mentre que un altre pseudogèn, *PMS2CL*, té una forta homologia en els exons 9 i 11-15 (De Vos, et al., 2004; Nakagawa, et al., 2004) (Figura 11). Per tal d'evitar l'amplificació d'aquests pseudogens s'han dissenyat diferents estratègies per millorar el seu anàlisis mutacional (Clendenning, et al., 2006; Senter, et al., 2008; Vaughn, et al., 2010).

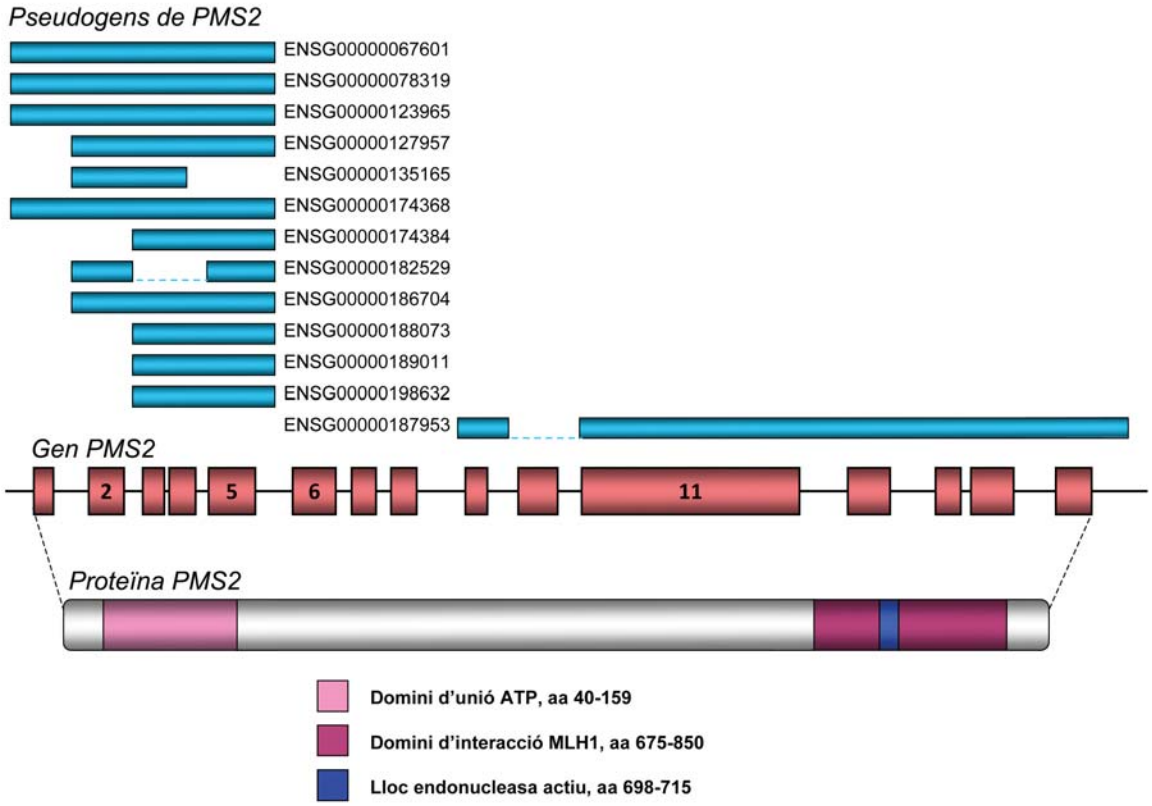


Figura 11. Representació gràfica del gen i la proteïna PMS2, així com els múltiples pseudogens detectats. En vermell es mostra el gen *PMS2* i en blau els seus pseudogens i les regions d'homologia. Adaptat de Clendenning, et al., 2006 i Deschenes, et al., 2007.

4. DIAGNÒSTIC DE LA SÍNDROME DE LYNCH

4.1 Criteris per a la selecció de famílies

Per tal de seleccionar famílies candidates de patir la síndrome de Lynch, el 1991 es van consensuar uns criteris exclusivament clínics per tal d'identificar aquests casos, els anomenats criteris d'Amsterdam I (Vasen, et al., 1991). Més tard, al 1999 es van definir els criteris d'Amsterdam II (Vasen, et al., 1999) on s'incloïen també els càncers extracolònics associats a la síndrome de Lynch (Taula 4). D'aquesta manera s'augmentava la sensibilitat al diagnòstic.

Taula 4. Criteris d'Amsterdam I i II. SL: síndrome de Lynch

CRITERIS D'AMSTERDAM I (l'individu els ha de complir tots)
-Tres o més familiars amb CCR, un d'ells familiar de 1r grau dels altres dos
-Dues generacions successives afectes
-Almenys un tumor diagnosticat abans dels 50 anys
-Exclusió de la Poliposi Adenomatosa Familiar
-Els tumors han de ser confirmats mitjançant un estudi histopatològic
CRITERIS D'AMSTERDAM II (l'individu els ha de complir tots)
-Tres o més familiars amb CCR o un tumor associat a SL (endometri, ovari, intestí prim, tracte biliar, tracte uriniari i sistema nerviós central), un d'ells familiar de 1r grau dels altres dos
-Dues generacions successives d'afectes
-Almenys un tumor diagnosticat abans dels 50 anys
-Exclusió de la Poliposi Adenomatosa Familiar
-Els tumors han de ser confirmats mitjançant un estudi histopatològic

Paral·lelament, amb la identificació de la inestabilitat de microsatèl·lits i la seva elevada freqüència en tumors associats a la síndrome de Lynch, es van proposar al 1997 els anomenats criteris de Bethesda (Rodriguez-Bigas, et al., 1997) (Taula 5), que pretenen identificar pacients amb una elevada probabilitat de patir la síndrome de Lynch, als quals estaria indicat determinar la presència del fenomen MSI en tumor. Més recentment, es van revisar aquests criteris de Bethesda, proposant unes noves recomanacions per a la selecció dels pacients. Aquests són els anomenats criteris de Bethesda revisats (Umar, et al., 2004) (Taula 5).

A part dels criteris clínics esmentats, existeixen models predictius que s'utilitzen clínicament per predir el risc d'un individu a patir la síndrome de Lynch: són el PREMM (Balmana, et al., 2006), MMRpro (Chen, et al., 2006), i MMR-Predict (Barnetson, et al.,

2006). Aquests utilitzen l'edat al diagnòstic, la història personal i familiar, la localització del tumor i la informació de la MSI (en el cas del MMRpro i MMRpredict) per predir la probabilitat de patir la síndrome de Lynch. Aquests models però encara no s'usen habitualment en l'algoritme diagnòstic de la síndrome de Lynch.

Taula 5. Criteris de Bethesda i criteris de Bethesda revisats. SL: síndrome de Lynch, MSH-I: Inestabilitat de microsatèl·lits elevada.

CRITERIS DE BETHESDA (l'individu ha de complir un dels requisits)
-Individus amb CCR que pertanyen a famílies que compleixen els criteris d'Amsterdam
-Individus amb dos neoplàsies associades a SL, incloent CCR sincrònic i metacrònic o càncer extracolònic (endometri, ovari, gàstric, hepatobiliar, budell prim, urèter o pelvis renal)
-Individus amb CCR i un familiar de primer grau amb CCR i/o neoplàsia extracolònica associada al SL i/o adenoma colorectal; un dels càncers diagnosticat abans dels 45 anys d'edat o l'adenoma abans dels 40 anys d'edat
-Individus amb CCR o càncer d'endometri diagnosticat abans dels 45 anys d'edat
-Individus amb CCR localitzat al còlon dret i histològicament indiferenciat diagnosticat abans dels 45 anys d'edat
-Individus amb CCR tipus cèl·lules en anell de segell (format per més del 50% de cèl·lules en anell de segell), mucinós o patró de creixement medul·lar diagnosticat abans dels 45 anys d'edat
-Individus amb adenoma colorectal diagnosticat abans dels 40 anys d'edat
CRITERIS DE BETHESDA REVISATS (l'individu ha de complir un dels requisits)
-CCR diagnosticat abans dels 50 anys
-Presència de CCR , metacrònic o un altre càncer associat a la síndrome de Lynch (endometri, ovari, gàstric, hepatobiliar, budell prim, urèter o pelvis renal), independentment de l'edat
-CCR amb histologia de MSI-H (presència de limfòcits tumorals infiltrats, reacció limfocitaria com la de Crohn, diferenciació mucinosa/anell de segell o patró de creixement medul·lar) diagnosticat abans dels 60 anys
-CCR diagnosticat en un o més familiars de primer grau amb un tumor associat, sent un dels càncers diagnosticats abans dels 50 anys
-CCR diagnosticat en dos o més familiars de primer o segon grau amb tumors associats a la síndrome de Lynch

4.2 Cribratge molecular dels tumors

El cribratge molecular dels tumors és essencial per l'avaluació de la síndrome de Lynch, i pretén identificar individus candidats a l'estudi genètic dels gens reparadors (Hendriks, et al., 2006a). Per aquest cribratge, s'analitza la inestabilitat de microsatèl·lits, l'expressió de les proteïnes reparadores, així com, en els casos amb pèrdua d'expressió de la proteïna MLH1, la hipermetilació del promotor del gen *MLH1* i/o la detecció de la mutació p.V600E del gen *BRAF*.

Anàlisi d'inestabilitat de microsatèl·lits

L'any 1997 l'Institut Nacional del càncer dels Estats Units va recomanar utilitzar un panell de cinc marcadors per detectar la inestabilitat de microsatèl·lits (MSI) en tumors colorectals: BAT26, BAT25, D2S123, D5S346 i D17S250. Els dos primers són marcador mononucleòtids, mentre que la resta són dinucleòtids (Boland, et al., 1998). D'aquesta manera es va determinar que sempre que els tumors mostrin dos o més microsatèl·lits alterats es classificaran com tumors amb inestabilitat alta (MSI-H) i quan els tumors mostrin una alteració en un sol marcador es classificaran com a inestabilitat baixa (MSI-L). Quan cap dels marcadors és positiu, la mostra es classifica com a estable (MSS) (Figura 12). Més recentment, un panell de cinc marcadors mononucleòtids (BAT26, BAT25, NR21, NR22, i NR24) ha resultat ser més sensible per detectar MSI-H que d'altres marcadors microsatèl·lits, a més de permetre realitzar l'anàlisi del tumor sense el teixit normal aparellat (Suraweera, et al., 2002).

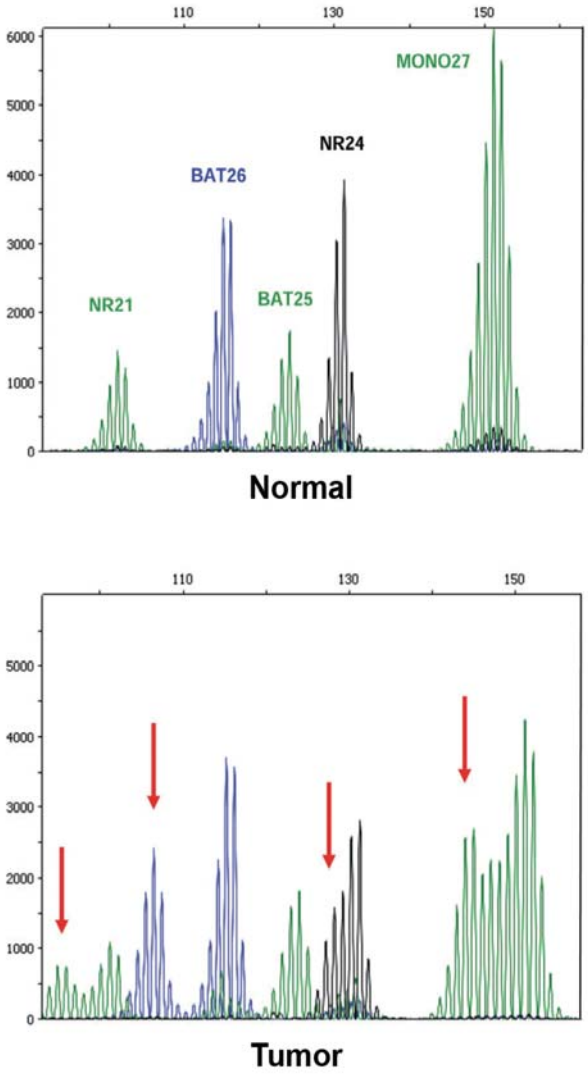


Figura 12. Estudi de la inestabilitat de microsatèl·lits en una mostra amb MSI-H. En la mostra s'observa l'anàlisi de 5 microsatèl·lits mononucleòtids en texit normal (quadre superior) en comparació amb el teixit tumoral (quadre inferior). Les fletxes vermelles indiquen delecions en 4 dels mononucleòtis analitzats. Extret de Iacopetta, et al., 2010.

Inmunohistoquímica

L'any 1994 es va començar la comercialització d'anticossos monoclonals contra les proteïnes MLH1 i MSH2. La inmunohistoquímica (IHQ) d'aquestes proteïnes es va proposar com a alternativa als estudis de microsatèl·lits en el diagnòstic (Debniak, et al., 2000; Dietmaier, et al., 1997; Thibodeau, et al., 1998). Posteriorment es va aconsellar realitzar la IHQ de les quatre proteïnes reparadores (MLH1, MSH2, MSH6 i PMS2) implicades en el síndrome de Lynch (Vasen, 2007).

Quan es realitza la IHQ en el teixit tumoral d'un individu amb sospita de síndrome de Lynch, la pèrdua d'expressió d'una o vàries proteïnes reparadores, suggereix que pot haver-hi hagut una inactivació germinal del gen corresponent i una posterior inactivació somàtica *(second hit)* (Figura 13).

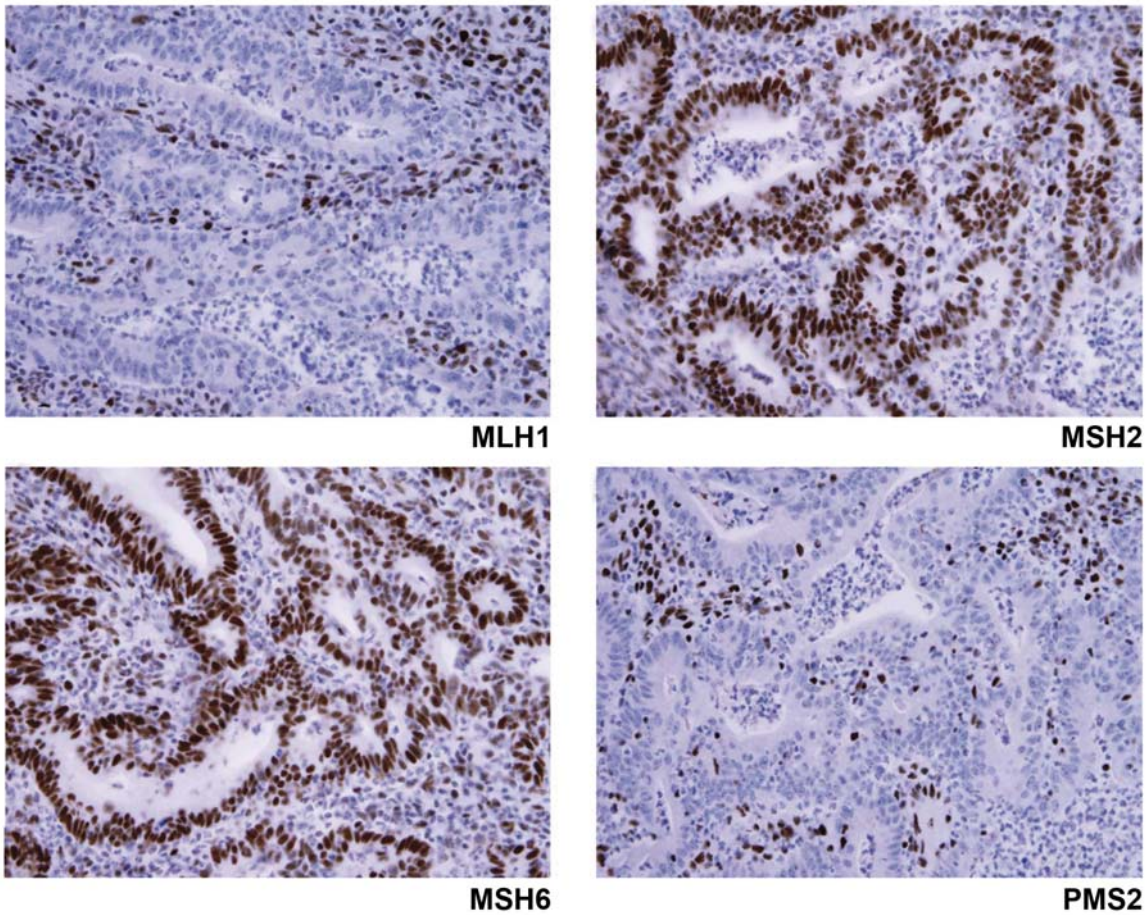


Figura 13. Inmunohistoquímica en de les proteïnes reparadores. En la mostra s'observa pèrdua d'expressió de les proteïnes MLH1 i PMS2. Extret de Bellizzi and Frankel, 2009.

Hipermetilació del promotor del gen MLH1 i la detecció de la mutació p.V600E en el gen BRAF

La inestabilitat de microsatèl·lits constitueix un marcador fenotípic de la síndrome de Lynch, però també es pot trobar en el 8-15% dels tumors esporàdics, on està fortament associada a la hipermetilació somàtica del promotor del gen *MLH1* i amb menor grau la mutació de *BRAF* p.V600E (Figura 14) (Herman, et al., 1998; Weisenberger, et al., 2006), com ja s'ha explicat anteriorment al apartat 1.4. Per aquest motiu, tant la hipermetilació del promotor de *MLH1* com la presència de la mutació *BRAF* p.V600E han estat utilitzades en la selecció de pacients amb CCR probablement esporàdic, els quals seran exclosos de l'anàlisi germinal de gens reparadors (revisat a Gausachs, et al., 2012). Ocasionalment però s'han detectat pacients amb síndrome de Lynch i mutacions en *BRAF* (Gausachs, et al., 2012; Walsh, et al., 2009) o hipermetilació del promotor de *MLH1* (Alemayehu, et al., 2008; Deng, et al., 2004; Nagasaka, et al., 2004; Rahner, et al., 2008).

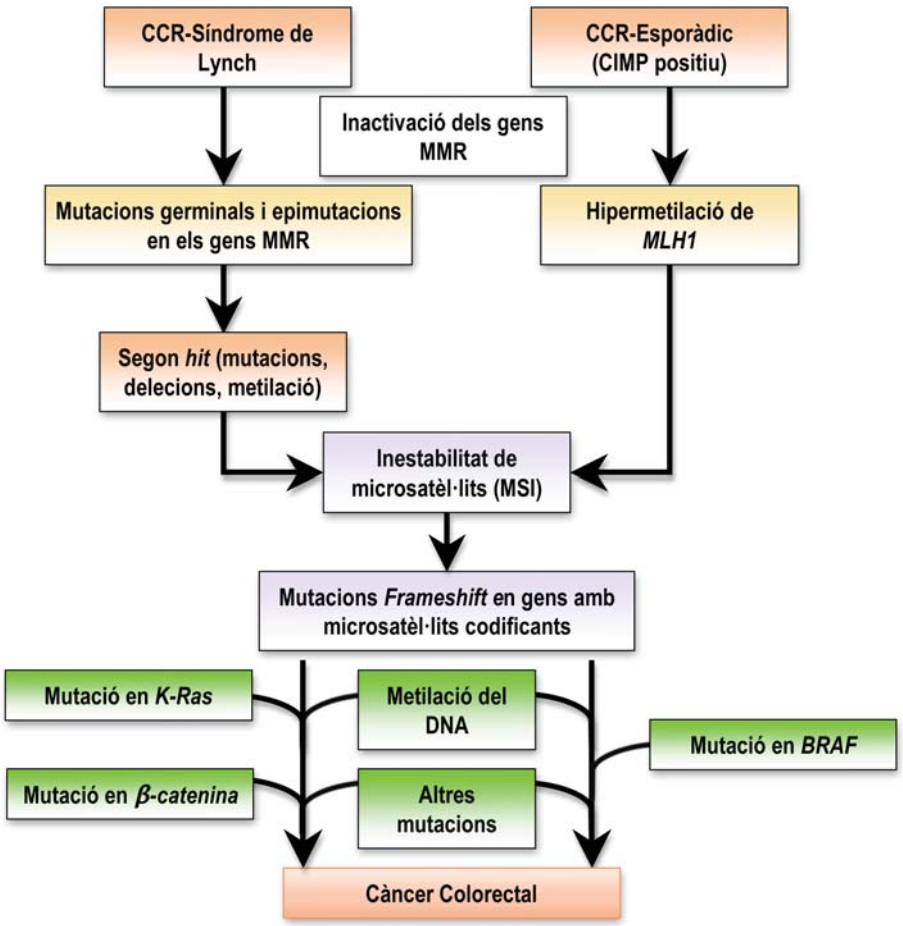


Figura 14. Esquema de les dues vies de desenvolupament del CCR amb MSI. Adaptat de Boland and Goel, 2010.

Estudis recents han demostrat que l'especificitat de l'estudi d'hipermetilació de *MLH1* és superior al de l'anàlisi de la mutació p.V600E de *BRAF* per la identificació dels tumors esporàdics i que l'estat de metilació del promotor de *MLH1* en el DNA dels tumors amb pèrdua d'expressió de MLH1 mitjançant la tècnica de MS-MLPA (Methylation-specific MLPA) és cost-efectiva (Gausachs, et al., 2012; Perez-Carbonell, et al.).

Cribratge poblacional

Actualment, hi ha un intens debat sobre quin mètode de cribratge utilitzar per la detecció de pacients amb síndrome de Lynch. Aquest debat està basat en estudis poblacionals (Hampel, et al., 2008; Hampel, et al., 2005a; Moreira, et al., 2012), on es planteja si és millor la utilització dels criteris d'Amsterdam o Bethesda en front del cribratge molecular, mitjançant estudi d'IHQ o MSI, a tots els casos incidents de CCR i CE, l'anomenat cribratge poblacional. Alguns grups suggereixen que el cribratge poblacional resultaria més cost-efectiu (Hampel, 2010; Hampel and de la Chapelle, 2011; Mvundura, et al., 2010). Tot i així, s'ha de tenir en compte, no només els costos, sinó també els aspectes ètic-legals sobre la transcendència de la informació que aporta el cribratge poblacional pel subjecte i la seva família sense l'assessorament genètic previ.

4.3 Algoritme diagnòstic de la síndrome de Lynch

Amb les premisses explicades en els punts anteriors, a la Figura 15 es mostra en detall un algoritme proposat per a la identificació de pacients amb síndrome de Lynch. Després de la selecció inicial segons els criteris clínics, les proves de IHQ i MSI en teixit tumoral són utilitzades com a estudis complementaris en la selecció de pacients candidats per l'estudi mutacional. Segons els criteris clínics que compleixen les famílies, es recomana la realització de IHQ i/o MSI com a primer pas (Vasen, et al., 2007). Cal destacar, que en el cas que es detecti manca d'expressió de la proteïna MLH1, es realitzarà l'anàlisi de la hipermetilació del seu promotor, per tal de d'avaluar la possibilitat que es tracti d'un CCR esporàdic. Si el resultat de la hipermetilació és negatiu es procedirà a l'anàlisi mutacional del gen *MLH1*. Mentre que si es detecta hipermetilació, es descartaria l'anàlisi mutacional del gen *MLH1*. Per altra banda, en el cas que el pacient presenti hipermetilació de *MLH1* però tumors a edat jove o múltiples tumors, pot considerar-se oportú procedir a l'anàlisi de mutacions en *MLH1* o metilació germinal de *MLH1* (Hitchins and Ward, 2009).

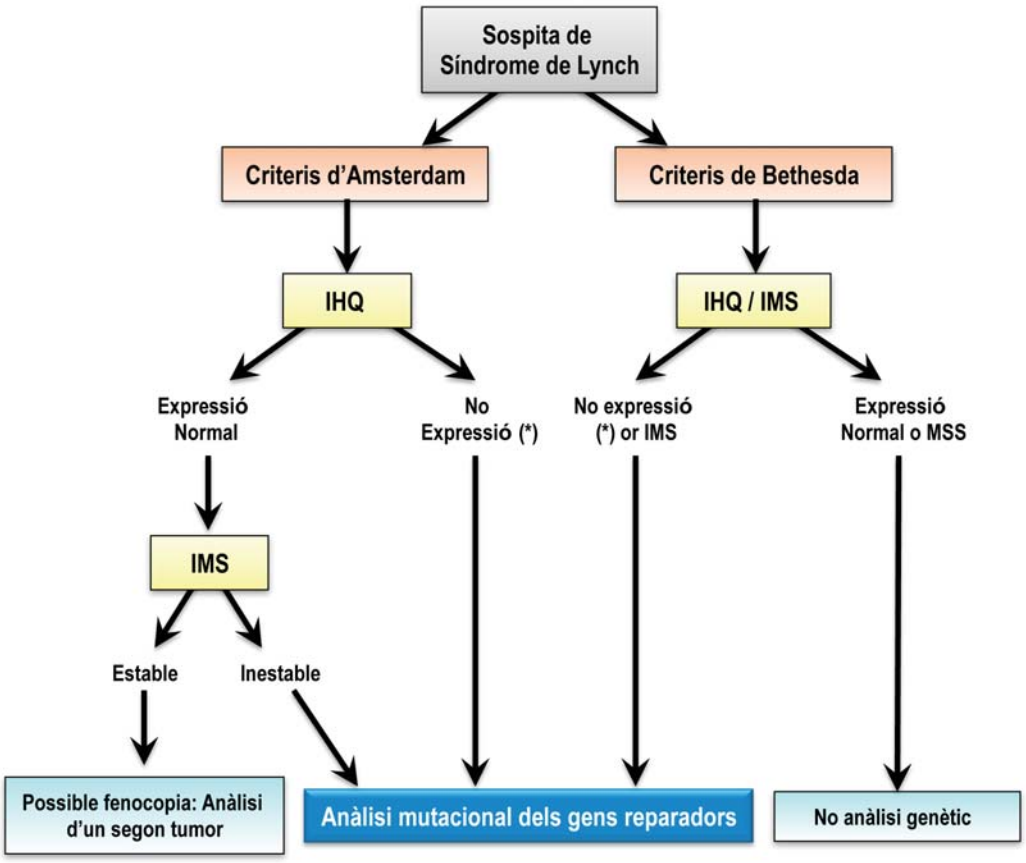


Figura 15. Algoritme diagnòstic proposat per l'estudi de la síndrome de Lynch. (*) En cas de manca d'expressió de MLH1, es realitza l'anàlisi d'hipermetilació de *MLH1* o de la mutació p.V600E de *BRAF*. Extret de Pineda, et al., 2010.

4.4 Estratègia d'estudi genètic

Independentment de la disponibilitat dels algoritmes de treball clínics i moleculars per seleccionar els pacients per a l'anàlisi de mutacions d'un gen en línia germinal, la seqüenciació completa de la regió codificant i les regions intròniques adjacents és la tècnica considerada *gold standard* en l'estudi de mutacions.

Durant la dècada passada molts dels laboratoris de diagnòstic van confiar en les tècniques de cribratge de mutacions puntuals d'alt rendiment, per tal d'obtenir una detecció més ràpida i econòmica de les mutacions. En aquest sentit, es van desenvolupar diferents tècniques basades en les propietats dels heterodúplex com el DGGE (*Denaturing Gradient Gel Electrophoresis)* (Fischer and Lerman, 1979), el CSCE (*Conformation Sensitive Capillary Electrophoresis*) (Davies, et al., 2006), o el DHPLC (*Denaturing High Performance Liquid Chromatography*) (Xiao and Oefner, 2001) entre d'altres. També s'han utilitzat les propietats de la cadena senzilla del DNA pel cribratge, és a dir SSCP (*Single-Strand Conformation Polymorphism*) (Orita, et al., 1989). Més

recentment, s'ha demostrat la utilitat de les tècniques a temps real per cribrar mutacions com ara el *High Resolution Melting* (HRM) (Liew, et al., 2004).

Els fragments identificats com a alterats amb aquestes eines de cribratge mutacional necessiten ser estudiats amb més detall, de manera que, en general, es requereix de la seqüenciació del DNA per identificar l'alteració concreta de cada fragment. Actualment, el cost de la seqüenciació s'està reduint, de manera que molts laboratoris estan analitzant els gens MMR mitjançant la seqüenciació directa sense un cribratge previ (Gruber, 2006). A més, el resultat de l'anàlisi immunohistoquímic de proteïnes reparadores en tumor s'utilitza en la selecció del gen candidat a analitzar genèticament. Cal remarcar però, que tant la seqüenciació directa com el cribratge previ són mètodes que no estan dissenyats per identificar delecions o duplicacions d'un o múltiples exons consecutius. El MLPA (*Multiplex Ligation dependent Probe Amplification*) és molt utilitzat en l'anàlisi de variacions en nombre de còpies a causa de la seva simplicitat relativa, el seu cost acceptable i la possibilitat d'obtenir un alt rendiment i robustesa (Schouten, et al., 2002). Sempre que sigui possible, és preferible confirmar qualsevol variació en el nombre de còpies per una altra metodologia diferent així com buscar la naturalesa molecular exacta de la deleció.

Per altra banda, quan es coneix l'espectre mutacional d'un gen en una població determinada és possible el disseny de determinades estratègies cost-efectives per a la detecció de mutacions fundadores o recurrents (Lynch, et al., 2009). L'algoritme proposat per l'anàlisi genètic de gens reparadors en la nostra població s'esquematitza a la Figura 16. L'estratègia es basa en el cribratge de grans reordenaments previ al cribratge de mutacions puntuals, degut a l'elevada freqüència d'aquests reordenaments en els gens reparadors (10-27%) (veure apartat 6.1, espectre mutacional dels gens reparadors). En cas de detectar un reordenament genòmic classificat com a patogènic es pot obviar el cribratge de mutacions puntuals (Gille, et al., 2002).

Actualment, el mètode convencional de seqüenciació comença a ser substituït parcialment per les tecnologies de *next-generation sequencing* (NGS). Recentment s'ha reportat la utilització del NGS en el diagnòstic del síndrome del càncer de mama i ovari hereditaris (Feliubadalo, et al., (In press); Hernan, et al., 2012), així com del càncer colorectal hereditari (Pritchard, et al., 2012). El NGS ens obre una nova etapa on s'ampliarà la informació del genoma humà així com el coneixement específic del genoma del càncer.

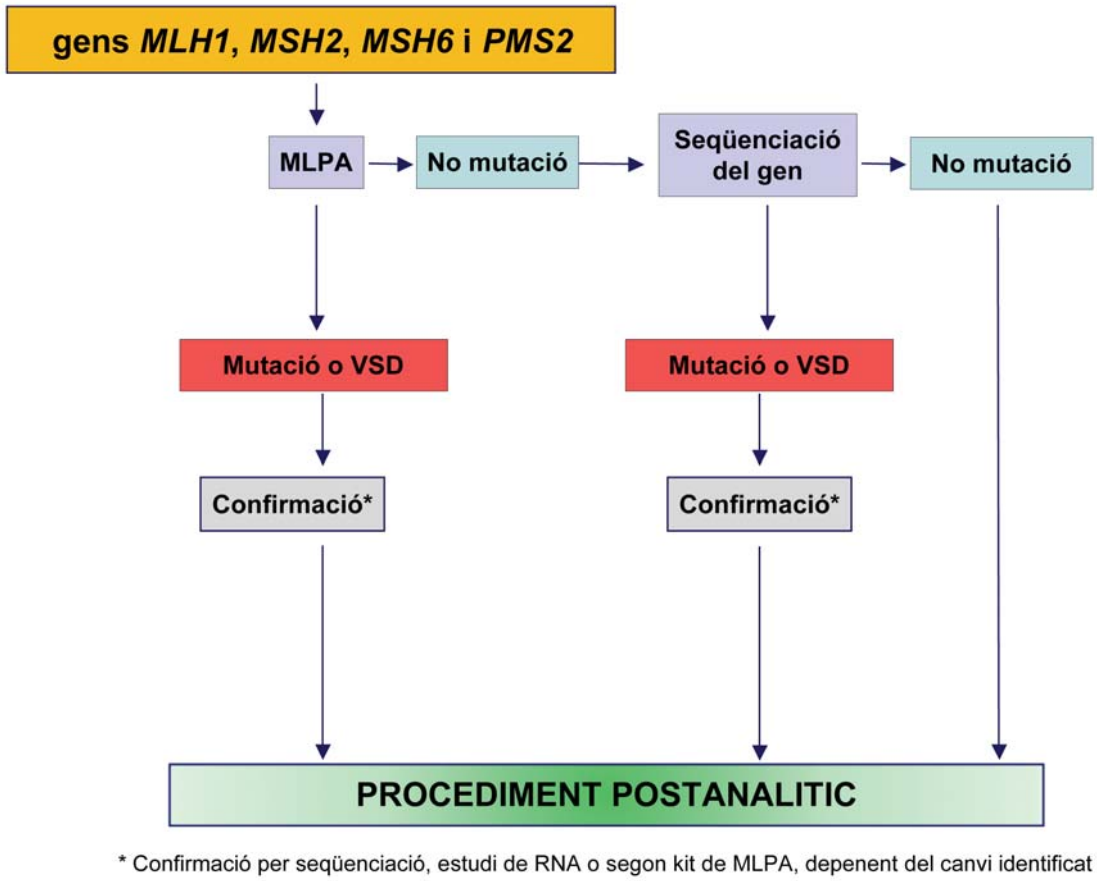


Figura 16. Algoritme de l'anàlisi genètic dels gens reparadors. Algoritme proposat en el Programa de Diagnòstic Molecular de Càncer Hereditari de l'Institut Català d'Oncologia (ICO). Cedit per la Unitat de Diagnòstic Molecular de l'ICO.

5. MUTACIONS EN ELS GENS REPARADORES

5.1 Espectre mutacional dels gens reparadors

El 90% de pacients amb la síndrome de Lynch presenten mutacions en *MLH1* o *MSH2*, i la resta en *MSH6* i *PMS2* (Lynch and de la Chapelle, 2003; Peltomaki and Vasen, 2004)(Figura 17). Els tipus de mutacions que es donen en els gens reparadors són molt variades i es distribueixen al llarg de la seqüència sense presència de *hotspots*. El 80% de les mutacions genètiques reportades en els gens reparadors són úniques, és a dir, reportades en una sola família (Peltomaki and Vasen, 2004). Tot i així, també s'han identificat diverses mutacions recurrents, algunes d'elles fundadores (veure apartat següent).

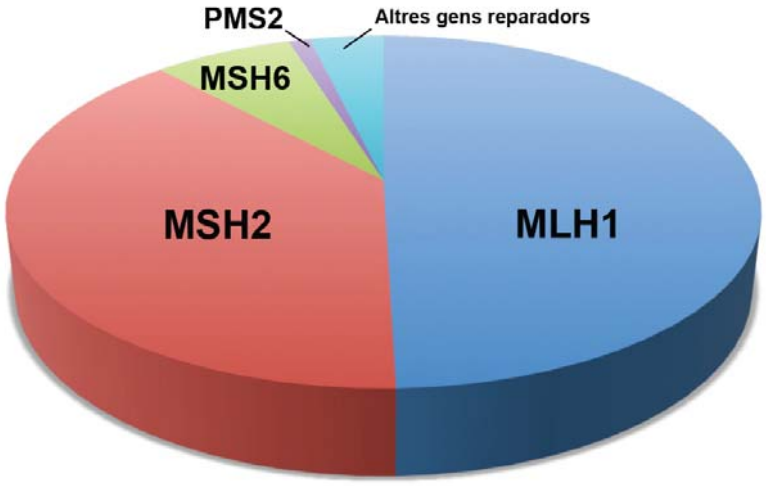


Figura 17. Distribució de mutacions en els gens reparadors. Adaptat de Peltomaki and Vasen, 2004.

La majoria de les mutacions germinals identificades en els gens reparadors són de tipus *nonsense, frameshift* o bé afecten la zona consens de *splicing* (posicions intròniques ±1 i ±2) (Figura 18). Els canvis *nonsense* són canvis que causen l'aparició directa d'un codó de terminació, mentre que les mutacions *frameshift* són insercions o delecions d'una o poques bases que causen el desplaçament de la pauta de lectura. Aquest dos tipus de mutacions, s'associen al truncament de la proteïna i pèrdua de la seva funció (Peltomaki and Vasen, 2004; Woods, et al., 2007), i es classifiquen com a patogèniques. Les mutacions que alteren el correcte processament del mRNA (mutacions de *splicing*), poden donar lloc a una proteïna truncada o a la pèrdua d'un domini o zona important per la seva funció (Auclair, et al., 2006; Lastella, et al., 2006; Spurdle, et al., 2008; Tournier, et al., 2008), i existeix un consens gairebé general de considerar-les mutacions patogèniques.

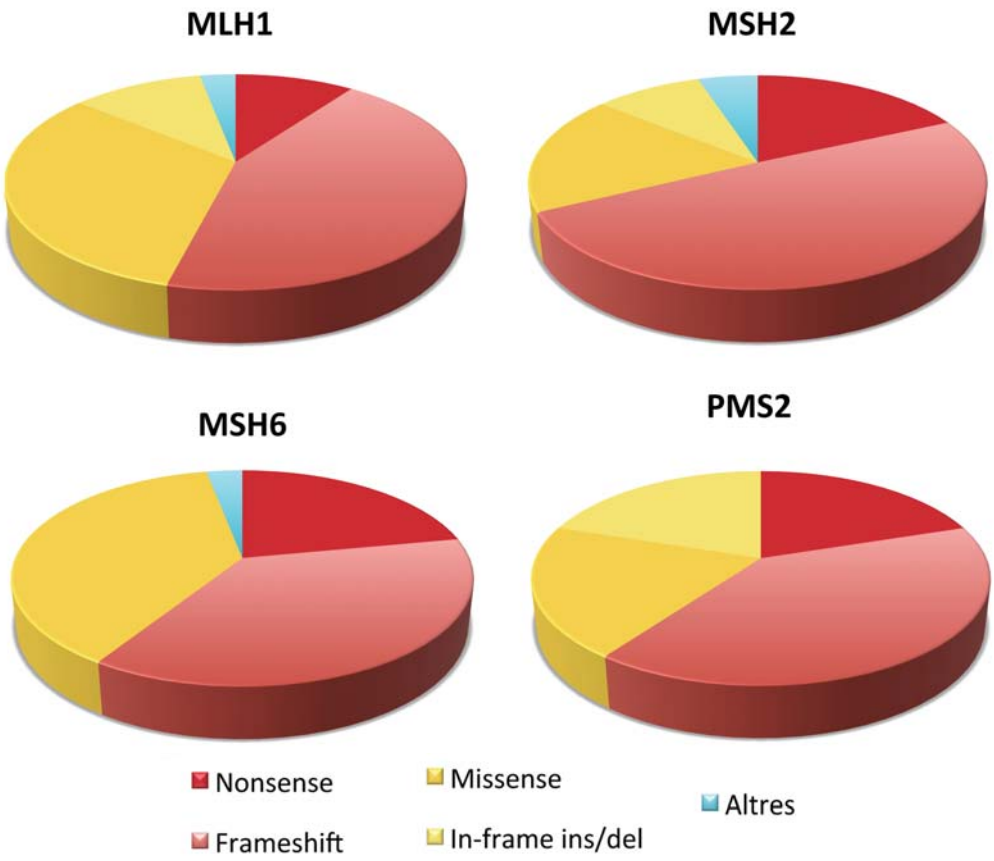


Figura 18. Espectre mutacional dels gens reparadors. Adaptat de Peltomaki and Vasen, 2004.

Posteriorment a Peltomaki s'ha reportat que entre un 10-27% de totes les mutacions patogèniques detectades en la síndrome de Lynch són causades per grans reordenaments, és a dir, pèrdues, amplificacions o reordenaments de fragments grans d'un o varis gens. Aquests reordenaments poden donar lloc a la pèrdua total de la proteïna (el que seria una haploinsuficiència) o bé a una modificació d'aquesta, donant lloc a una proteïna no funcional (Gylling, et al., 2009; van der Klift, et al., 2005; Woods, et al., 2007).

En canvi, una proporció significativa de casos, al voltant del 30% de les alteracions detectades (Peltomaki and Vasen, 2004), són canvis d'aminoàcid (*missense*) o canvis en zones no codificants (intrònics en una regió no consens de *splicing*, en el promotor o en el 3'UTR) que tenen sovint un efecte biològic desconegut (Figura 18). Aquestes són les anomenades variants de significat desconegut (VSD).

Actualment, existeixen diferents bases de dades on són reportats els canvis detectats en els gens reparadors. La *Leiden Open Variation Database* (LOVD) (http://chromium.liacs.nl/LOVD2/colon_cancer/home.php) és la base de dades de canvis de referència per tots els gens de susceptibilitat al CCR. En aquesta base de dades

figuren moltes de les alteracions reportades en els gens reparadors, tant les variants causants de la malaltia, com els canvis polimòrfics i les VSD. Addicionalment, hi ha dues bases de dades, on hi ha reportades diferents VSD estudiades en els gens reparadors, són la *MMR Gene Unclassified Variants Database* (http://www-mmrmissense.net) i la *MMR Genes Variants Database (*http://www.med.mun.ca/ mmrvariants).

En una proporció petita de casos s'han identificat epimutacions germinals o constitutives com a causa de Síndrome de Lynch. Aquests són canvis estables en l'expressió gènica que no afecten a la seqüència del DNA i que es donen en el teixit normal d'un individu. Fins al moment, s'han descrit epimutacions germinals en els gens reparadors *MLH1* i *MSH2* (revisat a Hitchins and Ward, 2009). En alguns casos s'han identificat alteracions genètiques responsables de la epimutació. Així, les epimutacions constitucionals en el gen *MSH2* s'han vist associades a la deleció d'exons en l'extrem 3' del gen *TACSTD1/EPCAM*, localitzat a 5' de *MSH2* i que codifica per la proteïna EPCAM (Ligtenberg, et al., 2009). En canvi, per la majoria de les epimutacions constitucionals del gen *MLH1*, no es coneix encara el mecanisme molecular pel qual es generen, ni tampoc el seu patró d'herència (Hitchins and Ward, 2009).

5.2 Mutacions fundadores en els gens reparadors

Les mutacions que trobem amb elevada freqüència en l'espectre mutacional d'un gen poden ser mutacions recurrents, és a dir, que ocorren repetidament però provenen d'esdeveniments mutacionals independents (*de novo*), o bé fundadores, que són aquelles que s'han donat en un únic individu o bé han estat introduïdes en una població per aquest únic individu. La probabilitat d'aparició d'aquestes mutacions augmenta si hi ha aïllament (absència d'immigració), endogàmia, creixement ràpid de la població o colls d'ampolla al tamany poblacional, ja que aquests fets acceleren la deriva genètica. La capacitat d'una mutació fundadora en els gens reparadors de persistir i arribar a estendre's en una població, és en part deguda al fet que la reproducció dels individus portadors no es veu afectada (de la Chapelle, 2004).

Els finlandesos, els islandesos, els jueus ashkenazi, els canadencs francesos i els amish, són poblacions conegudes com a poblacions fundadores típiques degut a que mostren alguna de les característiques descrites anteriorment. En algunes d'aquestes poblacions, les mutacions fundadores en els gens reparadors representen una part important del total de casos de síndrome de Lynch. Per exemple a Finlàndia, dues mutacions en el gen *MLH1* (la deleció de l'exó 16 i la c.454-1G>A) són les responsables del 63% de totes les mutacions causants de la síndrome de Lynch (Moisio, et al., 1996;

Nystrom-Lahti, et al., 1995). En la Taula 6 es mostren les mutacions fundadores descrites en gens reparadors en diferents poblacions fins al moment. La naturalesa molecular de les mutacions fundadores descrita en aquests gens és diversa, i s'observa un ampli espectre que inclou des de mutacions puntuals fins a grans reordenaments. El possible caràcter fundador d'aquestes mutacions s'ha de confirmar mitjançant l'anàlisi de l'haplotip associat en diferents famílies portadores (Chao, et al., 2008; Gonzalez-Perez and Lopez-Bigas, 2011; Ng and Henikoff, 2002).

Taula 6. Mutacions fundadores reportades en la síndrome de Lynch.

MLH1		
Població	**Mutació**	**Referències**
Finlàndia	Deleció exó 16	Nystrom-Lahti, et al., 1995; Moisio, et al., 1996
Finlàndia	c.454-1G>A	Nystrom-Lahti, et al., 1995; Moisio, et al., 1996
Suïssa	c. 2142G>A (p.W714X)	Hutte,r et al., 1996
Dinamarca	c.1667+2del7ins4	Jager, et al., 1997
Xina	Deleció exó 11	Chan, et al., 2001
EEUU	c. 1381A>T (p.K461X)	Stella, et al., 2001
Itàlia	c.2269_2270insT	Caluseriu, et al., 2004
Itàlia	c.1831delAT	Thiffault, et al., 2004
Itàlia	c.545+3A>G	Thiffault, et al., 2004
Corea	c.1757_1758insC	Shin, et al., 2004
Taiwan	c. 793C>T (p.R265C)	Tang, et al., 2009
Espanya	c.306+5G>A	Borràs, et al., 2010
Espanya	c.1865T>A (p.L622H)	Borràs, et al., 2010
Itàlia	c.731G<A	Lastella, et al., 2011
Itàlia	c.1558+1G>T	Lastella, et al., 2011
Itàlia	c.131G<A	Lastella, et al., 2011
Portugal	c.1896+280_oLRRFIP2: c.1750-678del	Pinheiro, et al., 2011
EEUU, Itàlia, Alemanya	c.589-2A>G	Tomsic, et al., 2012
MSH2		
Població	**Mutació**	**Referències**
New foundland	c.942+3A>T	Froggatt, et al., 1999
Jueus Azhkenazi	c.1906G>C (p.A636P)	Foulkes, et al., 2002; Sun et al .,2005
EEUU	Deleció de exó 1-6	Wagner, et al., 2003; Clendenning et al., 2008
Xina	c.1452_1455delAATG	Chan, et al., 2004
Danesa	c.1788_1790delAAT	Ripa, et al., 2005
Espanya (Tenerife)	c.2063T>G (p.M688R)	Medina-Arana, et al., 2006
Itàlia	Deleció exó 1-exó 6	Stella, et al., 2007
Espanya	c.[2635-3T>C;2635-5C>T]	Menendez, et al., 2010
Espanya	Deleció exó 4-exó 8	Pèrez-Cabornero, et al., 2011
Espanya	Deleció exó 7	Pèrez-Cabornero, et al., 2011
Itàlia	Deleció exó 8	Borelli, et al., 2012
MSH6		
Població	**Mutació**	**Referències**
Finlàndia	c.2983G>T (p.E995X)	Vahteristo, et al., 2005
Suïssa	c.1346T>C (p.L449P)	Cederquist, et al., 2005
Suïssa	c.2931C>G (p.Y997X)	Cederquist, et al., 2005
Jueus Azhkenazi	c.3984_3987dupGTCA	Goldberg, et al., 2010; Raskin et al., 2011
Jueus Azhkenazi	c.3959_3962delCAAG	Raskin, et al., 2011
PMS2		
Població	**Mutació**	**Referències**
Anglaterra, Suïssa	c.736_741delins	Clendenning, et al., 2008
EEUU	c.137G>T (p.S46I)	Tomsic, et al., 2012
EEUU	Deleció exó 10	Tomsic, et al., 2012
EPCAM		
Població	**Mutació**	**Referències**
Països Baixos/EEUU	c.859-1462_*1999del	Ligtenberg, et al., 2009
Espanya	del exó 8-9	Mur, et al.,(article 7, annex 1, sotmès a publicació)

6. AVALUACIÓ DE LA PATOGENICITAT DE LES VARIANTS DE SIGNIFICAT DESCONEGUT

Com s'ha esmentat anteriorment, una proporció important, al voltant del 30%, de les alteracions detectades en la síndrome de Lynch són variants de significat desconegut (VSD) (Peltomaki and Vasen, 2004). Aquestes VSD són generalment canvis *missense*, canvis *silents* o canvis en les zones no codificants. És especialment difícil predir l'efecte funcional de mutacions *missense*, que podran interferir en la funcionalitat de la proteïna o en el correcte processament del mRNA (Auclair, et al., 2006; Bray, et al., 2012; Winawer, et al., 1997). També pot ser problemàtica la interpretació de canvis en zones no codificants (com poden ser canvis intrònics no localitzats a les zones consens de *splicing*, canvis en el promotor, canvis en la zona 3'UTR) i de canvis *silents* en dominis codificants, que podran afectar l'estabilitat o el processament del mRNA. Així doncs, s'han descrit alguns criteris que poden suggerir patogenicitat de les variants associades a la síndrome de Lynch. A continuació es descriuen breument (Cotton and Scriver, 1998; Hofstra, et al., 1997; Syngal, et al., 1999):

- **Cosegregació de la variant amb la malaltia.** Per avaluar-la és necessari analitzar familiars de l'individu portador de la variant, tant afectats com sans, per tal de determinar si la variant segrega amb la malaltia o s'hereta de forma independent. Aquest criteri és aparentment fàcil de quantificar tot i que la seva utilitat clínica real pot ser limitada. Sovint no es disposa d'un elevat nombre d'individus afectats vius en una mateixa família (Dunlop, et al., 1997; Quehenberger, et al., 2005) o bé la penetrància de la malaltia és incomplerta i això dificulta l'anàlisi d'aquest criteri.
- **Absència de la variant en una sèrie d'individus control.** Si bé aquest és el mètode estàndard per a l'avaluació de les variants genètiques comunes, el principal desavantatge en el context de l'avaluació de la patogenicitat és que la majoria de les VSD són poc freqüents (en general les seves freqüències són <1%). Per tant, és necessari una mida mostral elevada (~10000 individus) per tenir la potència necessària per demostrar que la variant no és present a la població. A més, les variants d'interès sovint són específiques d'una regió geogràfica o grup ètnic, el que fa encara més difícil l'obtenció de mostres. Malgrat aquestes limitacions, aquest estudi s'utilitza com un mètode ràpid per detectar possibles variants neutrals. Es considera doncs, que si una variant és genotipada en uns pocs centenars de controls i es demostra que té una freqüència de 1% o superior, és molt poc probable que sigui una variant d'alt risc (Goldgar, et al., 2008).

- **No co-ocurrència** de la variant amb mutacions patogèniques en l'altre al·lel (en *trans*), donat que mutacions bial·lèliques en els gens reparadors donen lloc a la síndrome CMMR-D (veure l'apartat 2.1), amb un fenotip molt més agressiu (Wimmer and Etzler, 2008).
- **Aparició *de novo*** de la mutació en un individu amb clínica de la síndrome de Lynch.
- **Pèrdua de l'al·lel no mutat en el tumor**. Per la hipòtesi de Knudson, s'assumeix que la pèrdua total de la funció en els gens reparadors és deguda a la inactivació somàtica dels dos al·lels. En els tumors de la síndrome de Lynch, si es detecta la pèrdua del segon al·lel, aquest fet suggereix patogenicitat de la variant detectada a nivell germinal.
- **Absència de tinció inmunohistoquímica** de la proteïna corresponent en teixit tumoral. Aquesta absència és causada per la pèrdua d'expressió dels dos al·lels del gen. S'ha de destacar que, a vegades, les mutacions o VSD poden produir un defecte en la funcionalitat de la proteïna, tot mantenint la seva expressió (Heinen and Juel Rasmussen, 2012; Hofstra, et al., 2008).
- **Presència de MSI en el teixit tumoral del pacient,** com a conseqüència de la pèrdua de la funcionalitat de les proteïnes reparadores (Aaltonen, et al., 1993)
- Per a variants *missense*, **canvi de la polaritat de l'aminoàcid**.
- Per a variants *missense***, canvi d'aminoàcid en un domini que està evolutivament conservat** entre espècies i/o compartit entre proteïnes que pertanyen a la mateixa família (Tavtigian, et al., 2008).
- **Defecte a nivell funcional de la VSD estudiada** (veure apartat següent).

6.1 Estudis funcionals de les variants de significat desconegut

Per tal d'avaluar el defecte a nivell funcional de les VSD identificades en individus susceptibles a patir la síndrome de Lynch, s'han desenvolupat múltiples assaigs a nivell *in silico* e *in vitro* i també en diferents sistemes experimentals, que avaluen diferents aspectes tant a nivell de mRNA com de proteïna.

Per la majoria d'aquests assajos, no es necessària l'obtenció de mostres addicionals del pacient o dels seus familiars, el que facilita, en part, la seva implementació. Per contra, és necessari tenir uns bons coneixements del gen/proteïna que s'analitzarà, així com disposar d'uns bons dissenys i estratègies per al seu anàlisi. Cal tenir en compte però, que a vegades és difícil traslladar els resultat obtinguts en els anàlisis funcionals a la clínica (Couch, et al., 2008).

Anàlisi in silico

Aquests anàlisis permeten predir com pot afectar una VSD a nivell de processament del mRNA i a nivell de funció de la proteïna. Els programes de predicció dels llocs de *splicing* han estat desenvolupats per investigar la possibilitat de la creació, modificació o pèrdua d'un d'aquests llocs (revisat per (Spurdle, et al., 2008)). També hi ha programes que prediuen les modificacions en els elements exònics activadors de l'*splicing* (ESE). Els programes de predicció a nivell de proteïna sovint utilitzen la conservació evolutiva com a mesura per valorar la importància del residu, així com la severitat de la substitució o les diferències estructurals entre la proteïna salvatge i la mutada (Arnold, et al., 2009; Chao, et al., 2008; Olatubosun, et al., 2012; Tavtigian, et al., 2008). A la Taula 7 es llisten els programes més utilitzats per les prediccions *in silico* a nivell de RNA i proteïna a la literatura.

Taula 7. Programes de predicció *in silico* a nivell de RNA i proteïna més comunament utilitzats.

***Anàlisi in silico* a nivell RNA**		
Programes de predicció de llocs splicing		
Programa	Pàgina Web	*Referències*
NetGene2	http://www.cbs.dtu.dk/services/NetGene2/	Hebsgaard et al., 1996 Brunak et al., 1991
Spliceport	http://spliceport.cs.umd.edu	Dogan et al., 2007
NNSplice	http://www.fruitfly.org/seq_tools/splice.html	Reese et al., 1997
SoftBerry	http://linux1.softberry.com/berry.phtml	Burset et al., 2001
MaxEntScan	http://genes.mit.edu/burgelab/maxent/ Xmaxentscan_scoreseq.html	Eng L et al., 2004
GENESCAN	http://genes.mit.edu/GENSCAN	Burge and Karlin et al., 1998
Splice Site Finder	http://violin.genet.sickkids.on.ca/ali/ splicesite¢nder.html	Senapathy et al., 1990
GeneSplicer	http://www.tigr.org/tdb/GeneSplicer/gene_spl.html	Pertea et al., 2001
Human Splice Finder	https://splice.uwo.ca	Rogan et al., 1998, Nalla i Rogan et al., 2005
Splicing Sequene Finder	www.umd.be/SSF/	Desmet et al., 2009
Programes de predicció de llocs ESE		
Programa predicció	*Pàgina web*	*Referències*
RescueESE	http://genes.mit.edu/burgelab/rescue-ese/	Fairbrother et al., 2002
ESEfinder 3.0	http://rulai.cshl.edu/cgi-bin/tools/ESE3/ esefinder.cgi?process=home	Cartegni et al., 2003 Smith et al., 2006
PESX	http://cubweb.biology.columbia.edu/pesx/	Zhang i Chasin., 2004 Zhang et al.,2005

Anàlisi *in silico* a nivell proteïna		
Programa predicció	*Pàgina web*	*Referències*
Polyphen-2	http://genetics.bwh.harvard.edu/pph2/	Adzhubei et al., 2010
A-GVGD	http://agvgd.iarc.fr/	Mathe et al., 2006 Tavtigian et al., 2005
SIFT	http://sift.jcvi.org/	Ng i Henikoff et al., 2002
MAPP-MMR	http://mappmmr.blueankh.com/	Chao et al., 2008
CONDEL	http://bg.upf.edu/condel/home	Gonzalez-Perez et al., 2011
Mutation Assessor	http://mutationassessor.org/v1/	Reva et al., 2011

Assaigs a nivell de mRNA

L'estudi de l'efecte en el mRNA, es pot realitzar a diferents nivells:

- Estudi experimental sobre l'efecte en el processament: Aquests estudis es realitzen a partir d'un disseny específic per avaluar l'efecte de cada VSD en l'*splicing* del mRNA mitjançant RT-PCR (Arnold, et al., 2009; Sharp, et al., 2004; Tournier, et al., 2008).
- Estudi de l'expressió al·lèlica diferencial (ASE): Aquests estudis analitzen la quantificació relativa dels dos al·lels del mRNA (Castellsague, et al., 2010; Perera, et al., 2010; Santibanez Koref, et al., 2010).

Sempre que sigui possible, aquests assaigs es portaran a terme a partir del RNA dels limfòcits dels pacients. Tanmateix, molts laboratoris no disposen d'aquest tipus de mostra. Per aquest motiu s'han desenvolupat els assaigs de mRNA *in vitro,* que presenten una bona correlació amb les assaigs amb limfòcits (Betz, et al., 2010; Lastella, et al., 2006; Naruse, et al., 2009; Tournier, et al., 2008). Alguns d'aquests assaigs utilitzen plàsmids d'expressió on s'incorpora la VSD a estudiar, i posteriorment es transfecta aquest plàsmid en una línia cel·lular estable. El RNA extret de les cèl·lules s'utilitzarà per fer l'anàlisi a partir de RT-PCR (Betz, et al., 2010; Lastella, et al., 2006; Naruse, et al., 2009; Tournier, et al., 2008). Per altra banda, també s'han descrit els assaig de conversió. Aquests es basen en convertir una mostra diploide en una haploide, mitjançant l'aïllament dels dos al·lels en híbrids de cèl·lules somàtiques i posteriorment es realitza l'estudi del RNA de cada cèl·lula (Casey, et al., 2005; Nakagawa, et al., 2004; Nakagawa, et al., 2002; Yan, et al., 2000).

Assaig in vitro

Assaigs funcionals en llevats

Els assaigs funcionals en llevats s'han desenvolupat basant-se en el fet que el sistema MMR està altament conservat evolutivament (Ou, et al., 2007). A partir d'aquest concepte, s'analitza l'efecte de la VSD en la seva proteïna homòloga de llevat. S'han desenvolupat tres tipus d'estratègies diferents per aquests assaigs: (I) expressió del gen mutat de llevat en una soca de llevat haploide deficient en aquest gen (Shcherbakova and Kunkel, 1999; Wanat, et al., 2007), (II) expressió del gen mutat d'humà en una soca haploide salvatge de llevat (Clark, et al., 1999; Shimodaira, et al., 1998; Takahashi, et al., 2007), (III) expressió del gen mutat de llevat en una soca diploide de llevat que conté

només una copia funcional del gen a analitzar (Drotschmann, et al., 1999). A més, diversos mètodes han estat aplicats per avaluar les taxes de mutació espontània en llevats de les diverses soques que expressen mutants de les proteïnes MMR (Polaczek, et al., 1998; Shimodaira, et al., 1998; Takahashi, et al., 2007; Tran, et al., 1997; Wanat, et al., 2007).

Assaig de reparació in vitro que utilitzen línies cel·lular humanes

Alguns estudis analitzen la capacitat de les variants introduïdes en el cDNA dels gens reparadors per a complementar fenotips defectius d'aquests gens en línies cel·lulars (Blasi, et al., 2006; Mastrocola and Heinen, 2010a). A les cèl·lules complementades, l'estat de reparació és investigat mesurant la resposta cel·lular a agents metilants (Li, 2008), la taxa de mutació espontània, la capacitat de reparació, o la mesura de la inestabilitat de microsatèl·lits (revisat a Rasmussen, et al., 2012).

Assaig d'activitat reparadora in vitro

La reparació dels desaparellaments es reconstitueix *in vitro*, proporcionant una informació valuosa sobre les proteïnes necessàries i els seu paper durant el procés de reparació (Dzantiev, et al., 2004; Holmes, et al., 1990; Zhang, et al., 2005). Els extractes cel·lulars o les proteïnes recombinats són testades per observar la seva habilitat de reparació en d'un plàsmid que conté un desaparellament en un lloc de restricció. Aquest assaig de reparació ha estat utilitzat per estudiar la capacitat de reparació de moltes VSD tant en el gen *MLH1, MSH2* com *MSH6* (revisat a Heinen and Juel Rasmussen; 2012; Ou, et al., 2007; Rasmussen, et al., 2012). L'avaluació de la reparació dels desaparellaments és probablement l'assaig biològicament més rellevant per avaluar la funció d'una variant en els gens reparadors.

Estudi d'altres propietats de les proteïnes reparadores

Aquests assaigs ofereixen analitzar aspectes més específics de les propietats de la proteïna, que són de vital rellevància per la seva funció (revisat a Heinen and Juel Rasmussen, 2012; Rasmussen, et al., 2012). Aquest assaigs es poden dividir en:

- *Assaigs d'interacció de les proteïnes reparadores***:** Aquests assaigs avaluen la formació dels heterodímers entre les diferents proteïnes reparadores. Per estudiar aquestes interaccions disposem de varis assaigs: (I) L'assaig *pull-down* amb Glutatió S-transferasa (GST) (Belvederesi, et al., 2006; Guerrette, et al., 1999; Guerrette, et al.,

1998; Yuan, et al., 2002), (II) la Inmunoprecipitació (Kosinski, et al., 2010; Nystrom-Lahti, et al., 2002; Ollila, et al., 2006), (III) l'assaig de doble híbrid en llevats (Jager, et al., 2001; Kondo, et al., 2001; Kondo, et al., 2003; Pang, et al., 1997) i (IV) l'expressió transitòria dels gens reparadors en un línia MMR deficient. En aquest cas la transfecció de la proteïna mutada dins d'una línia cel·lular deficient en aquesta proteïna i en el la proteïna que forma l'heterodímer, ens permetrà valorar defectes d'interacció amb la proteïna que heterodimeritza (Belvederesi, et al., 2006; Brieger, et al., 2002; Kosinski, et al., 2010; Ollila, et al., 2006).

- *Assaigs d'expressió i estabilitat de les proteïnes reparadores:* La transfecció transitòria d'un plàsmid amb una VSD, l'estudiarem en una línia cel·lular deficient en el gen en qüestió. Aquest estudi ens permetrà avaluar l'expressió i l'estabilitat d'aquesta proteïna, mitjançant la seva extracció i posterior anàlisi per western-blot (Kosinski, et al., 2010; Lutzen, et al., 2008; Perera and Bapat, 2008; Takahashi, et al., 2007).
- *Assaigs d'unió al DNA:* Per avaluar la capacitat d'un heterodímer a unir-se al DNA, es poden utilitzar dos tipus d'assaigs: (I) Els assaigs de canvi de mobilitat del DNA (Cyr, et al., 2012; Drotschmann, et al., 1999; Heinen, et al., 2002b; Lutzen, et al., 2008; Ollila, et al., 2008), i (II) els assaigs d'unió al DNA immobilitzat (Heinen, et al., 2002b).
- *Assaigs basats en el cicle ATP-ADP:* Els quals valoren diferents aspectes relacionats amb els cicle de l'ATP-ADP, com la unió de les proteïnes a l'ATP, la hidròlisi d'ATP i l'eficiència catalítica de l'ATP, els canvis conformacionals en presència d'ATP, entre d'altres (Cyr, et al., 2012; Heinen, et al., 2002b; Lutzen, et al., 2008; Ollila, et al., 2008; Raevaara, et al., 2005).
- *Assaigs de localització cel·lular:* La localització subcel·lular de les proteïnes reparadores amb la variant a analitzar poden ser estudiades mitjançant inmunotinció o bé, per l'estudi de la proteïna fusionada amb un fluorocrom que serà expressada en cèl·lules humanes deficients en aquesta proteïna (Brieger, et al., 2005; Knudsen, et al., 2007; Lutzen, et al., 2008; Raevaara, et al., 2005).

Anàlisi d'altres funcions de les proteïnes reparadores

Les proteïnes reparadores tenen altres funcions a part de reparar els possibles desaparellaments erronis que es donen durant la replicació del DNA. Algunes d'aquestes funcions estan involucrades en l'activació dels punt de control durant el cicle cel·lular, en l'apoptosi en resposta a danys causat en el DNA per agents tòxics, entre d'altres. Hi ha varis estudis en els que s'estudia l'efecte de determinades VSD en relació amb aquestes funcions (Marinovic-Terzic, et al., 2008; Mastrocola and Heinen, 2010b).

6.2 Algoritmes de classificació proposats per l'avaluació de la patogenicitat de les variants de significat desconegut

Encara que un dels aspectes més rellevants per a l'avaluació de la patogenicitat d'una VSD són les característiques clinico-patològiques, sovint, resulten insuficients per establir el grau de rellevància que aquestes tenen en la síndrome de Lynch (Couch, et al., 2008). Per aquest motiu, la combinació d'aquests aspectes junt amb els diferents estudis funcionals s'està consolidant com a eina per classificar les VSD detectades en els gens reparadors.

Per tal avaluar de la patogenicitat de les VSD en els gens MMR, es va proposar un algoritme de classificació diagnòstic jeràrquic en tres etapes (Figura 19)(Couch, et al., 2008). El primer pas inclou el procediment diagnòstic dels pacients amb sospita de la síndrome de Lynch, on s'inclou l'anàlisi de IHQ i MSI en el teixit tumoral, seguit del test genètic dels gens reparadors. En cas que es detecti una VSD, el procediment de diagnòstic continua amb una sèrie d'assajos que en gran mesura es basen en el coneixement de la biologia del sistema MMR. En l'etapa 2 es proposa l'anàlisi in silico anivell de mRNA i de proteïna (Spurdle, et al., 2008; Tavtigian, et al., 2008), seguit de l'anàlisi in vitro de l'activitat reparadora. Si els resultats indiquen deficiència en l'activitat reparadora, es pot considerar com una forta evidència qualitativa de que aquesta variant és patogènica. Si la variant mostra activitat reparadora, s'analitzaran diferents propietats de la proteïna. La variant podrà ser considerada com a patogènica, si mostra defecte en alguna d'aquestes propietats.

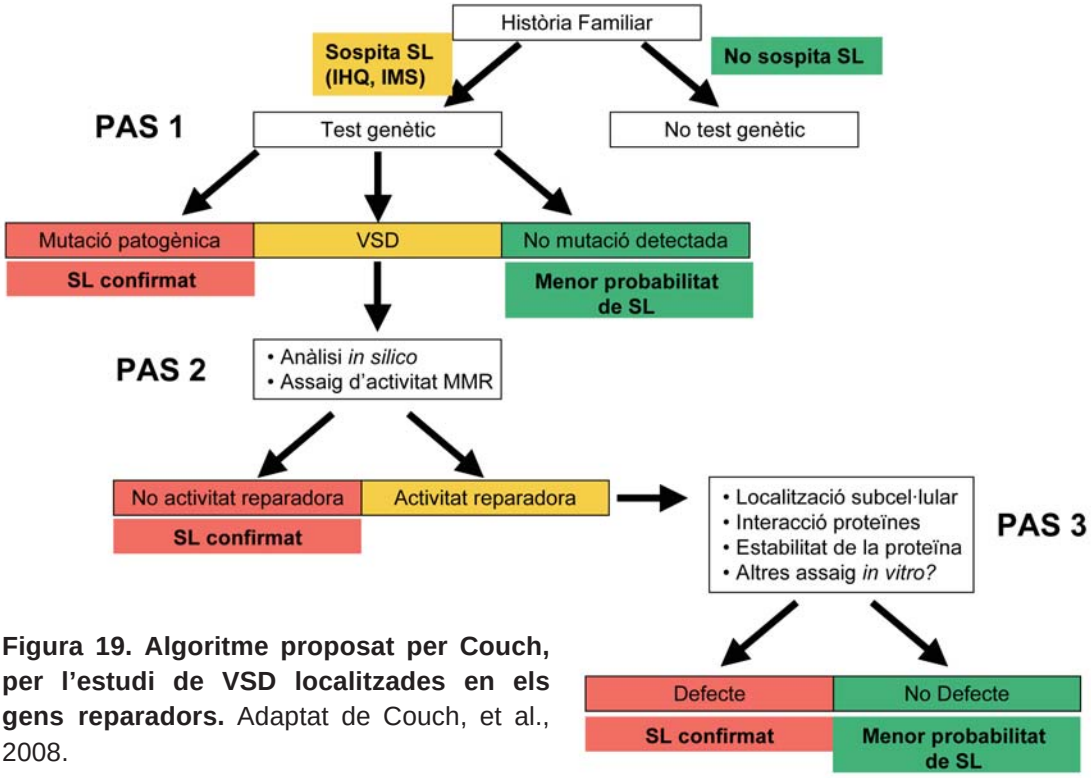


Figura 19. Algoritme proposat per Couch, per l'estudi de VSD localitzades en els gens reparadors. Adaptat de Couch, et al., 2008.

6.3 Sistemes de classificació de VSD

L'any 2007, es van proposar unes guies per a la classificació de variants, tal com es presenta a la pàgina web:
http://cmgsweb.shared.hosting.zen.co.uk/BPGs/Best_Practice_Guidelines.htm.
Aquestes directrius establien vàries normes de qualitat i línies d'evidència des d'un punt de vista molecular, per tal d'ajudar a determinar la possible patogenicitat de les VSD. Aquesta guia conclou que és essencial informar de totes les variants detectades en l'estudi de seqüenciació, i proposa informar de les VSD classificant-les en tres classes: (I) certament no patogènica, (II) poc probablement patogènica, i (III) probablement patogènica. No obstant això, els nivells de patogenicitat per classificar una VSD dins una categoria determinada, i les recomanacions clíniques associades amb cada categoria, no estaven incloses. Així mateix, la *American College of Medical Genetics (ACMG)* va proporcionar sis categories estàndard, per interpretar i reportar les diferents VSD sense especificar quin era la informació quantitativa necessària per categoritzar les VSD (Richards, et al., 2008).

Posteriorment, es va proposar la utilització de models Bayesians per a la integració d'evidències directe e indirectes de la patogenicitat d'una variant, en un model únic per obtenir un valor de probabilitat de patogenicitat. Aquests models van ser establerts per primer cop per trobar solució a la classificació de VSD detectades en els gens *BRCA1* i *BRCA2* (Goldgar, et al., 2004). Més recentment, també s'han proposat per avaluar el risc de càncer associat a VSD en els gens reparadors (Goldgar, et al., 2008; Spurdle, 2010).

Idealment, en aquest model integral, tant les dades quantitatives com les qualitatives serien adequadament ponderades per arribar a una classificació final. El resultat final seria la probabilitat general de patogenicitat, és a dir, la probabilitat de les dades observades sota la hipòtesi que la variant és patogènica en comparació amb la corresponent probabilitat que la variant sigui neutral. Cada tipus de prova depèn de diferents models i supòsits, i alguns són més adequats per a la quantificació i la formulació com un quocient de probabilitat que altres. Els primers estudis realitzats amb models Bayesians es centren principalment en les dades que poden ser avaluades directament sobre una base genètica/epidemiològica, ja que aquestes dades són fàcilment quantificables en termes de quocients de probabilitat i, a més, són les més directament relacionades amb els resultats clínics d'interès.

Plon, et al., 2008, va proposar un sistema de classificació en cinc categories en relació a la seva probabilitat de patogenicitat (Taula 8), que actualment, és el més comunament utilitzat. El major avantatge d'aquest sistema de classificació és la creació de la classe 2 (probablement no patogènica/poca significació clínica) i la classe 4 (probablement patogènica), amb una probabilitat de patogenicitat del 0.1-4.9% i del 95-99%, respectivament. Es considera important diferenciar aquests dos grups del grup de classe 3 (variant de significat desconegut), per diferenciar entre les variants en les que hi ha realment poca informació per donar alguna recomanació, i les variants en que hi ha certa evidència en contra (classe 2) o certa evidència a favor (classe 4) de la seva patogenicitat. Tot i així, aquest sistema no és ideal, ja que no ens permet classificar de forma qualitativa, i a més no es disposa d'un valor assignat per cada evidència.

Taula 8. Sistema de classificació proposat para les VSD identificades en els gens reparadors. Extret de Plon, et al., 2008.

Classe	Descripció	Probabilitat de ser patogènica
5	Definitivament patogènica	>0.99
4	Probablement patogènica	0.95-0.99
3	Variant de significat desconegut	0.05-0.949
2	Probablement no patogènica o amb poc significat clínic	0.001-0.049
1	No patogènica o sense siginificat clínic	<0.001

HIPÒTESI

L'anàlisi mutacional de gens reparadors en famílies amb la síndrome de Lynch és molts cops no informatiu. La detecció de Variants de Significat Desconegut (VSD), que representen fins un 30% de les alteracions detectades, ha merescut cada cop major atenció tant per l'increment de la freqüència en la seva detecció, com pel repte que representa obtenir una informació funcional rellevant en una situació ambigua.

L'estudi d'aquestes variants mitjançant l'anàlisi de cosegregació, freqüència en controls en població general, la naturalesa molecular del canvi i la conservació evolutiva del residu moltes vegades és no concloent per determinar la seva patogenicitat. Els estudis funcionals de variants són més informatius, encara que rarament són exhaustius. Aquests poden incloure estudis a nivell de mRNA i estudis funcionals a nivell de proteïna, que en moltes ocasions representen un gran repte en els laboratoris orientats a l'anàlisi mutacional tant per les limitacions dels models *in vitro* utilitzats com per la manca d'estudis que mostrin la seva reproduibilitat.

Hipotetitzem que l'estudi exhaustiu de VSD en els gens *MLH1* i *PMS2*, ens permetrà classificar la seva patogenicitat en una gran majoria dels casos. En conseqüència, aquests estudis permetran augmentar el percentatge d'individus amb sospita de síndrome de Lynch i mutació patogènica detectada. Aquest fet ens permetrà millorar el consell genètic d'aquestes famílies.

Per altra banda, les mutacions detectades amb elevada freqüència poden ser mutacions recurrents, és a dir, que ocorren repetidament provinents d'esdeveniments mutacionals independents (de novo), o bé fundadores, mutacions que s'han donat en un únic individu ancestre o bé han estat introduïdes en una població determinada per aquest únic individu. L'estudi d'haplotips en diferents famílies portadores d'una mateixa mutació s'utilitza per identificar el seu possible efecte fundador. En la nostra sèrie detecten dues variants amb una elevada freqüència en el gen *MLH1*.

Hipotetitzem que pot tractar-se de dues mutacions fundadores a la població espanyola, la seva identificació ens ajudarà a millorar el diagnòstic molecular de la síndrome de Lynch en la nostra població.

OBJECTIUS

Objectiu general

Aquesta tesi té per objectiu avaluar l'impacte funcional de les variants de significat desconegut identificades en el gens *MLH1* i *PMS2* en pacients amb sospita de síndrome de Lynch. La determinació de la patogenicitat d'aquestes variants serà clau en l'assessorament genètic d'aquestes famílies, i permetrà l'estudi predictiu de familiars de risc així com la personalització de les recomanacions de mesures de prevenció i seguiment adequades.

Objectius específics

1. Estudiar el possible efecte fundador de dues variants de significat desconegut en el gen *MLH1*, la c.306+5G>A i la 1865T>A, que hem identificat com a recurrents en la nostra població.

2. Estudiar funcionalment les variants de significat desconegut identificades en els gens *MLH1* i *PMS2* mitjançant una estratègia raonada per a l'estudi funcional, integrant estudis a nivell de RNA i proteïna, *in silico* i *in vitro.*

RESULTATS

La secció de Resultats d'aquesta tesi consta de 2 articles publicats i un sotmès a publicació, que es descriuen breument a continuació. Per altra banda, s'adjunten a l'Annex 1 altres articles en els quals la doctoranda ha col·laborat i n'és coautora.

Article 1

"Mutacions fundadores en el gen *MLH1* amb una moderada penetrància en famílies espanyoles amb síndrome de Lynch"
Cancer Research, 2010

Article 2

"Anàlisi funcional exhaustiu de variants de significat desconegut detectades en el gen *MLH1*"
Human Mutation, 2012

Article 3

"Refinament del rol de *PMS2* en la síndrome de Lynch: Anàlisi mutacional en línia germinal millorada per l'avaluació exhaustiva de les variants de significat biològicament desconegut"
Sotmès a publicació.

ARTICLE 1

MLH1 FOUNDER MUTATIONS WITH MODERATE PENETRANCE IN SPANISH LYNCH SYNDROME FAMILIES

Ester Borràs*, Marta Pineda*, Ignacio Blanco, Ethan M. Jewett, Fei Wang, Àlex Teulé, Trinidad Caldés, Miguel Urioste, Cristina Martínez-Bouzas, Joan Brunet, Judith Balmaña, Asunción Torres, Teresa Ramón y Cajal, Judit Sanz, Lucía Pérez-Cabornero, Sergi Castellví-Bel, Ángel Alonso, Ángel Lanas, Sara González, Víctor Moreno, Stephen B. Gruber, Noah A. Rosenberg, Bhramar Mukherjee, Conxi Lázaro, i Gabriel Capellá. *(*) Aquests autors han contribuït equitativament en aquest treball.*

Cancer Research; 70(19); 7379–91 (2010)

Resum: Les variants c.306+5G>A i c.1865T>A del gen *MLH1* havien estat freqüentment detectades en famílies espanyoles amb síndrome de Lynch. Per entendre la seva història ancestral i el seu efecte clínic, es van dur a terme anàlisis funcionals, anàlisis de penetrància i es van estudiar els seus orígens genètics i geogràfics. Es van estudiar detalladament la història familiar de 29 individus portadors. L'anàlisi funcional va incloure estudis *in silico* i *in vitro* tant a nivell de RNA com de proteïna. La penetrància de les mutacions es va calcular utilitzant una modificació de l'anàlisi de segregació que ajusta el possible biaix associat a la identificació del casos. L'efecte fundador es va avaluar a partir de l'anàlisi d'haplotips. Aquestes variants segregaven amb la malaltia i no van ser identificades en una sèrie control. Els tumors dels individus portadors mostraven inestabilitat de microsatèl·lits i pèrdua d'expressió de la proteïna MLH1. La variant c.306+5G>A afectava el correcte processament del mRNA i per això es va classificar com a patogènica. En canvi, la variant c.1865T>A (p.Leu622His) és patogènica degut al seu defecte en l'expressió i l'estabilitat de la proteïna MLH1. Per ambdues variants, la penetrància estimada és moderada i es troba dins del rang de la variabilitat estimada per les mutacions patogèniques del gen *MLH1* en la població espanyola. L'haplotip comú associat a cada una de les mutacions identificades confirmava l'origen fundador. L'edat de les mutacions estimada va ser de 53 a 122 per la mutació c306+5G>A i de 12 a 22 generacions per la mutació c.1865T>A. Els nostres resultats confirmen la patogenicitat, la moderada penetrància i l'origen fundador de les mutacions c.306+5G>A i c.1865T>A en el gen *MLH1*. Els resultats obtinguts tenen important implicacions tant en el consell genètic com en el diagnòstic molecular de la síndrome de Lynch.

Published OnlineFirst on September 21, 2010 as 10.1158/0008-5472.CAN-10-0570

Clinical Studies

Cancer Research

MLH1 Founder Mutations with Moderate Penetrance in Spanish Lynch Syndrome Families

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

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: [1]Laboratori de Recerca Translacional, Institut Català d'Oncologia, IDIBELL, Hospitalet de Llobregat; [2]Programa de Consell Genètic en Càncer, Institut Català d'Oncologia, Hospital Duran i Reynals, Hospitalet de Llobregat; [3]Servei d'Oncologia Mèdica, Hospital Vall d'Hebron; [4]Servei d'Oncologia Mèdica, Hospital de la Santa Creu i Sant Pau; [5]Institut d'Oncologia Corachan-Centre Mèdic; [6]Servei de Gastroenterologia, Institut de Malalties Digestives i Metabòliques, Hospital Clínic; [7]Unitat de Bioinformàtica i Bioestadística, Institut Català d'Oncologia i Universitat de Barcelona, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain; [8]Programa de Consell Genètic en Càncer, Institut Català d'Oncologia, Hospital Germans Trias i Pujol, Badalona, Spain; [9]Programa de Consell Genètic en Càncer, Institut Català d'Oncologia, Hospital Josep Trueta, Girona, Spain; [10]Center for Computational Medicine and Bioinformatics and [11]Department of Human Genetics, and the Life Sciences Institute, University of Michigan; [12]Department of Internal Medicine, University of Michigan Medical School and School of Public Health; [13]Departments of Internal Medicine and Human Genetics, University of Michigan Medical School; [14]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan; [15]Laboratorio de Oncología Molecular, Hospital Clínico San Carlos; [16]Departamento de Genética Humana, Centro Nacional de Investigaciones Oncológicas; [17]Centro de Investigación en Red de Enfermedades Raras, Madrid, Spain; [18]Laboratorio de Genética Molecular, Hospital de Cruces, Bizkaia, Spain; [19]Unitat de Consell Genètic, Hospital Universitari Sant Joan, Reus, Spain; [20]Instituto de Biología y Genética Molecular, Valladolid, Spain; [21]Servicio de Genética, Hospital Virgen del Camino, Pamplona, Spain; and [22]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

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 descibed 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)TGCAGAC-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). β-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 β-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 ≥50, assuming that $g(t) = \sum_{k=1}^{n} \exp[\beta_k]$, a piecewise constant HR in the k^{th} age band $k = 1, 2$. 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^{th} age interval, and β_k is the log(RR) in the k^{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 $1 - \exp[-\Lambda(t) \pm 1.96\sqrt{Var(\Lambda(t))}]$, where

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

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					L	
					CRC	49						
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	+	–	+	+		
					CRC	65					L	B
			B6.3	F	EC	55	+	–	+	+		
			B6.4	M	CRC	32					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 Universatario 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.

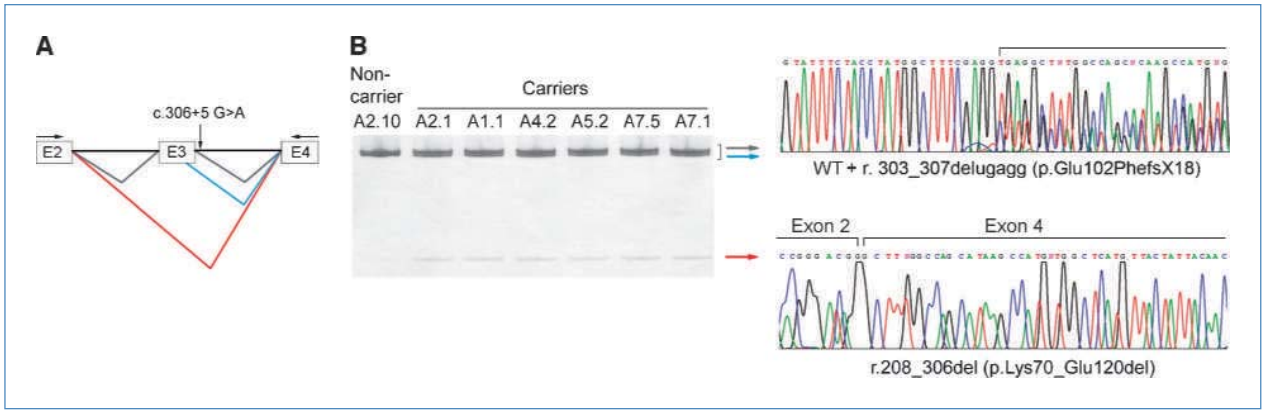


Figure 1. Characterization of the c.306+5G>A *MLH1* variant. A, schematic overview of *MLH1* exons 2 to 4 with a representation of the normal and aberrant transcripts caused by the c.306+5G>A mutation. Black arrows represent primers used for RT-PCR amplification. B, acrylamide gel showing RT-PCR products 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_307delugagg, 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.

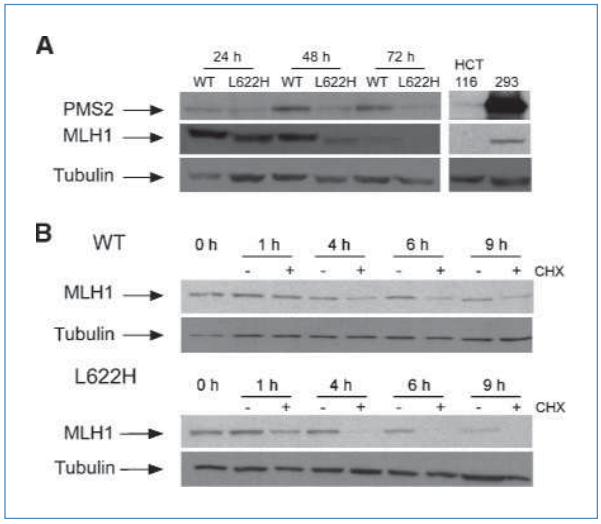


Figure 2. Expression and stability of the p.Leu622His variant of MLH1 in HCT116 cells. A, time course of expression of WT MLH1 and p.Leu622His (L622H) variants. Top and middle, cell lysates probed with anti-PMS2 and anti-MLH1 antibodies; bottom, to verify equal protein loading, cell lysates were probed 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 *MLH1* 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, 06–10.2) 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

A

Age (y)	CRC (males)				CRC (females)			
	% Cumulative risk population	% Cumulative risk *MLH1* carriers (95% CI)			% Cumulative risk population	% Cumulative risk *MLH1* carriers (95% CI)		
		c.306+5G>A	c.1865T>A	Other *MLH1* mutations		c.306+5G>A	c.1865T>A	Other *MLH1* 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)
		EC (females)				CRC and EC (females)		
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

Age (y)	HR (95% CI)					
	CRC (males)			CRC (females)		
	c.306+5G>A	c.1865T>A	Other *MLH1* mutations	c.306+5G>A	c.1865T>A	Other *MLH1* 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)
	EC (females)			CRC and EC (females)		
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.

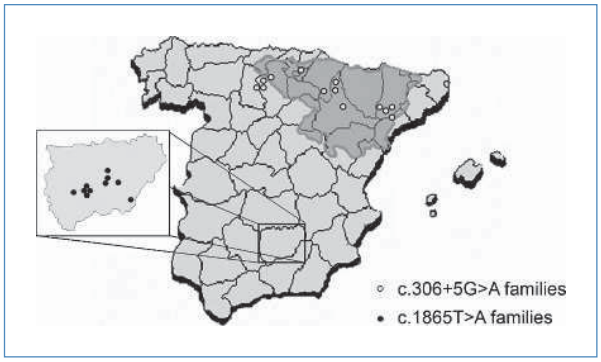


Figure 3. Map of Spain showing locations where Lynch syndrome families with the c. 306+5G>A mutation (white circles) and the c.1865T>A mutation (black circles) originated. For six families (four c.306+5G>A and two c.1865T>A), the original location is not known. The Ebro valley is highlighted in dark 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 5×10^7 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	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**199**	209
A2	250//254	106	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**199**	213
A3	250	96	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**205**	213
A4	254	92	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**205**	201/211
A5	254/256	92	**103**	**Yes**	A/**G**	**A**	250/**256**	A/**G**	**224**	**205**	201/213
A6	250	104	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**205**	201/213
A7	250	96/104	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**201**	211
A8	250/254	96/106	**103**	**Yes**	**G**	**A**	**256**	**G**	218/**224**	**199**/201	203/209
A9	254	96	**103**	**Yes**	A/**G**	**A**	**256**/260	A/**G**	216/**224**	197/**199**	213/215
A10	250/256	96/104	**101**	**Yes**	**G**	**A**/G	**256**	**G**	**224**	**199**	203/215
A11	250	96	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**199**	209/213
A12	250	96/104	**103**	**Yes**	**G**	**A**/G	**256**	**G**	216/**224**	195/**199**	201/215
A13	250	96	**103**	**Yes**	**G**	**A**/G	**256**	**G**	220/**224**	**199**/209	209/213
A14	252	96	101/**103**	**Yes**	**G**	**A**/G	**256**	**G**	**224**	**205**	211
A15	250//254	96	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**199**	213
A16	250//254	96	**103**	**Yes**	**G**	**A**	**256**	**G**	**224**	**199**	213
A17	254	96	**103**	**Yes**	A/**G**	**A**	**256**	**G**	**224**	**199**	209//213
			Max 1.55 Mb								
				Min 0.40 Mb							
Shared alleles			103	Yes	G	A	256	G	224	199	
Control frequencies (%)			**67**	**0**	**44**	**66**	**44**	**44**	**12**	**30**	

(*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	**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**205**	207
B2	256	**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**205**	207
B3	252//256	92/**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**205**	201//207
B4	256	**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**205**	201
B5	250	**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**197**	203
B6	250	**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**205**	207
B7	250/256	**104**	**101**/103	A/**G**	A/**G**	**256**	**G**	**Yes**	**222**	**205**	211
B8	250/256	**104**	**101**/103	**G**	A/**G**	**256**	**G**	**Yes**	**222**	**205**	201/207
B9	256	**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**197**	203
B10	250/254	92/**104**	**101**	A/**G**	A/**G**	**256**/260	A/**G**	**Yes**	214/**222**	**197**/199	201/203
B11	250/256	96/**104**	**101**	**G**	**G**	**256**	**G**	**Yes**	**222**	**197**/199	203/213
B12	250//256	**104**	**101**	**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 (%)		**32**	**31**	**44**	**34**	**44**	**44**	**0**	**22**	**10**	

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.

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.

References

1. 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.
2. 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.
3. 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.
4. 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.
5. Quehenberger F, Vasen HF, van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. J Med Genet 2005;42:491–6.
6. Pineda M, Gonzalez S, Lazaro C, Blanco I, Capella G. Detection of genetic alterations in hereditary colorectal cancer screening. Mutat Res 2009.
7. 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.
8. 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.
9. Nystrom-Lahti M, Kristo P, Nicolaides NC, et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat Med 1995;1:1203–6.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. Moreno V, Gemignani F, Landi S, et al. Polymorphisms in genes of

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-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.
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. Tuupanen S, Karhu A, Jarvinen H, Mecklin JP, Launonen V, Aaltonen LA. No evidence for dual role of loss of heterozygosity in hereditary nonpolyposis 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.

Supplementary Methods

Allele-specific expression (ASE) and loss of heterozygosity (LOH) analysis

PCR fragments encompassing the c.1865T>A change were generated from genomic DNA (gDNA) and cDNA obtained from peripheral blood lymphocytes, or tumor DNA (tDNA) from mutation carriers (Table S1). PCR products were purified using GFX^{TM} PCR DNA and Gel Band Purification kit (GE Healthcare). Allele-specific expression (ASE) of the variant c.1865T>A of *MLH1* at the cDNA level was analyzed by single-nucleotide primer extension (SNuPE) using the SNaPshot kit (Applied Biosystems) with primer 5'-TCTGAAGAAGAAGGCTGAGATGC-3', according to the manufacturer's instructions. Briefly, reactions were performed in a total volume of 10 µL containing 1.5 µL purified PCR product, 5 µL SNaPshot Ready Reaction Mix, and 0.2 µmol/L extension primer. Primer extension thermocycling conditions consisted of 25 cycles of 96ºC for 10 s, 50ºC for 5 s, and 60ºC for 30 s. SNaPshot reaction products were treated with 1 U shrimp alkaline phosphatase (usb) for 60 min at 37ºC and then 15 min at 75ºC. Products were run in an ABI Prism 3130 DNA sequencer and were analyzed by GeneMapper v4.0 (Applied Biosystems). Samples from c.1865T>A carriers showed a profile with two peaks (green and red, representing A and T alleles, respectively). ASE at cDNA level was calculated as the proportion of T allele between cDNA and gDNA (ASE = $f(T)_{cDNA}/ f(T)_{gDNA}$), were $f(T)$ was obtained from peak intensities using the formula $f(T)$ = [T allele/(T+A alleles)]. Loss of heterozygosity (LOH) at tumor DNA (tDNA) level was measured as the proportion of T allele between tDNA and gDNA (LOH = $f(T)_{tDNA}/ f(T)_{gDNA}$). We used gDNA samples from 8 carriers to establish a range for normal ASE and LOH between 0.89 and 1.11 (mean values ± 3·SD). Experiments were performed in duplicate.

Details on estimation of mutation age

Counting recombination events

For the c.306+5G>A mutation we counted recombination events in the region spanning microsatellites D3S2369 and D3S1298; for c.1865T>A we chose the region spanning D3S1612 and D3S1298. We only counted recombinations on a subset of disease haplotypes corresponding to those in bold in Table 3 (for further discussion see the subsection below on estimating the genealogy height). For each mutation, all diseased individuals had at least one allele in common in the immediate neighborhood of the disease allele, as would be expected if the mutation occurred on a single shared ancestral haplotype. We chose the most frequently occurring of the distinct disease haplotypes to be the ancestral one and counted the minimum number of recombination events necessary to obtain all of the other haplotypes. The choice of the ancestral disease haplotype does not affect the count of recombination events. When a haplotype was distinct from the ancestral disease haplotype we assumed that it resulted from a recombination, and we assumed that any two identical non-ancestral haplotypes arose from the same recombination event.

Estimating L from the count of recombination events

The length of the genealogy of the sampled copies of a mutation (L) was obtained by dividing the number of recombination events by the per-generation probability of a recombination on the

ancestral haplotype. The recombination map length of the c.306+5G>A haplotype was obtained using the Rutgers Map Interpolator (1). We supplied the interpolator with the position of D3S1298, which was included in the Rutgers smoothed map position file, and we interpolated the position of D3S2369, which was not included. For the interpolation we used the physical position of 36,472,209 bp from release 50 of the Ensembl database (2). We used the difference between the sex-averaged map position of D3S2369 and that of D3S1298 to obtain a map length of r=0.9796 cM for the haplotype. The procedure for the c.1865T>A mutation was similar except that both endpoints of the haplotype appeared in the smoothed map position file, and the haplotype length was r=2.2578 cM.

Estimating the height of the genealogy of the sampled copies of each mutant allele

To estimate TMRCA given the estimated genealogy length $\hat{L}$, we simulated the joint probability density of genealogy heights and lengths, $f(H,L;N,n)$. For a constant-sized population under the coalescent, this joint density is determined by the effective population size N and the number n of sampled copies of the mutation (3). Although N is unknown, the ratio of the most likely tree length L_m and the most likely height H_m given this length depends only on n and can be expressed as $L_m/H_m=1/c(n)$ for some function c. Assuming that our genealogy length estimate $\hat{L}$ is the most likely length under the true demographic history, we estimate TMRCA as $\hat{H}=c(n)\hat{L}$, where $c(n)$ was determined from coalescent simulations.

Our estimation procedure for $c(n)$ used 5×10^7 genealogies simulated under a constant population model with n=17 sampled lineages for the c.306+5G>A mutation and n=12 lineages for c.1865T>A. Simulations were carried out using the program ms (4). We estimated L_m from a histogram of the simulated tree lengths with 1,000 bins. Using a moving window of width 5 bins and step-size 1, we smoothed the histogram by averaging the values within each window, producing a vector of length 996. We then found the maximal element in this vector and averaged the centers of the bins corresponding to that element in order to estimate the location of the maximum of the distribution. Using this method, the estimated most likely gene tree length was $L_{m(\text{Ebro})}$ = 2.8031 in units of $4N$ generations ($L_{m(\text{Jaén})}$=2.3974).

To find H_m, we selected all simulated trees whose lengths fell within a window around L_m, [2.8005, 2.8057] for n=17 and [2.3947, 2.4001] for n=12. Each window was obtained by incrementally increasing the size of a symmetric window around L_m until the number of simulated genealogies with lengths in the window reached or exceeded 1,000. For each mutation the size of the increment was 2×10^{-4} units of $4N$ generations (10^{-4} units of $4N$ generations in each direction). To estimate H_m from these trees, we employed the same procedure used for estimating L_m, except that because of the smaller number of trees, the histogram for H used 100 bins and the moving window had width 3 and step-size 1. The most likely gene tree height for the given gene tree length was estimated to be $H_{m(\text{Ebro})}$=0.6514 in units of $4N$ generations ($H_{m(\text{Jaén})}$ = 0.6159). Dividing H_m by L_m yielded $\hat{c}_{\text{Ebro}}$=0.2324 and $\hat{c}_{\text{Jaén}}$=0.2569.

Note that although the number of sampled copies of the c.306+5G>A mutation is 42 (40 for the

c.1865T>A mutation), because no two of the 17 families (12 families for c.1865T>A) are believed to share a common individual within the last five generations, in the simulations we used n=17 (n=12 for c.1865T>A). Thus, to estimate the TMRCA of the sampled copies of a mutation we chose one haplotype per family to represent the haplotype in the fifth generation from that family, estimated the height of the genealogy that relates these disease haplotypes, and added four generations to the estimated height. The haplotype chosen was the one common to all family members, except in the single case of an intrafamilial recombination in the region considered. In this case, all individuals had a common ancestor in the fourth generation and the disease haplotype of this individual, assumed to be that of her two children, was chosen.

Confidence intervals for TMRCA

To obtain a 95% confidence interval for TMRCA for the c.306+5G>A mutation we simulated 100,000 trees under a coalescent of constant size with n=17 lineages (n=12 for c.1865T>A) and recorded the height H and length L of each tree. The height of each tree was then estimated from its length by H_{est}= cL and the ratio h=H/H_{est} was computed for each tree. We found the 2.5 and 97.5 percentiles of these 100,000 ratios by ordering them from smallest to largest $h_{(i)}$ (i =1…100,000) and taking $h_{(2,501)}$ and $h_{(97,500)}$. The lower bound of the 95% confidence interval for the estimate of the TMRCA was computed as $h_{(2,501)}\hat{L}$ +4 and the upper bound was computed as $h_{(97,500)}\hat{L}$ +4.

Supplementary References

1. Matise TC, Chen F, Chen W, *et al.* A second-generation combined linkage physical map of the human genome. Genome Res 2007;17:1783-6.
2. Flicek P, Aken BL, Beal K, *et al.* Ensembl 2008. Nucleic Acids Res 2008;36:D707-14.
3. Wakeley J. Coalescent Theory: An Introduction. Greenwood Village, Colorado: Roberts & Company Publishers; 2008.
4. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 2002;18:337-8.

Figure S1. Allelic loss of heterozygosity (LOH) and allele-specific expression (ASE) in samples from MLH1 c.1865T>A carriers. We analysed LOH in 6 tumor DNA samples and ASE in 2 cDNA samples. Normal LOH and ASE are shaded in grey. Ratio calculation was detailed in Supplementary Methods.

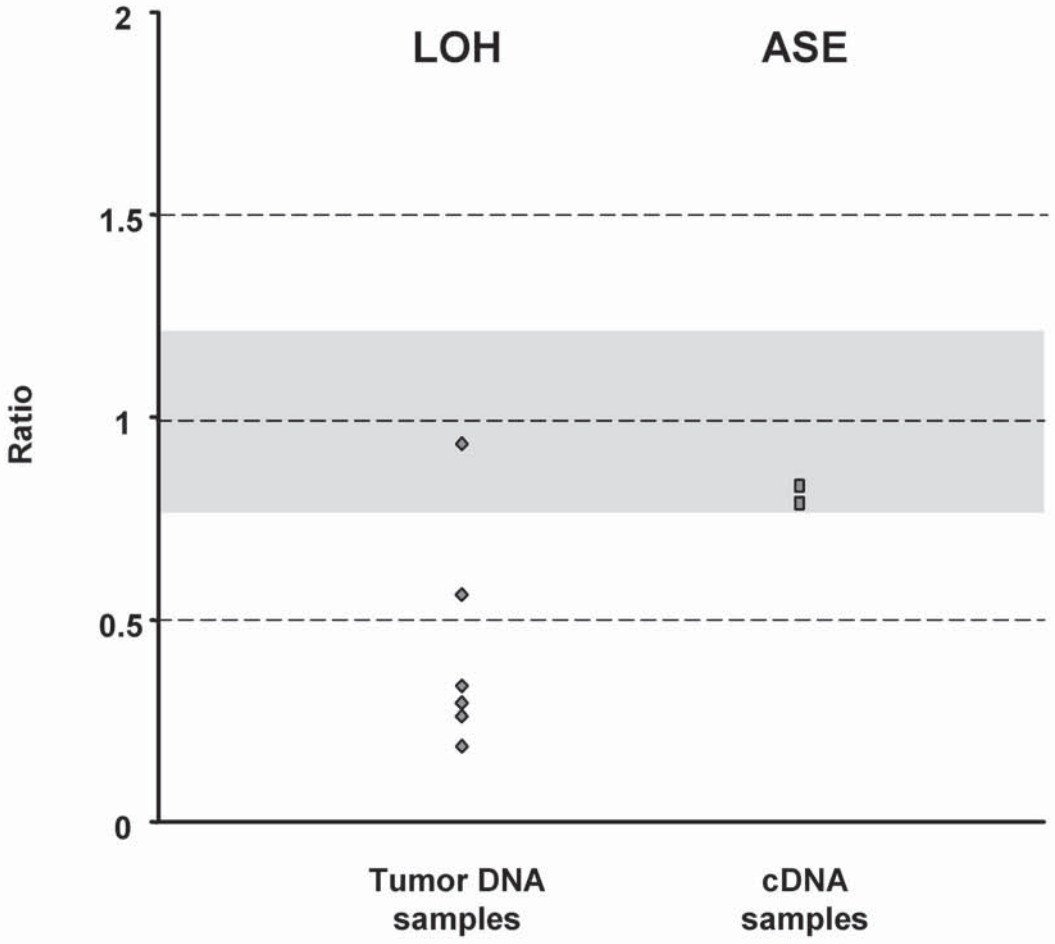


Table S1. Primers used in this study.

Primers	Application	Sequence	Hybridization temperature
D3S1611_up	Haplotype analysis	5' NED-CCCCAAGGCTGCACTT 3'	50 ºC
D3S1611_dw	Haplotype analysis	5' AGCTGAGACTACAGGCATTTG 3'	50 ºC
D3S3623_up	Haplotype analysis	5' VIC- CCCCCATGTTGGTTAAAGGC 3'	60 ºC
D3S3623_dw	Haplotype analysis	5' TCTCTGAACTGAAGTGACCTCCTG 3'	60 ºC
D3S2369_up	Haplotype analysis	5' FAM- CTCTATCCTTAATAAGTGGGTCACA 3'	56 ºC
D3S2336_dw	Haplotype analysis	5' AAATAGGGCCGTCCTTTG 3'	56 ºC
D3S1609_up	Haplotype analysis	5' NED-TGCTCTACACTGTGGCTTAATG 3'	58 ºC
D3S1606_dw	Haplotype analysis	5' ATATCCGTGGGCAAATGG 3'	58 ºC
D3S1612_up	Haplotype analysis	5' VIC-GCTCTCCTCAGTGGAAAATT 3'	58 ºC
D3S1612_dw	Haplotype analysis	5' ATGTAGAAGAGGATGATCTCCC 3'	58 ºC
D3S1298_up	Haplotype analysis	5' PET-AGCTCTCAGTGCCACCCC 3'	55 ºC
D3S1298_dw	Haplotype analysis	5' GAAAAATCCCCTGTGAAGCG 3'	55 ºC
D3S3564_up	Haplotype analysis	5' FAM-AGCTAAACACAGTCTAACTGCAT 3'	58 ºC
D3S3564_dw	Haplotype analysis	5' CCCACAGAGTGATAGGGA 3'	58 ºC
MLH1_E5_up	Haplotype analysis	5' TCTCTTTTCCCCTTGGGATT 3'	55 ºC
MLH1_E5_dw	Haplotype analysis	5' TGGAGGTCCCAAGATATCCTC 3'	55 ºC
MLH1_E8_up	Haplotype analysis	5' GTTTCAGTCTCAGCCATGAG 3'	55 ºC
MLH1_E8_dw	Haplotype analysis	5' CTGTGGCGTGAATCATGTGT 3'	55 ºC
MLH1_E15_up	Haplotype analysis	5' CATGAGTGGCAGCAAGCCAG 3'	60 ºC
MLH1_E15_dw	Haplotype analysis	5' TCCATATGCAAATCATACAATACAGCA 3'	60 ºC
MLH1_E3_CSCE_up	CSCE analysis	5' FAM-ATGGGAATTCAAAGAGATTTGG 3'	54 ºC
MLH1_E3_CSCE_dw	CSCE analysis	5' GTTTGCTCAGATTTGCATAC 3'	54 ºC
MLH1_E16_CSCE_up	CSCE analysis	5' HEX-GGATGCTCCGTTAAAGCT 3'	54 ºC
MLH1_E16_CSCE_dw	CSCE analysis	5 'GAAGTATAAGAATGGCTGTCAC 3'	54 ºC
MLH1_E3_up	Sequencing analysis	5' ATGGGAATTCAAAGAGATTTGG 3'	50 ºC
MLH1_E3_dw	Sequencing analysis	5' GTTTGCTCAGATTTGCATAC 3'	50 ºC
MLH1_E16_up	Sequencing analysis/SNuPE	5' GGATGCTCCGTTAAAGCTTG 3'	58 ºC
MLH1_E16_dw	Sequencing analysis/SNuPE	5' GAAGTATAAGAATGGCTGTAC 3'	58 ºC
MLH1_E2_exo_up	RNA analysis	5 'AGGGAGGCCTCAAGTTGATT 3'	56 ºC
MLH1_E4_exo_dw	RNA analysis	5' TCCATCAGCTGTTTTCGTTG 3'	56 ºC
MLH1_E15_exo_up	RNA analysis/SNuPE	5' GCCAATTTTGGTGTTCTCAG 3'	56 ºC
MLH1_E17_exo_dw	RNA analysis/SNuPE	5' CAAAGGGGGCACATAGTTGT 3'	56 ºC

Table S2. Age estimates for mutations c.306+5G>A and c.1865T>A. Quantities used in the age estimation and the haplotype frequencies in controls are also shown.

Mutation	Disease haplotype map length (cM)	Haplotype frequency in controls	Recombination count in diseased families	Number of families *n*	Multiplier *c*	Age estimate (generations)	95% confidence interval (generations)	Age estimate (years)
c.306+5G>A (Ebro)	0.9796	0.069	3	17	0.2324	75	[53-122]	1,879
c.1865T>A (Jaén)	2.2578	0.048	1	12	0.2569	15	[12- 22]	384

ARTICLE 2

COMPREHENSIVE FUNCTIONAL ASSESSMENT OF *MLH1* VARIANTS OF UNKNOWN SIGNIFICANCE

Ester Borràs, Marta Pineda, Angela Brieger, Innga Hinrichsen, Carolina Gómez, Matilde Navarro, Judit Balmaña, Teresa Ramón y Cajal, Asunción Torres, Joan Brunet, Ignacio Blanco, Guido Plotz, Conxi Lázaro i Gabriel Capellá

Human Mutation; doi: 10.1002/humu.22142 (2012).

Resum: La síndrome de Lynch està associada a mutacions germinals en els gens reparadors. Més del 30% dels canvis detectats en aquests gens són variants de significat biològic desconegut (VSD). El nostre objectiu era avaluar la patogenicitat de 8 VSD identificades en el gen *MLH1* en pacients amb sospita de síndrome de Lynch. Totes les variants eren noves o bé no havien estat prèviament caracteritzades. Per la seva classificació, es va seguir una estratègia que integra la història familiar, la patologia del tumor, la freqüència en controls i una varietat d'anàlisi *in silico* i *in vitro* tant a nivell de RNA com de proteïna, com per exemple els assaigs de reparació, l'expressió de les proteïnes MLH1 i PMS2, i la localització subcel·lular. Cinc de les variants van ser classificades com a patogèniques: les variants c.[248G>T(;)306G>C], c.[780C>G;788A>C] i c.791-7T>A afectaven el correcte processament del mRNA, mentre que les variants c.218T>C (p.L73P) i c.244G>A (p.T82A) mostraven una activitat reparadora deficient. Dues més van ser considerares probablement neutres: la variant silent c.702G>A no mostrava cap efecte a nivell de processament ni estabilitat del mRNA, i la variant c.974G>A (p.R325Q) no presentava cap defecte ni a nivell de mRNA ni de proteïna. Per contra, la variant c.25C>T (p.R9W) no va poder ser classificada, ja que estava associada a valors intermedis d'activitat reparadora. Els anàlisi funcionals exhaustius de les VSD en el gen *MLH1* han estat útils per a la classificació i són rellevants pel diagnòstic i el consell genètic de les famílies portadores.

RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Comprehensive Functional Assessment of *MLH1* Variants of Unknown Significance

Ester Borràs,[1] Marta Pineda,[1]* Angela Brieger,[2] Inga Hinrichsen,[2] Carolina Gómez,[1] Matilde Navarro,[1] Judit Balmaña,[3] Teresa Ramón y Cajal,[4] Asunción Torres,[5] Joan Brunet,[6] Ignacio Blanco,[1] Guido Plotz,[2] Conxi Lázaro,[1] and Gabriel Capellá[1]*

[1]*Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL, Hospitalet de Llobregat, Spain;* [2]*Medical Clinic 1, Johann Wolfgang Goethe-University Clinic, Frankfurt, Germany;* [3]*Medical Oncology Service, Hospital Universitari Vall d'Hebron, Barcelona, Spain;* [4]*Medical Oncology Service, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain;* [5]*Unitat de Consell Genètic, Hospital Universitari Sant Joan, Reus, Spain;* [6]*Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IdIBGI, Girona, Spain*

Communicated by Finlay A. Macrae

Received 19 January 2012; accepted revised manuscript 29 May 2012.
Published online 26 June 2012 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22142

ABSTRACT: Lynch syndrome is associated with germline mutations in DNA mismatch repair (MMR) genes. Up to 30% of DNA changes found are variants of unknown significance (VUS). Our aim was to assess the pathogenicity of eight *MLH1* VUS identified in patients suspected of Lynch syndrome. All of them are novel or not previously characterized. For their classification, we followed a strategy that integrates family history, tumor pathology, and control frequency data with a variety of in silico and in vitro analyses at RNA and protein level, such as MMR assay, MLH1 and PMS2 expression, and subcellular localization. Five *MLH1* VUS were classified as pathogenic: c.[248G>T(;)306G>C], c.[780C>G;788A>C], and c.791-7T>A affected mRNA processing, whereas c.218T>C (p.L73P) and c.244G>A (p.T82A) impaired MMR activity. Two other VUS were considered likely neutral: the silent c.702G>A variant did not affect mRNA processing or stability, and c.974G>A (p.R325Q) did not influence MMR function. In contrast, variant c.25C>T (p.R9W) could not be classified, as it associated with intermediate levels of MMR activity. Comprehensive functional assessment of *MLH1* variants was useful in their classification and became relevant in the diagnosis and genetic counseling of carrier families.
Hum Mutat 00:1–13, 2012. © 2012 Wiley Periodicals, Inc.

KEY WORDS: Lynch syndrome; *MLH1*; variants of unknown significance; functional characterization

Introduction

Lynch syndrome is the most common of the known colorectal cancer (CRC) predisposing syndromes. It is an autosomal-dominant hereditary disease caused by heterozygous germline mutations in DNA mismatch repair (MMR) genes: *MLH1* (MIM# 120436), *MSH2* (MIM# 609309), *MSH6* (MIM# 600678), or *PMS2* (MIM# 600259) [Lynch et al., 2008; Vasen, 2007]. Clinically, Lynch syndrome is characterized by early-onset CRC and an increased risk of other associated tumors [Lynch et al., 2008; Watson and Riley, 2005]. Tumors of Lynch syndrome spectrum are associated with microsatellite instability (MSI) and loss of MMR proteins expression [Lynch et al., 2008]. Amsterdam II and revised Bethesda criteria are currently used to identify patients at risk [Boland et al., 2008; Umar et al., 2004].

Many MMR gene mutations are short insertions, deletions, or nucleotide substitutions that create a premature stop codon, or nucleotide changes at the consensus splice site (SS) that affect mRNA splicing [Peltomaki and Vasen, 2004]. These mutations, frequently encoding a truncated protein, are classified as pathogenic offering a good clinical tool for diagnosis and genetic counseling. However, clinical implications of missense or silent variants or mutations close to SS are often not clear. Therefore, those variants are classified as variants of unknown significance (VUS) and represent up to 30% of the identified DNA changes in MMR genes [Peltomaki and Vasen, 2004].

DNA variants may exert their pathogenic effect by affecting transcription or protein function. At the RNA level, variants may affect transcript stability as well as mRNA processing by influencing consensus SS, by creating cryptic SS, or by altering binding sites for splicing enhancer proteins [Barnetson et al., 2008; Spurdle et al., 2008]. At the protein level, variants may affect MMR protein function as a result of altered protein expression or stability, defective subcellular localization, mismatch recognition, complex formation, or inactivation of enzymatic activity. Therefore, the assessment of pathogenicity of VUS is based on complex strategies that combine analyses both at RNA and protein levels. In addition, it is currently accepted that variant classification requires integration of data not only from functional assays, but also from clinicopathological features, frequency in control population and cosegregation, to determine a probability of pathogenicity [Barnetson et al., 2008; Couch et al., 2008; Goldgar et al., 2008; Kansikas et al., 2011; Ou et al., 2007; Pastrello et al., 2011; Plon et al., 2008].

The aim of this study was to assess the pathogenicity of 8 *MLH1* VUS: c.25C>T (p.R9W), c.218T>C (p.L73P),

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Gabriel Capellá and Marta Pineda, Hereditary Cancer Program, Catalan Institute of Oncology: ICO-IDIBELL, Gran Via 199-203, 08907- L'Hospitalet de Llobregat, Barcelona, Spain. E-mail: gcapella@iconcologia.net and mpineda@iconcologia.net

Contract grant sponsor: Ministerio de Ciencia e Innovación (SAF 06-06084, 09-07319); Fundació Gastroenterologia Dr. Francisco Vilardell (F05-01); Ministerio de Educación y Ciencia Spanish Networks RTICC (RD06/0020/1050, 1051); Acción en Cáncer (Instituto de Salud Carlos III); the Autonomous Government of Catalonia (2009-SGR-290); Fundación Científica de la AECC.

c.244A>G (p.T82A), c.[248G>T(;)306G>C] (p.[S83I(;)E102D]), c.[702G>A];[780C>G;788A>C] (p.[=];[N263T]), c.791-7T>A, and c.974G>A (p.R325Q). Web-prediction programs were combined with the analyses of mRNA splicing and stability and the evaluation of cosegregation and frequency of this VUS in controls. Furthermore, MMR activity, protein expression, and subcellular localization were studied.

Material and Methods

Patients and Samples

Mutational analysis of *MMR* genes was performed in 258 patients either fulfilling Amsterdam II or Bethesda criteria, associated with MSI phenotype in tumors, attended at distinct Spanish Genetic Counseling Units. In all, 119 pathogenic mutations and 18 VUS in any of the 4 MMR genes were identified (unpublished results). Sixty-nine pathogenic mutations were detected in the *MLH1* gene (NG_007109.1, NM_000249.3, and NP_000240.1), whereas eight additional *MLH1* changes were VUS (Table 1). The identified *MLH1* VUS have been submitted to the LOVD database (www.LOVD.nl). Mutation nomenclature of *MLH1* gene is according to HGVS recommendations (version 2.0) with nucleotide 1 corresponding to the A of the ATG-translation initiation codon.

Clinical data collection included the location of tumors, the age at diagnosis, and the result of MSI testing, immunohistochemistry analysis, and *BRAF* p.V600E mutation testing. Informed consent was obtained from all individuals. DNA samples from a hospital-based CRC case-control study (325 cases and 309 controls) [Moreno et al., 2006] were used to analyze the frequency of the detected *MLH1* VUS.

Analysis of Frequency in Controls and CRC Cases

Screening for the identified *MLH1* variants in control and sporadic CRC population was performed by Conformation-Sensitive Capillary Electrophoresis (CSCE) (Supp. Table S1, conditions available upon request). Patterns differing from those of controls were amplified using nonfluorescent forward primer and sequenced using the Big Dye Terminator v.3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA).

Cosegregation Analysis

DNA of relatives was screened for the familial associated *MLH1* variant by direct sequencing using genomic DNA primers (details available on request). Cosegregation was considered positive if two or more affected members were carriers of the studied variant.

Bioinformatic Analysis

DNA sequences containing the identified *MLH1* variants were analyzed with several bioinformatic tools addressed to evaluate its impact at RNA and protein level. To identify potential splicing mutations, disruption or creation of SS were evaluated using NNSplice [Reese et al., 1997], Splicerport [Dogan et al., 2007], NetGene2 Server [Hebsgaard et al., 1996], and SoftBerry [Burset et al., 2001]. Also, exonic variants were analyzed for putative ESE (exonic splicing enhancer) elements by ESEfinder [Cartegni et al., 2003], RESCUE-ESE [Fairbrother et al., 2002], and PESX [Zhang and Chasin, 2004; Zhang et al., 2005].

The impact of missense variants at the protein level was analyzed using PolyPhen-2 [Adzhubei et al., 2010], SIFT [Ng and Henikoff, 2002], and MAPP-MMR [Chao et al., 2008]. In addition, the putative effect was investigated in a structural model of the MutLα N-terminal heterodimer, which was based on a homology model of MLH1-NTD [Plotz et al., 2006] and a crystal structure of PMS2-NTD [Guarne et al., 2004]. Figures were created using the PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC. Conservation of residues was visualized using WebLogo [Crooks et al., 2004] based on an alignment of the MLH1 protein previously described [Plotz et al., 2008].

Lymphocyte Culture and mRNA Splicing Analysis

Human lymphocytes from *MLH1* variant carriers were cultured in RPMI 1640 medium (Invitrogen, Carlsbad, CA) supplemented with 10% fetal calf serum (PAA Laboratories, Pasching, Austria) and 1% penicillin-streptomycin (Sigma, Saint Quentin Favallier, France). Cells were cultured in the absence and presence of puromycin (Sigma). Puromycin was used to prevent potential degradation of unstable transcripts by the nonsense-mediated decay mechanism.

Total RNA was extracted from cultured lymphocytes using Trizol reagent (Invitrogen, Carlsbad, CA), 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 (Supp. Table S1; conditions available upon request). Polymerase chain reaction (PCR) products were sequenced on an ABI 3130 sequencer (Applied Biosystems) using Big Dye Terminator v3.1 cycle sequencing reaction kit (Applied Biosystems). To improve the detection and interpretation of splicing aberrations, transcripts from carriers were compared with transcripts from three control lymphocyte cultures.

Allele-Specific Expression Analysis

Allele-specific expression (ASE) was analyzed by single nucleotide primer extension (SNuPE) as previously described [Borras et al., 2010]. ASE was calculated by dividing the proportion of variant/wild-type allele in cDNA by the proportion of variant/wild-type allele in gDNA. We used $\leq$0.5 as a threshold value for ASE definition [Perera et al., 2010; Renkonen et al., 2003, 2005]. Experiments were performed in triplicate.

Plasmids and Site-Directed Mutagenesis

pcDNA3.1_MLH1 and pN1_PMS2 plasmids, kindly provided by Dr. Kolodner and Dr. Nyström-Lahti, were used in MMR assays and expression analyses. pECFP_C1_MLH1 and pDSred_C1_PMS2 vectors were used in subcellular localization assays [Brieger et al., 2011].

Selected *MLH1* missense variants were constructed by site-directed mutagenesis using QuikChange Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. Sequencing was used to verify the presence of the variants. In addition, three control plasmids were constructed: c.199G>A (p.G67R) and c.2246T>C (p.L749P), used as mutation controls, and c.655A>C (p.I219V) used as a polymorphism control.

Table 1. Clinical Characteristics of Cancer Affected Carriers of the Identified VUS in the *MLH1* Gene

MLH1 VUS	*MLH1* location	Reported at LOVD	Previously reported functional assays	Family ID	Criteria met	Cosegregation	Affected carrier ID	Gender	Type of tumor	Age of onset	MSI	IHC				BRAF
												MLH1	MSH2	MSH6	PMS2	
c.25C>T	Exon 1	No	No	A	BC	NE	A1*’	M	CRC	49	+	+	+			
c.218T>C	Exon 3	No	Ellison et al. (2004) (In yeast)	B	BC	NE	B1*	M	CRC	36	+	NE	+			
c.244A>G	Exon 3	No	No	C	BC	NE	C1*^	F	CRC	41	+	+	+	+	+	
c.[248G>T(;)306G>C]	Exon 3	No	No	D	BC	NE	D1*	M	CRC	45	+	–	+	+		WT
c.[702G>A];[780C>G;788A>C]	Exon 9	Yes, c.702G>A	No	E	AC	NE	E1*	F	EC	54	+	–	+	+		WT
c.791-7T>A	Intron 9	No	No	F	AC	Yes	F1*	M	CRC	62	+	–	+	NE		WT
									CRC	71						
									CRC	72						
							F2	F	OC	40	+	–	+	NE		
							F3	M	CRC	40						
							F4	F	BC	38	–					
c.974G>A	Exon 11	Yes	No	G	BC	Yes	G1	M	CRC	72	–					
							G2*	M	CRC	65	+	+	+	+	+	WT
									CRC	65						
				H	AC	Yes	H1	F	CRC	41	–	+	+	NE		
							H2*	M	CRC	70						

*Index case’ Individuals with an additional germline pathogenic mutation in *APC* gene (c.1958+3A>G); ^Individual with an additional germline VUS in *MSH6* gene (c.1186C>G; p.L396V).
Abbreviations: AC, Amsterdam criteria; BC, Bethesda criteria; M, male; F, female; BC, breast cancer; CRC, colorectal cancer; EC, endometrial cancer; OC, ovarian cancer; MSI, microsatellite instability: +, instable; –, stable; IHC, immunohistochemical analysis of MMR proteins in tumor tissue: + expression; – loss of expression; NE, not evaluable result; WT, wild type.

HEK293T Cells Culture and Cell Transfection

HEK293T cells (deficient for endogenous MLH1 and PMS2 [Trojan et al., 2002]) were kindly provided by Prof. Jiricny. Cells were grown in Dulbecco's Modified Eagle Medium (Invitrogen) with 10% fetal calf serum (PAA Laboratories) and 1% penicillin-streptomycin (Sigma). Cell lines were routinely tested for mycoplasma and verified by morphology, growth curve analysis, and expression of proteins (e.g., MLH1 and PMS2).

Transfection of HEK293T cells was carried out as described previously [Brieger et al., 2011]. In brief, HEK293T cells were transfected at 50–70% confluence with expression plasmids (1 μg/ml, respectively) using 10 μl/ml of the cationic polymer polyethylenimine (Polysciences, Warrington, PA; stock solution 1 mg/ml). After 48 hr, cells were prepared for confocal laser scanning microscopy or protein extraction [Plotz et al., 2006].

MMR Assay

MMR reactions were performed as previously described [Plotz et al., 2006] using a plasmid substrate with a G–T mismatch within an *EcoRV* restriction site. A nick is located at 83-bp distance to the mismatch in 3′-orientation, and serves to direct mismatch repair events. Restoration of the *EcoRV* site occurs when DNA is repaired. Digestion with *AseI* was used to linearize the plasmid and *EcoRV* was used to assess repair efficiency. Restriction site products were separated on 2% agarose gels and bands were quantified using ImageLab software (Bio-Rad, Hercules, CA). Repair efficiency was measured as the quotient of the intensities of those bands indicating repair divided by the sum of all band intensities. Relative repair efficiency was calculated by dividing the value of the tested variant protein through the value of a wild-type protein that had been expressed, processed, and tested in parallel. Experiments were performed in quadruplicate.

MLH1 and PMS2 Protein Expression Analysis

MLH1 and PMS2 expression levels were examined by SDS-PAGE, followed by Western blotting analysis with anti-MLH1 (clone G168-728 BD Biociences, San Jose, CA) and anti-PMS2 (clone E-19, Santa Cruz Biotechnology, Santa Cruz, CA) antibodies. Band intensities were quantified using MultiGauge v3.2 (Fujifilm, Tokyo, Japan). Beta-actin and EGFP expression was assessed in parallel and used as loading and transfection efficiency controls. All experiments were performed in quadruplicate.

Confocal Laser Microscopy

Confocal laser microscopy was carried out as previously described [Brieger et al., 2011]. Mock control for MLH1 and PMS2 was performed using empty pECFP-C1 and pDSRed-C1 vectors. Experiments were performed in triplicate.

RESULTS

Clinical Characteristics of Affected Carriers of *MLH1* VUS

Mutational analysis of *MMR* genes in 258 patients with suspected Lynch syndrome led to the identification of 8 index cases carriers of *MLH1* VUS. All of these variants are novel or not previously characterized (Table 1). Tumor location and age at diagnosis of affected carriers of *MLH1* variants are shown in Table 1. Most of the tumors in index cases were colorectal, showing MSI and/or loss of MLH1 protein expression.

In three Amsterdam II families, we detected four different VUS, c.791-7T>A, c.974G>A (p.R325Q), and c.[702G>A];[780C>G;788A>C] (p.[=];[N263T]). The c.791-7T>A variant was identified in family F (Supp. Fig. S1 and Table 1). This variant was present in two individuals affected by CRC (carriers F1 and F3), one affected by ovarian cancer (F2) and another diagnosed of breast cancer (carrier F4). One CRC and the ovarian cancer showed MSI and MLH1 loss of expression, but breast cancer tumor did not show MSI.

The c.974G>A variant was initially identified in a patient of family G (carrier G2), whose CRC tumor showed MSI without loss of expression of MLH1 protein. However, in another affected carrier of the same family (carrier G1), the CRC was microsatellite stable (MSS). Eventually, the same variant was identified in another family (family H) meeting Amsterdam criteria type II. In this pedigree, the variant was present in two individuals affected by CRC (carriers H1 and H2) (Supp. Fig. S1). The only tumor analyzed was MSS (carrier H1).

In family E, we identified three *MLH1* VUS in patient E1, affected by endometrial cancer showing MSI and loss of MLH1 expression (Table 1). Segregation analysis in this family demonstrated that c.702G>A variant co-occurred in-trans with variant c.[780C>G;788A>C] (p.N263T) in individual E1 (Supp. Fig. S1).

Three individuals were carriers of VUS located in *MLH1* exon 3 (Table 1). In family B, variant c.218T>C (p.L73P) was identified in an individual with CRC. This tumor showed MSI but MLH1 expression was not evaluable. In family C, individual C1 was a carrier of two variants in distinct MMR genes, the *MLH1* VUS c.244A>G (p.T82A) and *MSH6* VUS c.1186C>G (p.L396V). The CRC of this patient showed MSI without loss of expression of any MMR protein. In family D, two *MLH1* VUS were identified in patient D1, c.[248G>T(;)306G>C], for which allelic phase was unknown. Tumor of this patient showed MSI and loss of MLH1 expression.

Finally, the *MLH1* variant c.25C>T (p.R9W) was identified in one affected individual of family A showing mixed features of Lynch syndrome and adenomatous polyposis. MMR gene testing was performed in the index case after he was diagnosed of an MSI CRC at age 46 associated with multiple polyps. After colonoscopic screening, his son was diagnosed with familial adenomatous polyposis (FAP). Subsequently, mutational *APC* testing was performed and an additional germline mutation was detected in the *APC* gene (c.1958+3A>G), leading to increased exon 14 skipping [Aretz et al., 2004]. Both father and son were carriers of the *MLH1* variant c.25C>T and the *APC* mutation c.1958+3 A>G (Supp. Fig. S1).

Frequency of *MLH1* VUS in Healthy Controls and Sporadic CRC

The *MLH1* screening by CSCE in a series of control individuals and CRC cases was used to determine the allelic frequencies of the identified *MLH1* VUS (Table 2). None of the variants were identified at polymorphic frequency (>1%) in the control population. In sporadic CRC cases, c.[248G>T(;)306G>C] and c.702G>A variants were observed at 0.16% and 0.31%, respectively. The CSCE screening in the case-control series lead to the identification of other *MLH1* DNA changes (Supp. Table S2).

Table 2. Allelic Frequency in Control and Colorectal Cancer (CRC) Cases

Family ID	*MLH1* VUS	Number of control alleles	Control frequency (95% CI)	Number of CRC case alleles	CRC Case frequency (95% CI)
A	c.25C>T	604	0	648	0
B	c.218T>C	560	0	620	0
C	c.244G>A	560	0	620	0
D	c.[248G>T(;)306G>C]	560	0	620	0.00161(0.00008-0.01048)[a]
E	c.702G>A	604	0	648	0.00308 (0.00053 0.01237)
	c.[780C>G;788A>C]	604	0	648	0
F	c.791-7T>A	604	0	648	0
G and H	c.974G>A	604	0	648	0

[a]We detected c.248G>T and c.306G>C together.

Bioinformatic Prediction and mRNA Analysis Associated with *MLH1* Variants

Four programs were used to predict changes in scores of acceptor and donor *MLH1* SS (Table 3). For the c.[248G>T(;)306G>C] variant SplicePort program predicted the disruption of the donor site, whereas a decrease in the score was predicted by NetGene2 and Softberry (22% and 37%, respectively). For the c.[780C>G;788A>C] variant, a relevant decrease in the donor site score (45%) was only predicted by Spliceport. For the c.791-7T>A variant, NetGene2 predicted a 50% decrease in the acceptor site score, whereas the other programs predicted its complete disruption. In this case, the four programs predicted the creation of a new acceptor site, 5 base-pairs upstream. In silico analysis of the remaining variants predicted either no changes, small decreases (<14%), or increases in SS scores.

Three programs were used to predict the putative effect of exonic *MLH1* VUS on exonic splicing enhancers, ESE finder, RescueESE, and PESX (Table 3). All programs predicted ESE modifications for the c.702G>A variant, and two of the three programs predicted alterations for variant c.[780C>G;788A>C]. However, inconsistencies were observed regarding the predicted effect among programs.

Our reverse transcriptase (RT)-PCR analyses identified aberrantly expressed transcripts in three of the eight *MLH1* studied variants. *MLH1* c.[248G>T(;)306G>C] variant resulted in exon 3 skipping (r.208_306del), which is predicted to generate an in-frame deletion of 33 amino acids (p.K70_E102del) in the ATPase domain of MLH1 (Fig. 1A). *MLH1* c.[780C>G;788A>C] variant abolished the natural donor SS of exon 9 leading to exon 9 skipping (r.678_790del) (Fig. 1B). This change is predicted to generate a truncated protein (p.E227SfsX42). In both cases, sequencing of the RT-PCR products showed that variant alleles exclusively transcribed the aberrant transcripts (data not shown). The intronic *MLH1* c.791-7T>A variant lead to exon 10 skipping (r.791_884del) and is also predicted to truncate MLH1 protein (p.H264LfsX2) (Fig. 1C). Surprisingly, puromycin treatment of cultured lymphocytes did not seem to increase the amount of the frameshift transcripts for the latter two variants. The remaining five *MLH1* VUS included in this study had no apparent effect on mRNA splicing after gel analysis and sequencing of the RT-PCR products (data not shown).

We evaluated the allelic expression of *MLH1* in lymphocytes derived from individuals carrying the missense variants c.218T>C, c.244A>G, and c.974G>A, and the silent c.702G>A, which had not shown any apparent effect on mRNA processing in our RT-PCR analyses (Table 3). The expression of variant/wild-type allele in cDNA samples from lymphocytes cultured in the presence or absence of puromycin did not show allelic imbalances (Supp. Table S3). ASE of the c.25C>T variant was not assessed because of the poor quality of its RNA.

In summary, our RNA analyses allowed to classify *MLH1* variants [248G>T(;)306G>C], c.[780C>G;788A>C], and c.791-7T>A as pathogenic mutations affecting the mRNA processing. The silent c.702G>A variant was classified as likely neutral because it did not affect mRNA processing or stability.

In Silico and In Vitro Protein Analyses of *MLH1* Missense Variants

MLH1 c.218T>C (p.L73P) and c.244A>G (p.T82A) missense variants were classified as pathogenic mutations by the three in silico prediction programs used (Table 4). We also evaluated the putative role of the affected residues on MLH1 protein structure (Fig. 2). The p.L73P variant affects a highly conserved amino acid located in helix B, which is involved in forming the hydrophobic core of the extensive β-sheet that resembles the backbone of the ATPase pocket. The mutation to proline likely disturbs the constitution of this backbone. The threonine 82 is located in a completely conserved loop (TSK-loop) which is directly involved in nucleotide binding in the MLH1 ATPase pocket with a side-chain hydrogen bond [Ban et al., 1999].

The c.25C>T (p.R9W) variant, located at the N-terminus, was classified as pathogenic by PolyPhen-2 and SIFT but not by MAPP-MMR. This arginine residue shows an intermediate degree of conservation, where positively charged residues predominate (Fig. 2). It is located in the flexible, nonstructured N-terminal tail of MLH1, which gets ordered after N-terminal dimerization [Ban et al., 1999]. Finally, c.974G>A (p.R325Q) variant affects a nonconserved residue located at a α helix in the MLH1 transducer domain. Concordantly, it was predicted as nonpathogenic by the three programs analyzed (Table 4 and Fig. 2).

As reported, control pathogenic mutations c.199G>A (p.G67R) and c.2246T>C (p.L749P) were inactive in MMR assays, whereas the control polymorphism c.655A>C (p.I219V) was proficient [Drost et al., 2010; Kosinski et al., 2010; Plotz et al., 2008; Raevaara et al., 2005; Takahashi et al., 2007; Trojan et al., 2002]. Our results showed a significantly reduced MMR activity for c.218T>C (p.L73P) and c.244A>G (p.T82A) variants (8.24% ± 4.20 and 12.2% ± 3.97, respectively). The c.25C>T (p.R9W) variant instead showed intermediate MMR activity values (43.66% ± 12.64) and c.974G>A (p.R325Q) did not show differences compared to MLH1 wild type (99.97% ± 9.92) (Fig. 3).

Protein expression analysis detected low MLH1 and PMS2 expression levels for c.199G>A (p.G67R) and c.2246T>C (p.L749P) control mutations (Fig. 3); again, the control polymorphism c.655A>C (p.I219V) displayed MLH1 and PMS2 levels similar to wild type [Blasi et al., 2006; Plotz et al., 2008; Raevaara et al., 2005; Takahashi et al., 2007], confirming the reliability of the method. The missense

Table 3. Bioinformatic Prediction and In Vitro Analysis of Splicing Aberrations Associated with *MLH1* Variants

Family ID	MLH1 VUS	Predicted Protein	Distance to the nearest SS	SS	Splice Site Prediction									Enhancer site prediction				cDNA splicing analysis	cDNA stability analysis (ASE)	Classification at RNA level
					NNSplice		Spliceport		NetGene2		SoftBerry		Interpretation	ESEfinder	RESCUE-ESE	PESX	Interpretation			
					wild-type	variant	wild-type	variant	wild-type	variant	wild-type	variant								
A	c.25C>T	p.R9W	-92	D	0.93	0.93	NR	NR	0.95	0.95	7.36	7.36	No effect	No change	No change	No change	No effect	C>T substitution only	NA	Likely neutral
B	c.218T>C	p.L73P	+11	A	0.67	0.73	0.36	0.76	0.28	0.38	6.85	6.85	No effect	No change	No change	ESE destroyed.	Inconclusive	T>C substitution only	No effect	Likely neutral
				D	1.00	1.00	0.85	0.85	0.86	0.86	13.24	13.24								
C	c.244G>A	p.T82A	+37	A	0.67	0.67	0.36	0.31	0.28	0.33	6.85	6.85	No effect	1 site destroyed. 2 sites increased. 1 site reduced	No change	No change	Inconclusive	G>A substitution only	No effect	Likely neutral
				D	1.00	1.00	0.85	0.89	0.86	0.87	13.24	13.24								
D	c.[248G>T(;)306G>C]	p.[S83I(;)E102D]	+41; 0	A	0.67	0.67	0.36	0.55	0.28	0.31	6.85	6.85	Inconclusive	1 site destroyed. 1 site increased	No change	No change	Inconclusive	r.208_306del; p.K70_E102del (exon 3 skipping)	NP	Pathogenic
				D	1.00	0.93	0.85	NR	0.86	0.67	13.24	8.34								
E	c.702G>A	p.(=)	+25	A	0.88	0.88	0.32	0.61	0.41	0.43	NR	NR	No effect	3 sites destroyed	1 site destroyed. 3 sites created	ESE destroyed. 3 ESE created	Aberrant ESE	G>A substitution only	No effect	Likely neutral
				D	0.96	0.96	0.98	0.98	0.97	0.99	12.54	12.54								
	c.[780C>G;788A>C]	p.N263T	-5; -2	A	0.88	0.88	0.32	0.32	0.41	0.41	nr	nr	Inconclusive	1 site created. 4 sites increased	1 site destroyed	No change	Inconclusive	r.678_790del; p.E227SfsX42 (exon 9 skipping)	NP	Pathogenic
				D	0.96	0.96	0.98	0.53	0.97	0.99	12.54	12.40								
F	c.791-7T>A		-7	A	0.99	NR	0.48	NR	0.33	0.16	8.75	NR	Aberrant SS	NA	NA	NA	NA	r.791_884del; p.H264LfsX2 (exon 10 skipping)	NP	Pathogenic
				A*	NR	0.99	NR	0.10	NR	0.16	NR	9.57								
				D	0.93	0.93	0.53	0.53	0.44	0.41	12.68	12.68								
G and H	c.974G>A	p.R325Q	-65	A	0.86	0.86	1.13	1.13	0.95	0.95	7.47	7.47	No effect	1 site destroyed. 2 sites created	No change	No change	Inconclusive	G>A substitution only	No effect	Likely neutral
				D	0.98	0.98	0.67	0.67	0.94	0.94	7.64	7.64								

Results highlighted in gray indicate major alterations. For splice site prediction, major alternations are destroyed or created SS or score modifications ≥45%; For enhancer site prediction, major alterations are destroyed or created ESE. Predictions are interpreted as inconclusive when the same results are not obtained by all the programs used. Abbreviations: SS, splice site; A, acceptor consensus splice site; D, donor consensus splice site; A*, new acceptor splice site; NR, consensus splice site not recognized; NA, not available; NP, not performed.

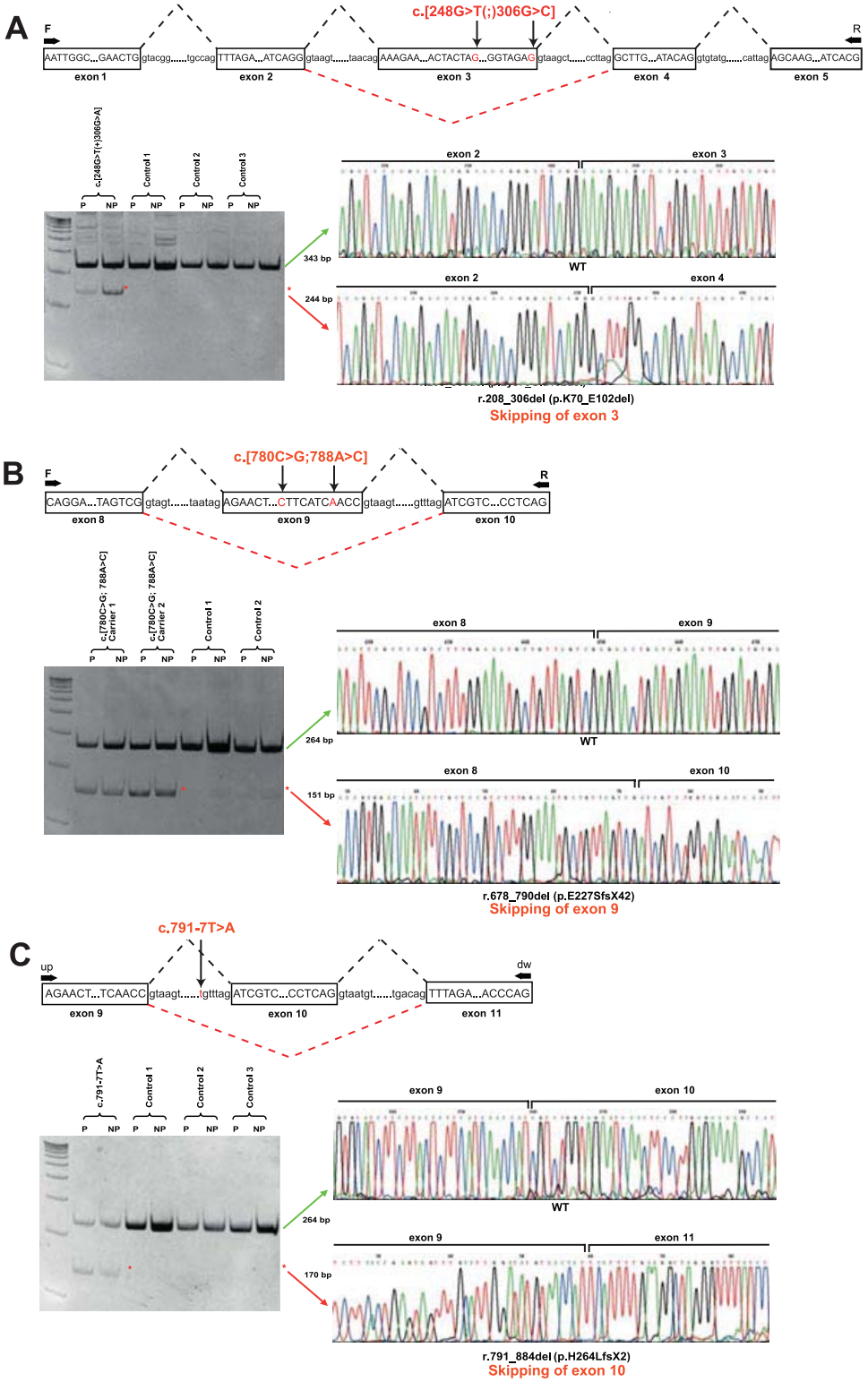


Figure 1. Analysis of *MLH1* VUS at mRNA level. Characterization of *MLH1* variant: **A:** c.[248G>T(+)306G>C], **B:** c.[780C>G;788A>C], and **C:** c.791-7T>A. Upper lines: Schematic overview of *MLH1* exons; left: agarose gel showing RT-PCR products; right: direct sequencing of the RT-PCR products. Abbreviations: P, puromycin; NP, no puromycin; F, forward primer; R, reverse primer.

Table 4. Summary of the In Vitro Functional and In Silico Analyses at Protein Level in *MLH1* Exonic Variants without Associated Splicing Aberrations

Family ID	*MLH1* VUS	Predicted Protein	Predicted impact on protein function				MMR assay	MLH1 expression	PMS2 expression	Subcellular localization		Classification at Protein level
			Functional domain	PolyPhen-2 (score)	SIFT (score)	MAPP-MMR (score)	Interpretation				MLH1	PMS2	
A	c.25C>T	p.R9W	ATPase	PrD (0,97)	Aff (0,01)	Bor (4,48)	Inconclusive	++	+	+	N	N	Inconclusive
B	c.218T>C	p.L73P	ATPase	PrD (0,94)	Aff (0,00)	Del (17,150)	Impaired	+	+	+	N	N	**Pathogenic**
C	c.244G>A	p.T82A	ATPase	PrD (0,99)	Aff (0,00)	Del (36,720)	Impaired	+	+++	+++	N	N	**Pathogenic**
G and H	c.974G>A	p.R325W	Transducer Domain	Ben (0,07)	Tol (0,53)	Neu (2,040)	Neutral	+++	+++	+++	N	N	Likely neutral

Results in bold with gray highlight indicate major alterations. Predictions are interpreted as inconclusive when the same results are not obtained by all the programs used. Functional domain, as described in Mohd et al. (2006). In vitro MMR activity: +++ corresponds to MMR activity >60% compared to wild type, ++ 30–60%, + 0–30%. Expression of MLH1 and PMS2 +++ corresponds to >60% comparing to wild type, ++30–60%, + <30%.
Abbreviations: PrD, Probably Damaging; Ben, Benign; Aff, Affected; Tol, Tolerant; Bor, Borderline; Del, deleterious; Neu, neutral; N, nuclear.

Figure 2. Localization of MLH1 variants in the N-terminal domain of MLH1-PMS2 dimer. **A:** Mapping of amino acid substitutions onto the structural model of N-terminal domain of MLH1-PMS2 dimer. Substitutions occurred in the residues whose Cα atoms are indicated as spheres, and colored according to their previous classification (red: mutation control, green: polymorphism, yellow: VUS). **B:** Magnification of the ATPase domain. **C:** Conservation graphic for variants analyzed at protein level. Affected residues are indicated by yellow letters or arrows. Overall/individual font sizes indicate overall/individual conservation, respectively.

p.R9W and p.L73P variants showed a strong reduction of MLH1 and PMS2 expression (13.29% ± 4.75 and 20.56% ± 18.21 for p.R9W and 18.61% ± 11.41 and 30.02% ± 23.8 for p.L73P) (Fig. 3). In contrast, no decrease in protein expression was observed for p.T82A and p.R325Q variants.

The results of subcellular localization analysis in HEK293T cells are shown in Figure 4. As previously described [Brieger et al., 2012], in the absence of PMS2, MLH1 was almost entirely detected in the nucleus. In contrast, single expressed PMS2 was located in the cytoplasm. Coexpression of MLH1 and PMS2 proteins lead to nuclear localization of both proteins. Control mutation p.G67R coexpressed with PMS2 showed nuclear as well as cytoplasmatic location of MLH1 and PMS2, as previously described [Raevaara et al., 2005], whereas both other controls (p.I219V and p.L749P) showed only nuclear location of MLH1 and PMS2 (data not shown). All

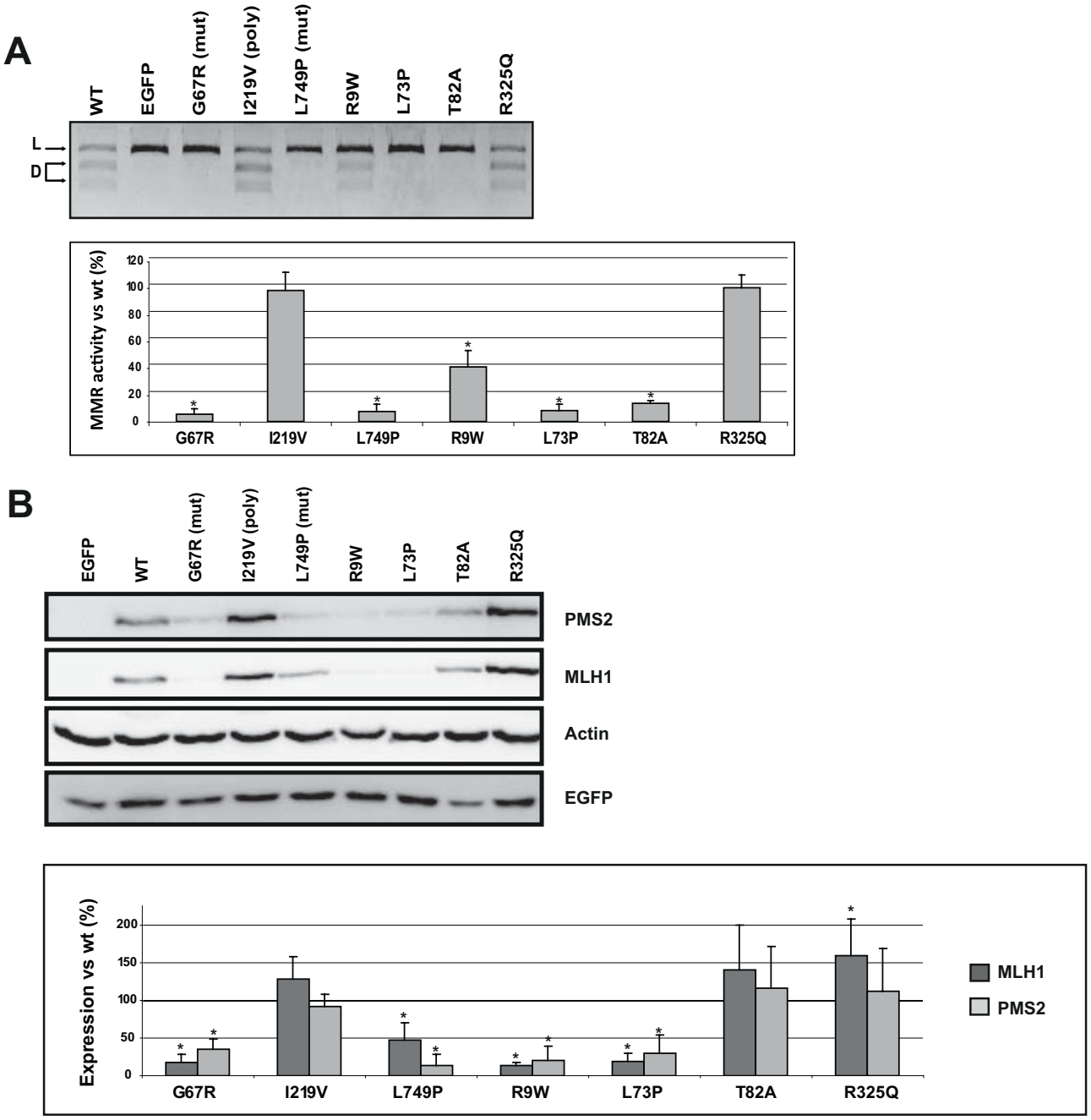


Figure 3. Mismatch repair activity and protein expression levels of MLH1 variants. **A:** Agarose gel showing digestion products of MMR assay. Digestion of the unrepaired plasmid yielded linearized vector (L), whereas successful repair produced two additional fragments of lower molecular weight (D, digestion fragments). Quantification of repair levels of MLH1 variants in direct comparison to MLH1 wild type is shown. Statistically significant reductions of expression are marked by asterisks (*, $P < 0.005$). **B:** Western-blot analysis of MLH1 and PMS2. Quantification of MLH1 or PMS2 is shown in dark and light gray, respectively. Statistically significant variations of expression are marked by asterisks (*, $P < 0.01$).

MLH1 VUS analyzed showed exclusive nuclear location of MLH1 and PMS2 (Fig. 4).

Taken together, functional assays at protein level demonstrated that p.L73P and p.T82A are pathogenic mutations significantly impairing MMR activity. In contrast, we classified the c.974G>A (p.R325Q) variant as likely neutral because it had no effect on protein function. Finally, as the c.25C>T (p.R9W) variant showed intermediate values in MMR activity, its classification remains unclear.

Discussion

In this study, we have performed a comprehensive in-depth analysis of eight *MLH1* VUS, either not previously reported or with no functional information available, allowing the classification of seven of them (Table 5).

RNA analyses allowed classifying three *MLH1* variants as pathogenic mutations affecting mRNA processing. The concomitant association of the variants c.[248G>T(;)306G>C] produced

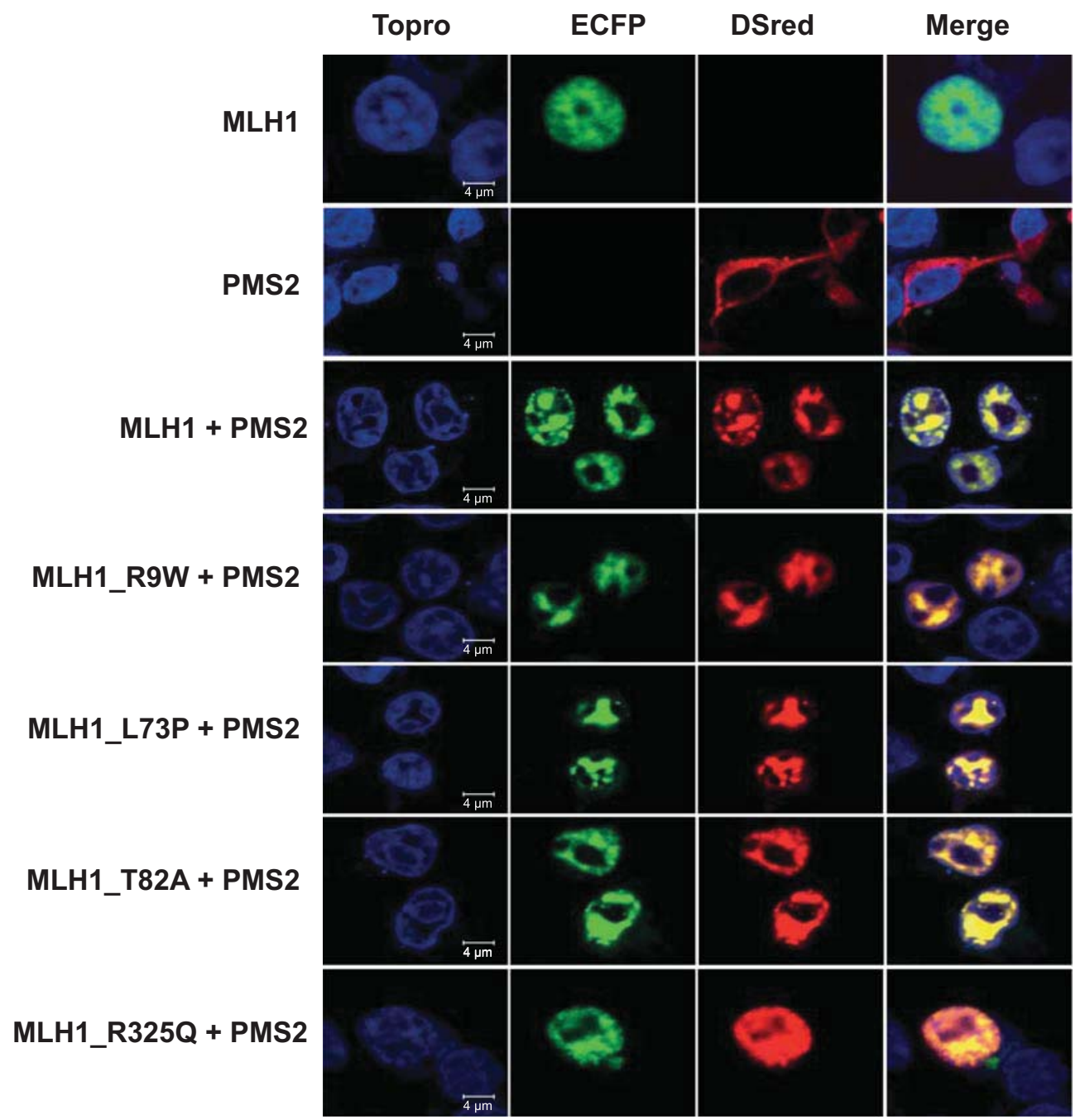


Figure 4. Subcellular localization of MLH1 variants. *MLH1* variants were transfected in HEK293T cells and detected using confocal laser microscopy. Nuclei were counterstained with TO-PRO-3 (left column) after fixation and resulting overlay is shown in the right column; ECFP: depicts subcellular location of wild type and variant MLH1 protein; DSRed: depicts subcellular location of wild-type and variant PMS2 protein.

a pathogenic in-frame deletion affecting the ATPase domain associated with loss of MLH1 protein expression in the tumor. In silico prediction suggests that variant c.306G>C could be the culprit. Noteworthy, the variant c.306G>T (p.E102D), located in the same nucleotide, did not associate with defective protein expression or MMR activity [Takahashi et al., 2007]. Unfortunately, no RNA analysis for the c.306G>T mutation has been reported.

The other two variants with an impact at the RNA level induced specific out-of-frame deleted transcripts leading to truncated proteins. On the one hand, the c.[780C>G;788A>C] variant led to exon 9 skipping. This is an example of variant located out of the consensus SS that affects RNA processing, as reported by other authors [Arnold et al., 2009; Ars et al., 2000; Auclair et al., 2006; Cartegni et al., 2002; Lastella et al., 2006; Naruse et al., 2009; Tournier et al., 2008]. The novel c.791-7T>A variant resulted in exon 10 skipping. Likewise, the neighboring c.791-5T>G variant also lead to the same RNA alteration [Auclair et al., 2006; Tournier et al., 2008]. In these two cases, tumors showed loss of MLH1 protein expression. In the family of the latter case, the mutation cosegregated with the disease.

Functional assays at protein level demonstrated that p.L73P is a pathogenic mutation significantly impairing MMR activity. Assays performed in yeast had previously revealed an intermediate

Table 5. Summary of Results

MLH1 VUS	Predicted Protein	Selection criteria			Clinico-pathological characteristics		Control Frequency (%)	mRNA analysis		Protein analysis		Final Classification
		Family ID	Criteria met	MSI	MLH1 expression (IHC)	Cosegregation		mRNA in silico prediction	cDNA splicing and stability analysis	Protein in silico prediction	Protein analysis	
c.25C>T	p.R9W	A	BC	+	Conserved	NE	0	No effect	No effect	Inconclusive	Inconclusive	Inconclusive
c.218T>C	p.L73P	B	BC	+	Not evaluable	NE	0	Inconclusive ESE	No effect	**Impaired**	**Impaired**	**Pathogenic**
c.244G>A	p.T82A	C	BC	+	Consverved	NE	0	Inconclusive ESE	No effect	**Impaired**	**Impaired**	**Pathogenic**
c.[248G>T(;)306G>C]	p.[S83I(;)E102D]	D	BC	+	**Lost**	NE	0	**Inconclusive SS** and ESE	**Aberrant transcript**	**Impaired; Impaired**	NP	**Pathogenic**
c.702G>A	p.(=)	E	AC	+	**Lost**	NE	0	**Aberrant ESE**	No effect	NP	NP	Likely Neutral
c.[780C>G;788A>C]	p.N263T					NE	0	**Inconclusive SS** and ESE	**Aberrant transcript**	**Impaired**	NP	**Pathogenic**
c.791-7T>A		F	AC	+	**Lost**	**Yes**	0	**Aberrant SS**	**Aberrant transcript**	NP	NP	**Pathogenic**
c.974G>A	p.R325Q	G and H	BC and AC	+/-	Conserved	**Yes**	0	Inconclusive ESE	No effect	Neutral	Normal	Likely Neutral

Summary of the clinical characteristics, control frequency, and mRNA and protein analyses of the identified *MLH1* VUS. Information concerning these variants is detailed in Tables 1–4.

(34–66%) decrease in MMR activity [Ellison et al., 2004]. Other variants in the same residue have been reported and, for all of them, yeast assays yielded inconclusive results [Ellison et al., 2004]. The observed functional defect is in agreement with structural and conservation analyses. Interestingly, other alterations also affecting the integrity of the structural backbone of the ATPase pocket (p.G67R and p.T117M, Fig. 2) had similar effects by destabilizing and inactivating the protein [Blasi et al., 2006; Drost et al., 2010; Ellison et al., 2004; Kosinski et al.,; Raevaara et al., 2005; Shcherbakova and Kunkel, 1999; Shimodaira et al., 1998; Takahashi et al., 2007; Trojan et al., 2002; Vogelsang et al., 2009].

The variant c.244A>G (p.T82A) destroys two vital hydrogen bonds within the ATPase pocket and, therefore, disables ATPase activity, accounting for the repair deficiency of the mutant. p.T82I variant, located in the same position, has been shown to be MMR defective in yeast and human cells while showing normal expression levels [Takahashi et al., 2007]. Also, p.T82K and p.T82M variants affect MMR activity in yeast [Ellison et al., 2004]. The p.T82A carrier had a concomitant *MSH6* p.L396V variant that did not have an impact at RNA level (data not shown) and did not show a defective mismatch repair activity [Drost et al., 2012; Kolodner et al., 1999]. The p.T82A variant is deemed as pathogenic in spite of the conserved MLH1/PMS2 protein expression in tumor. This observation is in agreement with the normal expression levels observed in transfected cells.

Two variants, c.702G>A and c.974G>A (p.R325Q) were classified as likely neutral. The silent c.702G>A did not affect mRNA processing or stability. The coexistence of variants c.[702G>A];[780C>G;788A>C] in the same individual belonging to a family with typical Lynch syndrome characteristics further supports the classification of the silent c.702G>A variant as likely neutral based on the results obtained in mRNA analyses. The c.974G>A (p.R325Q) variant was previously reported as a somatic mutation in two sporadic MSI tumors [Kuismanen et al., 2000; Wu et al., 1997], but not as a germline mutation. In agreement with its proficient function, this variant was detected in two families where two of the three affected carriers analyzed harbored MSS tumors. It should be noted that the expression levels for this variant were above wild type in the functional assays, as previously reported for other variants such as p.E663G, p.R755S, and p.E102D [Kosinski et al., 2010; Takahashi et al., 2007].

Finally, variant c.25C>T (p.Arg9Trp) could not be classified based on the presence of some repair activity in spite of low protein expression levels. Of note, a concomitant *APC* mutation, considered as pathogenic was detected in the index case, where multiple adenomas coexisted with an MSI CRC [Aretz et al., 2004; Castellsague et al., 2010].

In the series reported here only Lynch syndrome suspected patients, showing MSI tumors were considered. All these patients have attended Genetic Counseling Units where collection of pathological information has been comprehensive, in combination with *BRAF* mutation or methylation analysis in MLH1-negative cases. While protein immunostaining is often informative discrepancies may be observed between expression patterns and putative functional impact. As shown, pedigree sizes are often small and clinical information for segregation analysis is scarce. In this setting, the analysis of variant frequency in healthy controls is performed based on the assumption that a variant present in a frequency higher than 1% is unlikely pathogenic.

Analysis at the RNA level offered a good performance allowing the classification of four of eight (50%) variants. This is a high yield that further reinforces the notion that RNA analysis must be an early step when assessing the functional impact of variants.

When fresh lymphocytes cannot be obtained, a number of in vitro methodologies have been developed to assess the functional impact of variants on RNA [Betz et al., 2010; Lastella et al., 2006; Naruse et al., 2009; Tournier et al., 2008]. Single amino acid substitutions at *MLH1* and other genes have been associated with allelic imbalances [Castellsague et al., 2010; Perera et al., 2010; Santibanez Koref et al., 2010].

The putative clinical usefulness of in silico predictions is a matter of controversy. In our study, the comparison of splicing prediction and experimental observations indicates that only Spliceport, out of the four programs used, offers a perfect match between prediction and observation. Nonetheless and for the sake of robustness, we support the use of several software packages to make a better prediction [Arnold et al., 2009; Betz et al., 2010]. In contrast, ESE prediction programs did not offer concordant results and their prediction did not always correlate with experimental evidence [Arnold et al., 2009; Auclair et al., 2006; Chenevix-Trench et al., 2006; Lastella et al., 2006; Tournier et al., 2008]. Assessing the putative impact of in silico prediction of single amino acid substitutions on RNA splicing should be always performed, as it can help in prioritizing RNA or protein analysis.

At the protein level, a strong correlation between experimental findings and in silico prediction packages has been observed in this set of cases. When all programs offered concordant predictions the functional assays went in the same direction. These observations support the weight that Kaniskas and collaborators give to in silico prediction programs [Kansikas et al., 2011].

Based on the evidence linking MMR defects to clinical Lynch syndrome, assays showing loss of MMR activity in vitro could provide strong evidence supporting the pathogenicity of an MMR variant. Several forms of MMR assays have been developed, based on the production of a variant protein or the introduction of a variant-encoding gene in cells with an endogenous deficiency of the gene, followed by complementation of a cell extract in an MMR assay [Drost et al., 2010; Kosinski et al., 2010; Ollila et al., 2006]. However, when variants show proficient MMR activity or assays yield inconclusive results further characterization using a selection of biochemical assays may be required. These assays focus on the analysis of protein expression, subcellular localization, and protein–protein interaction. We evaluated MLH1 and PMS2 expression in the same protein extracts of MLH1/PMS2 transfected cells used in MMR assays. Although in our study in vitro interaction assays has not been performed, the evaluation of PMS2 expression after MLH1/PMS2 transfection has been proposed as a reliable method for identifying dimerization defects [Kosinski et al., 2010; Mohd et al., 2006].

The coexistence of variants within the same gene or in two distinct cancer predisposition genes poses a clinical challenge when attempting to identify which is the real culprit. This can only be addressed by a comprehensive characterization of the identified changes. In our setting, a combination of *MLH1* variants was identified in two families. In two additional cases, mutations or variants in other genes (*APC* or *MSH6*) were later identified. At a time when next-generation sequencing may foster the detection of multiple VUS in distinct cancer genes, their functional characterization will become more relevant to the genetic counseling of these families.

The assessment of pathogenicity of VUS is always based on complex strategies that combine analyses both at RNA and protein levels and integrates other type of data. The use of Bayesian models can be necessary for the integration of direct and indirect evidence of pathogenicity in a single model to obtain a probability of pathogenicity from the likelihood ratio [Goldgar et al., 2008; Pastrello et al., 2011; Plon et al., 2008]. Based on our observations and the decision tree proposed by [Couch et al., 2008] for the in vitro analysis of MMR VUS found in putative LS families, we propose a preliminary algorithm for the evaluation of *MLH1* VUS (Supp. Fig. S2, Supp. Tables S4 and S5).

Attempts are ongoing aiming to reach consensus on the interpretation of the relative weight of all clinical and experimental evidence on MMR variants. In the meanwhile, it is necessary both to increase the knowledge on specific variants found, compare the evidence obtained at different levels (RNA splicing, protein function and stability), and devise strategies for the incorporation of this knowledge in the clinical setting. Validation of functional tests used will be mandatory prior to a generalized use in the clinical setting.

Acknowledgments

EB is a recipient of an IDIBELL fellowship and of travel grant by RTICC. Authors have no conflict of interest to declare.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7:248–249.

Aretz S, Uhlhaas S, Sun Y, Pagenstecher C, Mangold E, Caspari R, Moslein G, Schulmann K, Propping P, Friedl W. 2004. Familial adenomatous polyposis: aberrant splicing due to missense or silent mutations in the APC gene. Hum Mutat 24:370–380.

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian SV, Golgar DE, Young JP, Spurdle AB. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30:757–770.

Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X. 2000. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Hum Mol Genet 9:237–247.

Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S, et al. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. Hum Mutat 27:145–154.

Ban C, Junop M, Yang W. 1999. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell 97:85–97.

Barnetson RA, Cartwright N, van Vliet A, Haq N, Drew K, Farrington S, Williams N, Warner J, Campbell H, Porteous ME, et al. 2008. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. Hum Mutat 29:367–374.

Betz B, Theiss S, Aktas M, Konermann C, Goecke TO, Moslein G, Schaal H, Royer-Pokora B. 2010. Comparative in silico analyses and experimental validation of novel splice site and missense mutations in the genes MLH1 and MSH2. J Cancer Res Clin Oncol 136:123–134.

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

Boland CR, Koi M, Chang DK, Carethers JM. 2008. The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch syndrome: from bench to bedside. Fam Cancer 7:41–52.

Borras E, Pineda M, Blanco I, Jewett EM, Wang F, Teule A, Caldes T, Urioste M, Martinez-Bouzas C, Brunet J, et al. 2010. MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. Cancer Res 70:7379–7391.

Brieger A, Adam R, Passmann S, Plotz G, Zeuzem S, Trojan J. 2011. A CRM1-dependent nuclear export pathway is involved in the regulation of MutLalpha subcellular localization. Genes Chromosomes Cancer 50:59–70.

Brieger A, Plotz G, Hinrichsen I, Passmann S, Adam R, Zeuzem S. 2012. C-terminal fluorescent labeling impairs functionality of DNA mismatch repair proteins. PLoS One 7:e31863.

Burset M, Seledtsov IA, Solovyev VV. 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 29:255–259.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31:3568–3571.

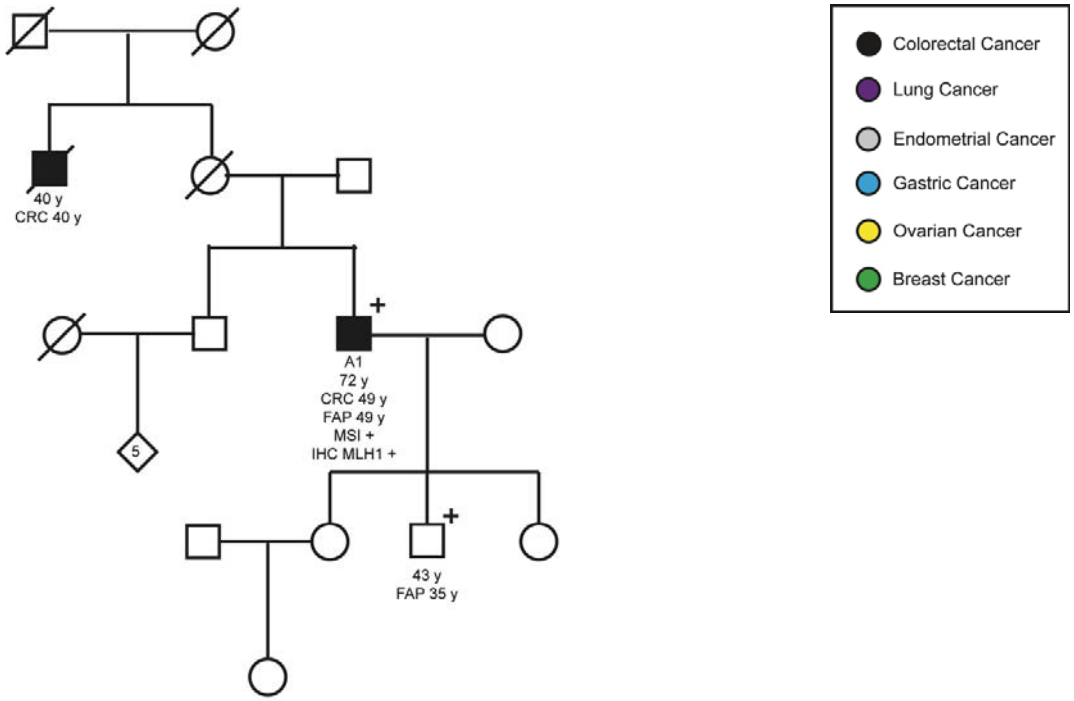
Castellsague E, Gonzalez S, Guino E, Stevens KN, Borras E, Raymond VM, Lazaro C, Blanco I, Gruber SB, Capella G. 2010. Allele-specific expression of APC in adenomatous polyposis families. Gastroenterology 139:439–447.

Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, Ng P, Gruber SB, Watson P, Rennert G, Anton-Culver H, et al. 2008. Accurate classification of MLH1/MSH2

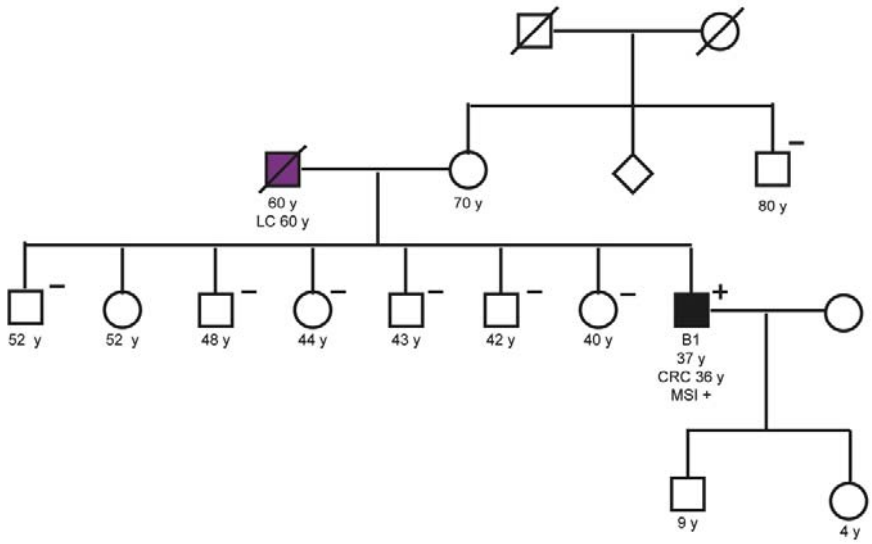
missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). Hum Mutat 29:852–860.
Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T, et al. 2006. Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. Cancer Res 66:2019–2027.
Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de Wind N. 2008. Assessment of functional effects of unclassified genetic variants. Hum Mutat 29:1314–1326.
Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190.
Dogan RI, Getoor L, Wilbur WJ, Mount SM. 2007. SplicePort–an interactive splice-site analysis tool. Nucleic Acids Res 35(Web Server issue):W285–W291.
Drost M, Zonneveld JB, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind N. 2010. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum Mutat 31:247–253.
Drost M, Zonneveld JB, van Hees S, Rasmussen LJ, Hofstra RM, de Wind N. 2012. A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. Hum Mutat 33:488–494.
Ellison AR, Lofing J, Bitter GA. 2004. Human MutL homolog (MLH1) function in DNA mismatch repair: a prospective screen for missense mutations in the ATPase domain. Nucleic Acids Res 32:5321–5338.
Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. Science 297:1007–1013.
Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS. 2008. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat 29:1265–1272.
Guarne A, Ramon-Maiques S, Wolff EM, Ghirlando R, Hu X, Miller JH, Yang W. 2004. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. Embo J 23:4134–4145.
Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. 1996. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24:3439–3452.
Kansikas M, Kariola R, Nystrom M. 2011. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. Hum Mutat 32:107–115.
Kolodner RD, Tytell JD, Schmeits JL, Kane MF, Gupta RD, Weger J, Wahlberg S, Fox EA, Peel D, Ziogas A, et al. 1999. Germ-line msh6 mutations in colorectal cancer families. Cancer Res 59:5068–5074.
Kosinski J, Hinrichsen I, Bujnicki JM, Friedhoff P, Plotz G. 2010. Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. Hum Mutat 31:975–982.
Kuismanen SA, Holmberg MT, Salovaara R, de la Chapelle A, Peltomaki P. 2000. Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. Am J Pathol 156:1773–1779.
Lastella P, Surdo NC, Resta N, Guanti G, Stella A. 2006. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. BMC Genomics 7:243.
Lynch HT, Lynch JF, Lynch PM, Attard T. 2008. Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. Fam Cancer 7:27–39.
Mohd AB, Palama B, Nelson SE, Tomer G, Nguyen M, Huo X, Buermeyer AB. 2006. Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. DNA Repair (Amst) 5:347–361.
Moreno V, Gemignani F, Landi S, Gioia-Patricola L, Chabrier A, Blanco I, Gonzalez S, Guino E, Capella G, Canzian F. 2006. Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. Clin Cancer Res 12:2101–2108.
Naruse H, Ikawa N, Yamaguchi K, Nakamura Y, Arai M, Ishioka C, Sugano K, Tamura K, Tomita N, Matsubara N, et al. 2009. Determination of splice-site mutations in Lynch syndrome (hereditary non-polyposis colorectal cancer) patients using functional splicing assay. Fam Cancer 8:509–517.
Ng PC, Henikoff S. 2002. Accounting for human polymorphisms predicted to affect protein function. Genome Res 12:436–446.
Ollila S, Sarantaus L, Kariola R, Chan P, Hampel H, Holinski-Feder E, Macrae F, Kohonen-Corish M, Gerdes AM, Peltomaki P, et al. 2006. Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. Gastroenterology 131:1408–1417.
Ou J, Niessen RC, Lutzen A, Sijmons RH, Kleibeuker JH, de Wind N, Rasmussen LJ, Hofstra RM. 2007. Functional analysis helps to clarify the clinical importance of unclassified variants in DNA mismatch repair genes. Hum Mutat 28:1047–1054.
Pagenstecher C, Wehner M, Friedl W, Rahner N, Aretz S, Friedrichs N, Sengteller M, Henn W, Buettner R, Propping P, et al. 2006. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. Hum Genet 119:9–22.
Pastrello C, Pin E, Marroni F, Bedin C, Fornasarig M, Tibiletti MG, Oliani C, Ponz de Leon M, Urso ED, Della Puppa L, et al. 2011. Integrated analysis of unclassified variants in mismatch repair genes. Genet Med 13:115–124.
Peltomaki P, Vasen H. 2004. Mutations associated with HNPCC predisposition–Update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 20:269–276.
Perera S, Li B, Tsitsikotas S, Ramyar L, Pollett A, Semotiuk K, Bapat B. 2010. A novel and rapid method of determining the effect of unclassified MLH1 genetic variants on differential allelic expression. J Mol Diagn 12:757–764.
Plon SE, Cooper HP, Parks B, Dhar SU, Kelly PA, Weinberg AD, Staggs S, Wang T, Hilsenbeck S. 2008. Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. Genet Med 13:148–154.
Plotz G, Raedle J, Spina A, Welsch C, Stallmach A, Zeuzem S, Schmidt C. 2008. Evaluation of the MLH1 I219V alteration in DNA mismatch repair activity and ulcerative colitis. Inflamm Bowel Dis 14:605–611.
Plotz G, Welsch C, Giron-Monzon L, Friedhoff P, Albrecht M, Piiper A, Biondi RM, Lengauer T, Zeuzem S, Raedle J. 2006. Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic Acids Res 34:6574–6586.
Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lonnqvist KE, Holinski-Feder E, Sutter C, McKinnon W, Duraisamy S, et al. 2005. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology 129:537–549.
Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4:311–323.
Renkonen ET, Nieminen P, Abdel-Rahman WM, Moisio AL, Jarvela I, Arte S, Jarvinen HJ, Peltomaki P. 2005. Adenomatous polyposis families that screen APC mutation-negative by conventional methods are genetically heterogeneous. J Clin Oncol 23:5651–5659.
Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, Aittomaki K, Jarvinen HJ, Mecklin JP, Lindblom A, et al. 2003. Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. J Clin Oncol 21:3629–3637.
Santibanez Koref M, Wilson V, Cartwright N, Cunnington MS, Mathers JC, Bishop DT, Curtis A, Dunlop MG, Burn J. 2010. MLH1 Differential allelic expression in mutation carriers and controls. Ann Hum Genet 74:479–488.
Shcherbakova PV, Kunkel TA. 1999. Mutator phenotypes conferred by MLH1 overexpression and by heterozygosity for mlh1 mutations. Mol Cell Biol 19:3177–3183.
Shimodaira H, Filosi N, Shibata H, Suzuki T, Radice P, Kanamaru R, Friend SH, Kolodner RD, Ishioka C. 1998. Functional analysis of human MLH1 mutations in Saccharomyces cerevisiae. Nat Genet 19:384–389.
Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM. 2008. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 29:1304–1313.
Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. 2007. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. Cancer Res 67:4595–4604.
Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine MP, Soret J, Tazi J, et al. 2008. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat 29:1412–1424.
Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, Plotz G, Jiricny J, Marra G. 2002. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology 122:211–219.
Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, et al. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 96:261–268.
Vasen HF. 2007. Review article: The Lynch syndrome (hereditary nonpolyposis colorectal cancer). Aliment Pharmacol Ther 26(Suppl 2):113–126.
Vogelsang M, Comino A, Zupanec N, Hudler P, Komel R. 2009. Assessing pathogenicity of MLH1 variants by co-expression of human MLH1 and PMS2 genes in yeast. BMC Cancer 9:382.
Watson P, Riley B. 2005. The tumor spectrum in the Lynch syndrome. Fam Cancer 4:245–248.
Wu Y, Nystrom-Lahti M, Osinga J, Looman MW, Peltomaki P, Aaltonen LA, de la Chapelle A, Hofstra RM, Buys CH. 1997. MSH2 and MLH1 mutations in sporadic replication error-positive colorectal carcinoma as assessed by two-dimensional DNA electrophoresis. Genes Chromosomes Cancer 18:269–278.
Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev 18:1241–1250.
Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA. 2005. Exon inclusion is dependent on predictable exonic splicing enhancers. Mol Cell Biol 25:7323–7332.

Supp. Figure S1. Pedigrees of families harboring MLH1 *VUS*. Age at diagnosis, molecular characteristics of tumors (MSI, microsatellite instability: +, instable; -, stable; IHC MLH1, immunohistochemical analysis of MLH1 protein in tumour tissue: +, expression; -, loss of expression), and mutation status are shown. Carriers are depicted by "+" and non-carriers by "–" symbols. CRC, colorectal cancer; BC, breast cancer; EC, endometrial cancer; GC, gastric cancer; LC, lung cancer; OC, ovarian cancer.

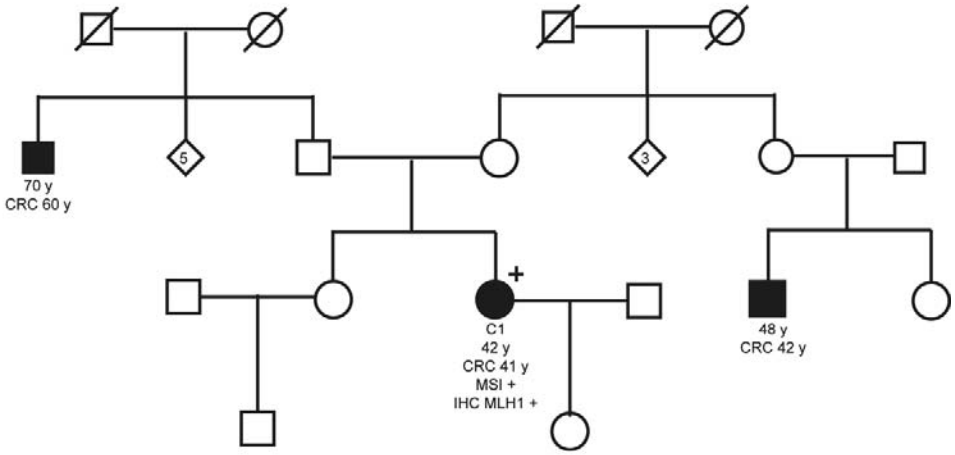
A



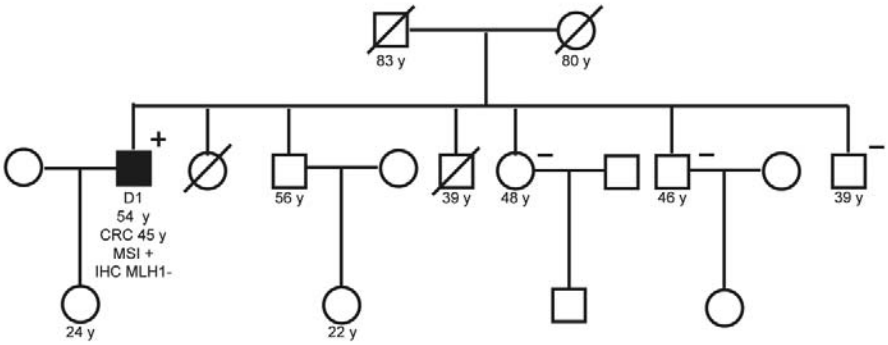
B



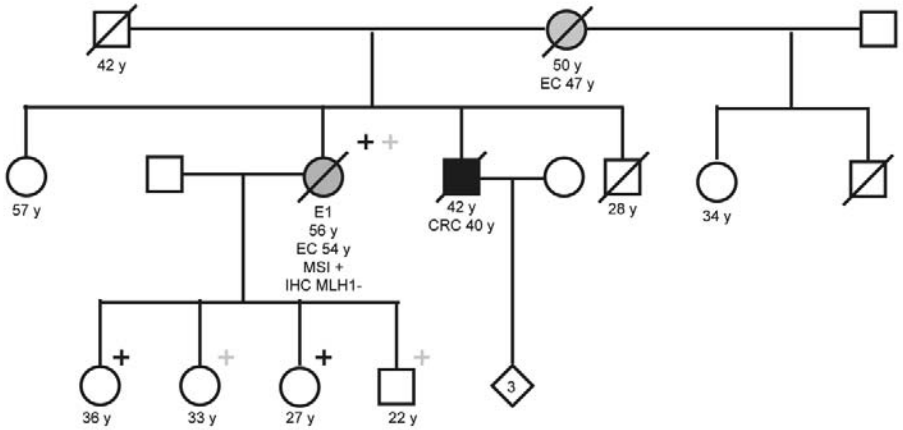
C


70 y
CRC 60 y
5
3
+
C1
42 y
CRC 41 y
MSI +
IHC MLH1 +
48 y
CRC 42 y

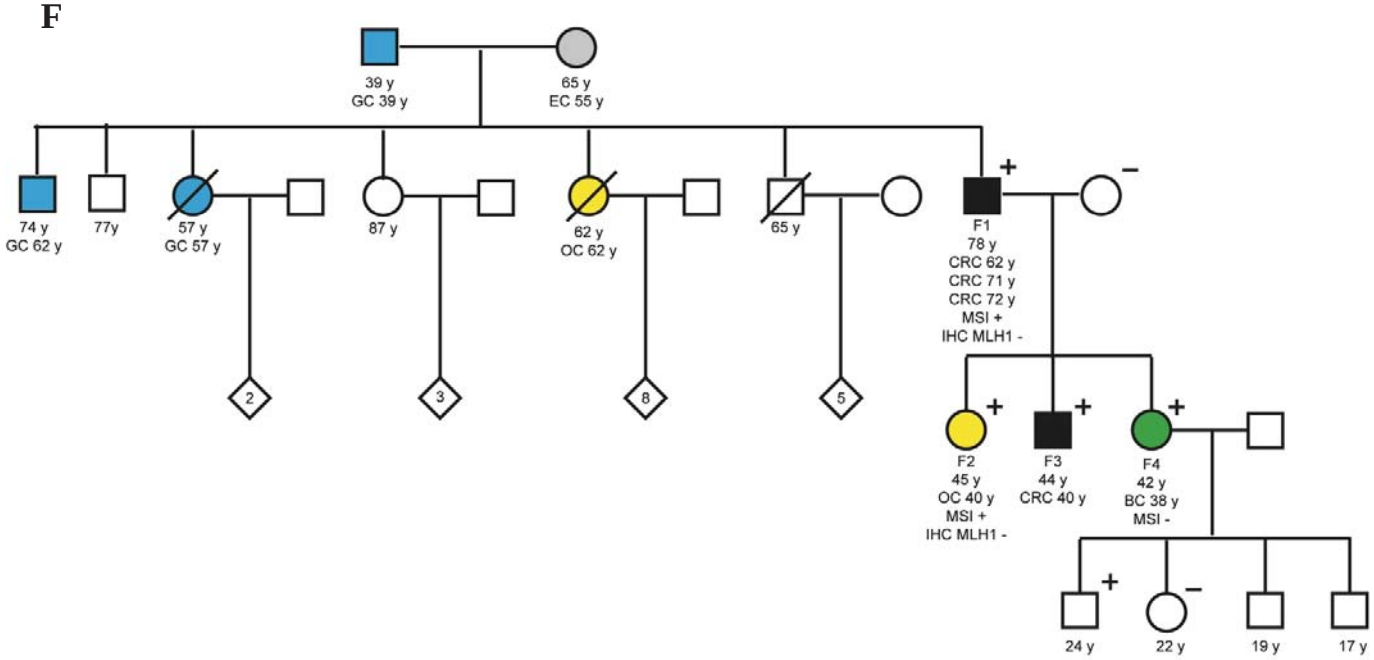
D


83 y
80 y
+
D1
54 y
CRC 45 y
MSI +
IHC MLH1-
56 y
39 y
48 y
46 y
39 y
24 y
22 y

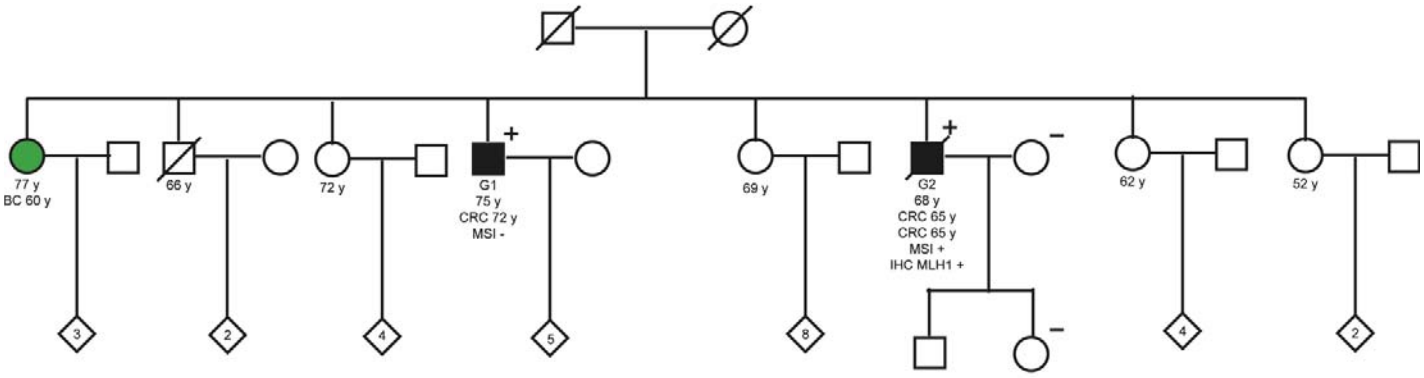
E


42 y
50 y
EC 47 y
57 y
E1
56 y
EC 54 y
MSI +
IHC MLH1-
42 y
CRC 40 y
28 y
34 y
36 y
33 y
27 y
22 y
3

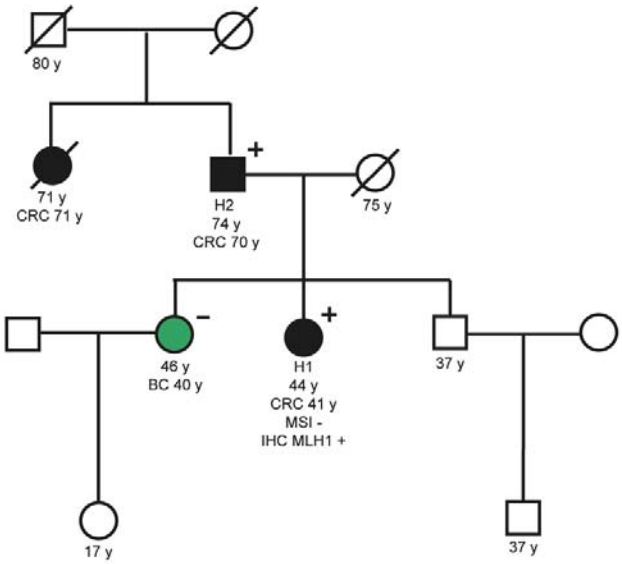
F


39 y
GC 39 y
65 y
EC 55 y
74 y
GC 62 y
77y
57 y
GC 57 y
87 y
62 y
OC 62 y
65 y
F1
78 y
CRC 62 y
CRC 71 y
CRC 72 y
MSI +
IHC MLH1 -
2
3
8
5
F2
45 y
OC 40 y
MSI +
IHC MLH1 -
F3
44 y
CRC 40 y
F4
42 y
BC 38 y
MSI -
24 y
22 y
19 y
17 y

G

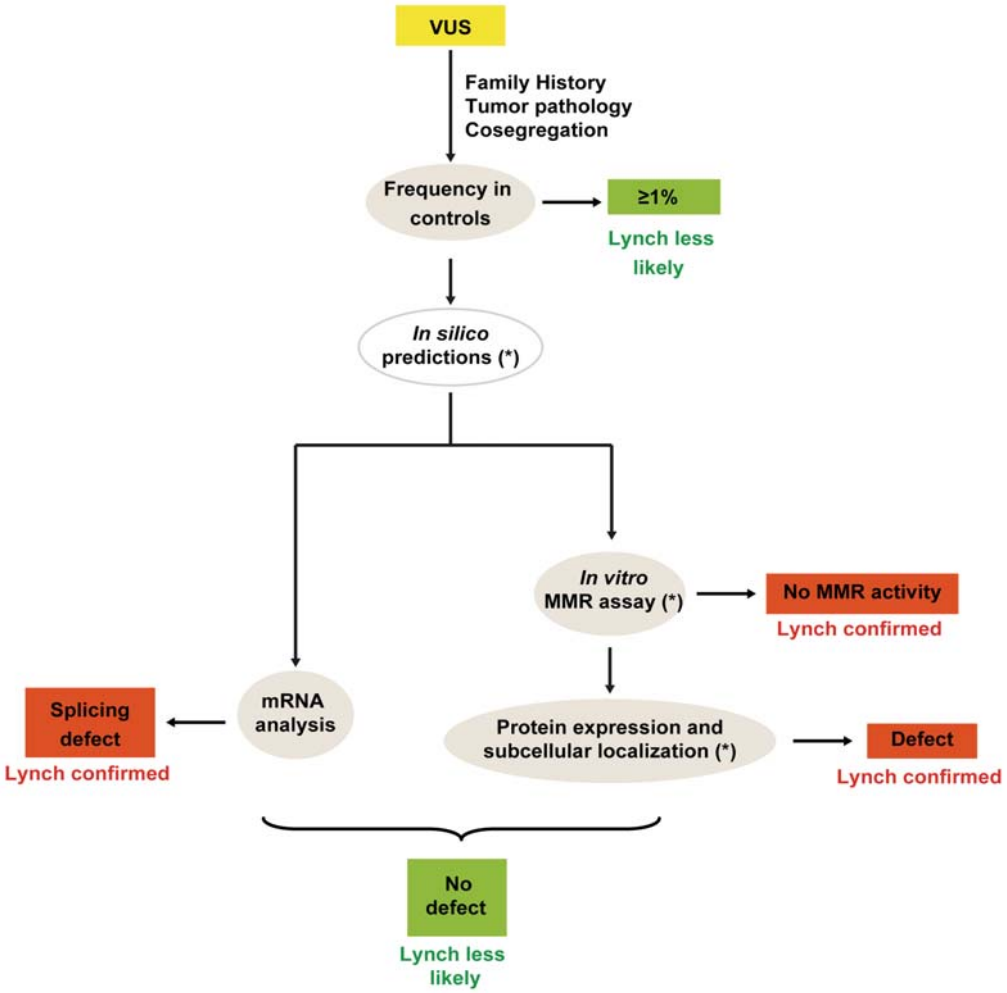

77 y
BC 60 y
66 y
72 y
G1
75 y
CRC 72 y
MSI -
69 y
G2
68 y
CRC 65 y
CRC 65 y
MSI +
IHC MLH1 +
62 y
52 y
3
2
4
5
8
4
2

H


80 y
71 y
CRC 71 y
+
H2
74 y
CRC 70 y
75 y
–
46 y
BC 40 y
+
H1
44 y
CRC 41 y
MSI -
IHC MLH1 +
37 y
17 y
37 y

Supp. Figure S2. A proposed algorithm for *MLH1* VUS analysis. Based on our observations and the decision tree proposed by Couch et al., (2008) for the *in vitro* analysis of MMR VUS found in putative LS families, we propose an algorithm for the evaluation of *MLH1* VUS. We posit that this strategy provides a useful framework that (i) emphasizes the importance of clinical information (family history, tumor pathology and control frequency data); (ii) suggest that *in silico* predictions can be of help in prioritizing RNA or protein assays in a mutually exclusive manner; and (iii) provides a useful framework for including all information when assessing a VUS.

The putative usefulness of the proposed approach was validated in a set of 25 *MLH1* variants which had been previously studied both at protein and RNA levels (Supp. Tables S4 and S5). Had we applied the algorithm after the initial assessment of clinico-pathological data, 20 VSD of validation set data would have been classified either as pathogenic (n=13) or likely neutral (n=7). For these variants RNA *in silico* predictions [aberrant SS (complete agreement for the 4 programs utilized) or inconclusive SS (any of 4 programs predicted aberrations)] were highly predictive of cDNA aberrations. No predictive power was observed for ESE *in silico* predictions. Regarding prediction of protein defects, correlation with experimental observations was good but not perfect (7 of 10 predicted aberrations were experimentally confirmed). Thus, *in silico* predictions would have been useful in directing to RNA or protein *in vitro* analysis.



Supp. Table S1. Additional variants identified in exons 1, 2, 9, 10 and 11 in *MLH1 g*ene by CSCE in control and CRC cases

CSCE Fragment	Change identified	Allelic frequency in controls (%)	Allelic frequency in sporadic cases (%)	Previously Reported
exon 1	**c.116+103A>G**	0/604 (0.00%)	1/648 (0.15%)	
	c.116+69G>A	1/604 (0.16%)	0/648 (0.00%)	
	c.1-93G>A	100/604 (16.56 %)	112/648 (17.28%)	rs: 1800734
exon 2	**c.117-109A>G**	0/608 (0.00%)	1/648 (0.15%)	
	c.195C>T (p.=)	0/608 (0.00%)	1/648 (0.15%)	
	c.198C>T (p.=)	1/608 (0.16%)	1/648 (0.00%)	Reported at LOVD
	c.207+41A>G	0/608 (0.00%)	1/648 (0.15%)	
exon 9	**c.790+68A>T**	0/604 (0.00%)	1/648 (0.15%)	
exon 10	**c.791-63G>A**	1/604 (0.16%)	0/648 (0.00%)	
	c.791-79A>G	9/604 (1.49%)	5/648 (0.77%)	
	c.884+39G>A	1/604 (0.16%)	0/648 (0.00%)	
	c.885-81G>C	11/604 (1.8%)	21/648 (3.24%)	
exon 11	**c.954C>T (p.=)**	0/604 (0.00%)	1/648 (0.15%)	
	c.977T>C (p.V326A)	1/604 (0.15%)	1/604 (0.15%)	Reported at LOVD
	c.1038+51C>T	1/604 (0.15%)	0/648 (0.00%)	Reported at LOVD

Supp. Table S2. Primers used in this study

Primer Name	Sequence
CSCE_MLH1_E1_fam_F	5' FAM-TGACTGGCATTCAAGCTGTC 3'
CSCE_MLH1_E1_R	5' TCGGGGGAGAGCGGTAAA 3'
CSCE_MLH1_E2_hex_F	5' HEX- CCCGTCTCTTCCCTCTCTCT 3'
CSCE_MLH1_E2_R	5' ATCCCATCTGCAAAAGCCTA 3'
CSCE_MLH1_E3_fam_F	5' FAM-ATGGGAATTCAAAGAGATTTGG 3'
CSCE_MLH1_E3_R	5' GTTTGCTCAGATTTGCATAC 3'
CSCE_MLH1_E10_fam_F	5' FAM-CTTTCTTCCTGGGGATGTGA 3'
CSCE_MLH1_E10_R	5' GCCAGTGGTGTATGGGATTC 3'
CSCE_MLH1_E11_hex_F	5' HEX-TCGATCCTGAGGTTTTGACC 3'
CSCE_MLH1_E11_R	5' GCAAAAATCTGGGCTCTCAC 3'
CSCE_MLH1_E16_hex_F	5' HEX-GGATGCTCCGTTAAAGCT 3'
CSCE_MLH1_E16_R	5' GAAGTATAAGAATGGCTGTCAC 3'
CSCE_MLH1_E9_fam_F	5' FAM-TGGGAAGGAACCTTGTGTTT 3'
CSCE_MLH1_E9_R	5' GCCATGAGGTTTCTCCATGT 3'
ASE_MLH1_218	5' AGTGAACCTTTCACATACAATATCC 3'
ASE_MLH1_244	5' CTGGATATTGTATGTGAAAGGTTCACT 3'
ASE_MLH1_702	5' GAGAACTGATAGAAATTGGATGTGA 3'
ASE_MLH1_974	5' CGAGGAGAGCATCCTGGAGC 3'
RNA_MLH1_5UTR_F	5' CAGCAACCCACAGAGTTGAG 3'
RNA_MLH1_E1_F	5' TATCCAGCGGCCAGCTAA 3'
RNA_MLH1_E2_R	5' TCAGGCCTCCCTCTTTAACA 3'
RNA_MLH1_E5_R	5' AGGGGCTTTCAGTTTTCCAT 3'
RNA_MLH1_E8_F	5' CACTACCCAATGCCTCAACC 3'
RNA_MLH1_E9_F	5' AGAACTGATAGAAATTGGATGTGAGG 3'
RNA_MLH1_E10_R	5' AGGTACAGGAATGGGTGTGTG 3'
RNA_MLH1_E11_R	5' CCACATTCTGGGGACTGATT 3'
RNA_MLH1_E12_R	5'GTATCCCCCTCCAAGCTCTG 3'

Supp. Table S3. Summary of Allelic Ratios

MLH1 VUS	Predicted Protein	Puromycin Treatment	Specific design	Average ASE	SD
c.218T>C	**p.L73P**	-	E1_E5	0.82	0.05
		+	E1_E5	0.98	0.05
c.244A>G	**p.T82A**	-	E1_E5	1.01	0.14
		+	E1_E5	1.02	0.04
c.702G>A	**p.=**	-	E8-E10	0.95	0.06
		+	E8-E10	0.88	0.16
c.974G>A	**p.R325Q**	-	E9_E12	0.90	0.02
		+	E9_E12	0.83	0.03

Supp. Table S4. Summary of the mRNA and protein analyses in a validation set of 25 *MLH1* variants

		Selection criteria		Clinico-pathological characteristics		Control Frequency	mRNA analysis		Protein analysis		
MLH1 VUS	Predicted Protein	Amsterdam criteria	MSI	MLH1 expression	Cosegregation		mRNA in silico prediction	cDNA splicing analysis	Protein in silico prediction	Protein analysis	Final Classification
c.544A>G	p.R182G	No	NA	NA	NA	NA	**Aberrant SS**	**Aberrant transcript**	**Impaired**	Inconclusive*	**Pathogenic**
c.793C>T	p.R265C	Yes#	NA	NA	NA	NA	**Inconclusive SS** and ESE	**Aberrant transcript**	**Impaired**	Inconclusive*	**Pathogenic**
c.794G>A	p.R265H	NA	NA	NA	NA	NA	**Inconclusive SS** and ESE	**Aberrant transcript**	**Impaired**	Normal*	**Pathogenic**
c.1989G>T	p.E663D	Yes	NA	NA	NA	NA	**Inconclusive SS**	**Aberrant transcript**	Neutral	Normal*	**Pathogenic**
c.304G>A	p.E102K	Yes	NA	NA	NA	NA	No effect	No effect	**Impaired**	**Impaired***	**Pathogenic**
c.1646T>C	p.L549P	Yes	NA	NA	NA	NA	No effect	No effect	**Impaired**	**Impaired***	**Pathogenic**
c.199G>A	p.G67R	Yes#	+	NA	NA	0	Inconclusive ESE	No effect	**Impaired**	**Impaired**	**Pathogenic**
c.229T>C	p.C77R	Yes	+	NA	NA	NA	Inconclusive ESE	No effect	**Impaired**	**Impaired**	**Pathogenic**
c.350C>T	p.T117M	Yes#	+	Lost	Yes	NA	Inconclusive ESE	No effect	**Impaired**	**Impaired***	**Pathogenic**
c.464T>G	p.L155R	Yes	+	Lost	NA	NA	Inconclusive ESE	No effect	**Impaired**	**Impaired**	**Pathogenic**
c.739T>C	p.S247P	Yes#	+	Lost	NA	0.01	Inconclusive ESE	No effect	**Impaired**	**Impaired**	**Pathogenic**
c.977T>C	p.V326A	Yes#	NA	NA	NA	0.01	Inconclusive ESE	No effect	Inconclusive	**Impaired***	**Pathogenic**
c.988_990del	p.I330del	Yes	+	Lost	NA	NA	Inconclusive ESE	Inconclusive	NA	**Impaired**	**Pathogenic**
c.1820T>A	p.L607H	Yes	NA	NA	NA	NA	No effect	No effect	**Impaired**	Normal*	Likely Neutral*
c.2041G>A	p.A681T	Yes#	+	NA	NA	0	No effect	No effect	Inconclusive	Normal*	Likely Neutral*
c.1421G>A	p.R474Q	No	NA	NA	NA	NA	**Aberrant ESE**	No effect	Neutral	Normal*	Likely Neutral*
c.1852_1853AA>GC	p.K618A	Yes#	+	Lost	Yes	NA	**Aberrant ESE**	No effect	Inconclusive	Normal	Likely Neutral
c.1963A>G	p.I655V	NA	NA	NA	NA	0.005	**Aberrant ESE**	No effect	Inconclusive	Normal*	Likely Neutral*
c.637G>A	p.V213M	No	+	NA	NA	NA	Inconclusive ESE	No effect	Neutral	Normal	Likely Neutral
c.1808C>G	p.P603R	No	NA	NA	NA	NA	Inconclusive ESE	No effect	**Impaired**	Normal*	Likely Neutral
c.2059C>T	p.R687W	Yes#	+	Lost	Yes	NA	**Aberrant ESE**	No effect	Inconclusive	Inconclusive*	Inconclusive*
c.1652A>C	p.N551T	No	NA	NA	NA	NA	No effect	No effect	**Impaired**	Inconclusive*	Inconclusive*
c.2146G>A	p.V716M	Yes#	NA	NA	NA	NA	Inconclusive ESE	No effect	Neutral	Inconclusive*	Inconclusive*
c.1984A>C	p.T622P	Yes	NA	NA	NA	NA	No effect	Inconclusive	Inconclusive	Inconclusive*	Inconclusive*
c.2066A>G	p.Q689R	No	NA	NA	NA	0.006	Inconclusive ESE	No effect	Neutral	Inconclusive*	Inconclusive*

Information concerning these variants is detailed in Supp. Table S5. Results in bold indicate major alterations. Gray highlighting indicate correlation between *in silico* predictions and *in vitro* results. Abbreviations: AC, Amsterdam criteria; BC, Bethesda criteria; MSI, microsatellite instability: +, instable; -, stable; IHC, immunohistochemical analysis; NE, not evaluable; NP, not performed. #Amsterdam criteria fulfilled in some families, *Subcellular localization analysis is not performed.

Supp. Table S5. Bioinformatic predictions and results of mRNA and protein analyses in a validation set of 25 *MLH1* variants. Results highlighted in grey indicate major alterations.

MLH1 VUS	Predicted Protein	SS	Splice Site Prediction								Enhancer site prediction				Summary of mRNA in silico predictions	cDNA splicing analysis	Classification at RNA level	Polyphen	Sift	Mapp-MMR	Interpretation of protein prediction	MMR assay	MLH1 and PMS2 expression	MLH1 and PMS2 Subcellular localization	Classification at Protein level
			NNSplice		Spliceport		NetGene2		SoftBerry		Interpretation of SS prediction	ESEfinder	RESCUE-ESE	PESX	Interpretation of ESE prediction											
			wild-type	variant	wild-type	variant	wild-type	variant	wild-type	variant																
c.199G>A	p.G67R	A	0.86	0.86	NR	NR	NR	NR	0.66	0.66	No effect	2 sites created	No change	No change	Inconclusive	Inconclusive ESE	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Impaired	Impaired	Impaired
		D	1	1	1.94	1.85	0.89	0.87	0.81	0.81																
c.229T>C	p.C77R	A	0.67	0.67	0.36	0.72	0.28	0.39	0.71	0.71	No effect	No change	No change	2 sites destroyed	Inconclusive	Inconclusive ESE	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Impaired	Impaired	Impaired
		D	1	1	0.85	1.02	0.86	0.89	0.79	0.79																
c.304G>A	p.E102K	A	0.67	0.67	0.36	0.36	NR	NR	0.71	0.71	No effect	No change	No change	No change	No effect	No effect	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Likely neutral	NP	Impaired*
		D	1	1	0.85	0.88	0.86	0.89	0.79	0.81																
c.350C>T	P.T117M	A	1	1	2.36	2.05	0.82	0.78	0.73	0.72	No effect	1 site reduced	2 sites destroyed	1 site destroyed	Inconclusive	Inconclusive ESE	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Impaired	NP	Impaired*
		D	0.88	0.88	0.73	0.44	0.63	0.56	0.79	0.79																
c.464T>G	p.L155R	A	0.8	0.83	0.74	1.12	0.33	0.49	NR	NR	No effect	No change	No change	1 site destroyed	Inconclusive	Inconclusive ESE	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Impaired	Impaired	Impaired
		D	0.84	0.84	NR	NR	0.44	0.49	NR	NR																
c.544A>G	p.R182G	A	0.8	0.8	0.74	0.74	0.33	0.33	NR	NR	Aberrant	No change	No change	No change	No effect	Inconclusive SS	Skipping exon 6	Aberrant transcript	PrD	Aff	Del	Impaired	Likely neutral	Inconclusive	NP	Inconclusive*
		D	0.66	NR	NR	NR	0.44	NR	NR	NR																
c.637G>A	p.V213M	A	0.68	0.68	0.22	0.2	0.39	0.34	NR	NR	No effect	2 sites destroyed	No change	No change	Inconclusive	Inconclusive ESE	No effect	No effect	Ben	Tol	Neu	Neutral	Likely neutral	Likely neutral	Likely neutral	Normal
		D	0.98	0.98	0.89	0.87	0.88	0.88	0.86	0.86																
c.739T>C	p.S247P	A	0.88	0.88	0.32	0.17	0.41	0.42	NR	NR	No effect	No change	1 site destroyed	No change	Inconclusive	Inconclusive ESE	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Impaired	Impaired	Impaired
		D	0.96	0.96	0.98	1.05	0.97	0.97	0.81	0.8																
c.793C>T	p.R265C	A	0.99	0.99	0.48	NR	0.33	0.26	0.71	0.7	Inconclusive	1 site increased	No change	No change	Inconclusive	Inconclusive SS and ESE	Skipping exon 10	Aberrant transcript	PrD	Aff	Del	Impaired	Likely neutral	Inconclusive	NP	Inconclusive*
		D	0.93	0.93	0.35	0.35	0.44	0.41	0.79	0.79																
c.794G>A	p.R265H	A	0.99	0.99	0.48	0.15	0.33	0.25	0.71	0.68	Inconclusive	2 sites destroyed	No change	No change	Inconclusive	Inconclusive SS and ESE	Skipping exon 10	Aberrant transcript	PrD	Aff	Del	Impaired	Likely neutral	Likely neutral	NP	Normal*
		D	0.93	0.93	0.35	0.35	0.44	0.41	0.79	0.79																
c.977T>C	p.V326A	A	0.94	0.94	1.14	1.14	0.67	0.56	0.71	0.71	No effect	1 site destroyed	No change	No change	Inconclusive	Inconclusive ESE	No effect	No effect	PsD	Aff	Del	Inconclusive	Impaired	Inconclusive	NP	Impaired*
		D	0.98	0.98	0.67	0.86	0	0	0.83	0.83																
c.988_990del	p.I330del	A	0.94	0.94	1.14	1.14	NR	NR	0.71	0.71	No effect	1 site created	No change	No change	Inconclusive	Aberrant splicing and inconclusive	Weak exon 11 skipipng	Inconclusive	NA	NA	NA	NA	Impaired	Impaired	Impaired	Impaired
		D	0.86	0.86	0.67	0.57	0.94	0.92	0.83	0.83																
c.1421G>A	p.R474Q	A	0.99	0.94	2.21	1.18	0.97	0.95	0.89	0.86	No effect	1 site created	6 sites created and 2 sites	1 site destroyed	Aberrant ESE	Aberrant ESE	No effect	No effect	Ben	Tol	NA	Neutral	Likely neutral	Likely neutral	NP	Normal*
		D	0.99	0.99	0.91	0.91	0.81	0.81	0.78	0.78																
c.1646T>C	p.L549P	A	0.98	0.98	2.03	2.03	0.68	0.68	0.77	0.77	No effect	No change	No change	No change	No effect	No effect	No effect	No effect	PrD	Aff	Del	Impaired	Impaired	Inconclusive	NP	Impaired*
		D	0.99	0.99	0.4	0.34	NR	NR	0.81	0.81																
c.1652A>C	p.N551T	A	0.98	0.98	2.03	2.03	0.68	0.68	0.77	0.77	No effect	No change	No change	No change	No effect	No effect	No effect	No effect	PrD	Aff	Del	Impaired	Likely neutral	Inconclusive	NP	Inconclusive*
		D	0.99	0.99	0.4	0.4	NR	NR	0.81	0.81																
c.1808C>G	p.P603R	A	0.99	0.99	0.88	1.16	0.97	0.97	0.79	0.79	No effect	1 site destroyed	No change	1 site created	Inconclusive	Inconclusive ESE	No effect	No effect	PrD	Aff	NA	Impaired	Likely neutral	Likely neutral	NP	Normal*
		D	0.9	0.9	NR	NR	0.71	0.71	nr	nr																
c.1820T>A	p.L607H	A	0.99	0.99	0.88	0.88	0.97	0.97	0.79	0.79	No effect	No change	No change	No change	No effect	No effect	No effect	No effect	PrD	Aff	NA	Impaired	Likely neutral	Likely neutral	NP	Normal*
		D	0.9	0.9	NR	NR	0.71	0.71	nr	nr																
c.1852_1853AA>GC	p.K618A	A	0.99	0.99	0.88	0.88	0.97	0.97	0.79	0.79	No effect	2 sites created	3 sites destroyed	3 sites destroyed	Aberrant ESE	Aberrant ESE	No effect	No effect	PsD	Aff	Del	Inconclusive	Likely neutral	Likely neutral	Likely neutral	Normal
		D	0.41	0.41	NR	0.1	0.71	0.71	NR	NR																
c.1963A>G	p.I655V	A	0.43	0.43	0.62	0.53	0.43	0.39	0.66	0.66	No effect	2 sites destroyed	1 site destroyed	1 site created	Aberrant ESE	Aberrant ESE	No effect	No effect	Ben	Tol	Bor	Inconclusive	Likely neutral	Likely neutral	NP	Normal*
		D	0.91	0.91	0.9	0	0	0	0.76	0.77																
c.1984A>C	p.T662P	A	0.43	0.43	0.62	0.62	0.43	0.42	0.66	0.66	No effect	No change	No change	No change	No effect	No effect	Partial deletion of exon 17	Inconclusive	Ben	Tol	Bor	Inconclusive	Likely neutral	Inconclusive	NP	Inconclusive*
		D	0.91	0.9	0.9	0.79	0	0	0.76	0.76																
c.1989G>T	p.E663D	A	0.43	0.43	0.62	0.62	0.42	0.41	0.66	0.66	Inconclusive	No change	No change	No change	No effect	Inconclusive SS	Skipping exon 17	Aberrant transcript	Ben	Tol	Neu	Neutral	Likely neutral	Likely neutral	NP	Normal*
		D	0.91	0.76	0.9	NR	0.88	NR	0.76	NR																
c.2041G>A	p.A681T	A	0.89	0.89	0.92	0.92	0.71	0.77	NR	NR	No effect	No change	No change	No change	No effect	No effect	No effect	No effect	PsD	Aff	Bor	Inconclusive	Likely neutral	Likely neutral	Likely neutral	Normal*
		D	0.92	0.92	1	1	0.95	0.95	0.8	0.8																
c.2059C>T	p.R687W	A	0.89	0.89	0.92	0.55	0.71	0.56	NR	NR	No effect	1 site destroyed	1 site created	2 sites created	Aberrant ESE	Aberrant ESE	No effect	No effect	PrD	Aff	Bor	Inconclusive	Likely neutral	Inconclusive	Likely neutral	Inconclusive*
		D	0.92	0.92	1	0.68	0.95	0.93	0.8	0.8																
c.2066A>G	p.Q689R	A	0.89	0.89	0.92	0.87	0.82	0.82	nr	nr	No effect	No change	2 sites destroyed	No change	Inconclusive	Inconclusive ESE	No effect	No effect	Ben	Tol	Bor	Inconclusive	Likely neutral	Inconclusive	Likely neutral	Inconclusive
		D	0.92	0.92	1	1	0	0	0.8	0.8																
c.2146G>A	p.V716M	A	0.58	0.58	0.01	0.01	NR	NR	0.67	0.67	No effect	No change	1 site created	1 site created and 2 sites destroyed	Inconclusive	Inconclusive ESE	No effect	No effect	Ben	Tol	Neu	Neutral	Likely neutral	Inconclusive	NP	Inconclusive*
		D	NR	NR	NR	NR	NR	NR	NR	NR																

For splice site prediction, major alternations are destroyed or created SS, or score modifications ≥45%; For enhancer site prediction, major alterations are destroyed or created ESE. Predictions are interpreted as inconclusive when the same results are not obtained by all the programs used. Abbreviations: SS, splice site; A, acceptor consensus splice site; D, donor consensus splice site; NR, consensus splice site not recognized; NA, not available; NP, not performed;* Subcelular localization analysis is not performed. Data on variants have been extracted from (Arnold, et al., 2009; Auclair, et al., 2006; Kansikas, et al., 2011; Pagenstecher, et al., 2006; Takahashi, et al., 2007; Tournier, et al., 2008).

ARTICLE 3

REFINING THE ROLE OF *PMS2* IN LYNCH SYNDROME: GERMLINE MUTATIONAL ANALYSIS IMPROVED BY COMPREHENSIVE ASSESSMENT OF VARIANTS

Ester Borràs, Marta Pineda, Juan Cadiñanos, Jesús del Valle, Angela Brieger, Innga Hinrichsen, Matilde Navarro, Joan Brunet, Ruben Cabanillas, Van der Klift, Conxi Lázaro, Guido Plotz, Ignacio Blanco i Gabriel Capellá.

Manuscrit sotmès a publicació

Resum: Les mutacions en heterozigosis detectades en els gens reparadors són les causants de la síndrome de Lynch. Tot i que la majoria d'aquestes mutacions ocorren en els gens *MLH1* o *MSH2*, la contribució de *PMS2* no està tan clarament definida. L'objectiu d'aquest estudi era analitzar la contribució de *PMS2* en el síndrome de Lynch, a partir de l'avaluació de la patogenicitat de les variants de significat desconegut (VSD) detectades en l'anàlisi mutacional del gen *PMS2* en una sèrie de pacients espanyols. D'una cohort de 202 pacients amb sospita de síndrome de Lynch, tretze pacients mostren pèrdua d'expressió de PMS2 en els tumors, on es va analitzar les mutacions germinals en PMS2, mitjançant una estratègia basada en *Long-Range* PCR i MLPA. L'anàlisi de patogenicitat de les VSD es va analitzar mitjançant les dades clinico-patològiques, la freqüència en una població control i en una varietat de assaigs *in silio* i *in vitro*, tant a nivell de RNA com de proteïna. Un total de 25 canvis van ser detectats en el gen *PMS2.* Dotze d'ells van ser classificats coma polimorfismes després de l'anàlisi en una sèrie control. Nou d'ells es van ser classificats com a patogènics: set alteracions basades en la seva naturalesa molecular i dues d'ells després de demostrar el seu defecte funcional (c.538-3C>G afectava el correcte processament del mRNA i c.137G>T mostrava un defecte en l'activitat reparadora). Les VSD, c.59G>A i c.1569C>G, van ser classificades com a probablement neutrals, mentre que les variants c.384G>A i c.1789A>T no es va poder classificar. En un 69% del pacients amb tumors associats a la síndrome de Lynch i amb pèrdua d'expressió de la proteïna PMS2, es van detectar mutacions patogèniques en el gen *PMS2*. L'anàlisi funcional exhaustiu ha estat útil per la classificació de VSD en el gen *PMS2* i contribueix a refinar l'impacte real de *PMS2* en la síndrome de Lynch.

Refining the Role of *PMS2* in Lynch Syndrome: Germline Mutational Analysis Improved by Comprehensive Assessment of Variants

E. Borràs[1], M. Pineda[1], J. Cadiñanos[2], J. del Valle[1], A. Brieger[3], I. Hinrichsen[3], R. Cabanillas[2], M. Navarro[1], J. Brunet[4], X. Sanjuan[5], H. van der Klift[6], C. Lázaro[1], G.Plotz[3], I. Blanco[1] and G. Capellá[1.]

(1) Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL, Hospitalet de Llobregat, Barcelona, Spain; (2) Institute of Molecular and Oncological Medicine of Asturias, IMOMA, Oviedo, Spain; (3) Medical Clinic 1, Johann Wolfgang Goethe-University, Frankfurt, Germany; (4) Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IdIBGI, Girona, Spain; (5) Department of Pathology, University Hospital of Bellvitge, Hospitalet de Llobregat, Barcelona, Spain (6) Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

ABSTRACT

BACKGROUND & AIM: The majority of MMR gene mutations causing Lynch syndrome (LS) occur either in *MLH1* or *MSH2***.** However, the relative contribution of *PMS2* is less well defined. The aim of this study was to evaluate the role of *PMS2* in LS by assessing the pathogenicity of variants of unknown significance (VUS) detected in the mutational analysis of *PMS2* in a series of Spanish patients. **METHODS:** From a cohort of 202 LS-suspected patients, thirteen patients showing loss of PMS2 expression in tumors were screened for germline mutations in *PMS2,* using a long-range PCR-based strategy and MLPA. Pathogenicity assessment of *PMS2* VUS was performed evaluating clinico-pathological data, frequency in control population and *in silico* and *in vitro* analyses at the RNA and protein level. **RESULTS:** Overall 25 different *PMS2* DNA changes were detected. Twelve changes were classified as polymorphisms. Nine of them were classified as pathogenic: seven alterations based on their molecular nature and two after demonstrating a functional defect (c.538-3C>G affected mRNA processing and c.137G>T impaired MMR activity). Two VUS (c.59G>A and c.1569C>G) were classified as likely neutral while the c.384G>A and c.1789A>T variants remained as VUS. **CONCLUSION:** Pathogenic *PMS2* mutations were detected in 69% of patients harboring LS associated tumors with loss of PMS2 expression. In all, *PMS2* mutations account for 6% of the LS cases identified. The comprehensive functional analysis shown here has been useful in the classification of *PMS2* VUS and contributes in refining the role of *PMS2* in LS.

INTRODUCTION

Heterozygous germline mutations in either one of the mismatch repair (MMR) genes are responsible for Lynch syndrome (LS)[1]. Although *MLH1* and *MSH2* are the most frequently mutated genes in LS, the contribution of *PMS2* is less well defined, representing 1-13.7% of all MMR gene mutations[2-6]. The analysis of *PMS2* is especially complex due to the presence of multiple pseudogenes. Fourteen of these pseudogenes share a high homology with the *PMS2* 5' end[7], and the pseudogene *PMS2CL* has strong homology to *PMS2* exons 9 and 11-15[7, 8]. In order to avoid spurious amplification of pseudogenes, long-range (LR) PCR strategies have improved the yield of mutational analysis[9-11]. When patients are selected for mutation analysis on the basis of negative PMS2 staining in tumors, the reported mutation rate varies considerably (0%-62%)[2, 9, 11, 12], likely linked to differences in the selection criteria and in the strategy used for genetic analysis.

Most of the reported mutations identified in LS suspected patients are truncating and classified as pathogenic. However, a significant number (30%) are classified as variants of unknown significance (VUS)[13]. The assessment of pathogenicity is difficult and requires the integration of clinical, pathological and functional data (reviewed by Rasmussen[14]). While functional analyses of MMR variants are frequently performed for *MLH1* and *MSH2* genes, the impact of *PMS2* VUS has been analyzed only incompletely in a limited number of cases. Studies have been done at mRNA level[15, 16], on protein expression[8, 15], and on protein interaction[17, 18]. Also, site-directed mutagenesis of residues located in the PMS2 ATPase domain has been performed in function-structure studies[19, 20], and the impact of PMS2 VUS in other processes such as apoptosis has been explored[17, 21, 22].

The aim of this study was to evaluate the contribution of *PMS2* in LS by assessing the

pathogenicity of VUS detected in the mutational analysis of *PMS2* in a series of Spanish patients with LS-associated tumours showing loss of PMS2 expression. For *PMS2* VUS classification a strategy was followed, which integrates frequency in controls, family history and tumor pathology data with a variety of *in silico* and *in vitro* analyses at RNA and protein level[23].

MATERIAL AND METHODS

Patients and samples

Mutational screening of MMR genes was performed in 202 probands with LS-associated tumours showing MSI phenotype or loss of MMR protein expression. In all, 133 pathogenic mutations in MMR genes (*MLH1*, *MSH2* or *MSH6)* were identified (unpublished results). Mutational analysis of *PMS2* was performed in 13 probands with tumors showing loss of PMS2 but normal MLH1 expression by IHC. One patient fulfilled Amsterdam criteria, 8 revised Bethesda criteria[24, 25] and the remaining patients were referred to the Genetic Counseling Unit after MMR staining in tumors that displayed characteristic MSI histology[26].

Genomic DNA was extracted from whole blood using the FlexiGene DNA kit (Qiagen). DNA samples from 95 controls in a hospital based CRC case-control study were used to analyze the frequency of the detected *PMS2* VUS[27]. Informed consent was obtained from all individuals, and ethics committee approved this study.

Mutational analysis in *PMS2*

Genomic rearrangements in *PMS2* were analyzed using MLPA commercial kit P008-A1 (MRC-Holland, Amsterdam, the Netherlands). PCR products were analyzed on an ABI 3730 sequencer using GeneMapper software (Applied Biosystems, Forster City, CA, USA). The proportion of each peak relative to the height of all peaks was calculated for each sample and then compared to proportions for the corresponding peak averaged for a set of at least ten normal DNA samples. Ratios of 1.0 (±0.2) were treated as normal copy number. Ratios of 0.5 were considered as deletions and ratios of 1.5 as duplications. Each positive result was confirmed in a second independent MLPA reaction. In one case, confirmation of the identified rearrangement was performed using the recently redesigned version of the *PMS2* MLPA kit (P008-B1) that has been shown to improve the reliability of copy number detection in the 3' end of the gene[28, 29]. In another case, LR-PCR amplification and sequencing were used to characterize the breakpoint (see below). Identification of repeat elements in genomic sequences was performed using the Repeat Masker program.

Point mutations were analyzed using either of two previously described LR-PCR procedures[9, 11]. In brief, amplicons spanning exons 1–6 (long-range amplicon LR1), 6–10 (LR2), 10-12 (LR3) and 13-15 (LR4)[9] or, alternatively, 10-15 (LR3', substituting LR3 and LR4)[11],were generated by amplification using LaTaq polymerase (TaKaRa Bio Inc., Otsu, Shiga, Japan). Fifteen µl of each PCR product were diluted in TE 1x up to a final volume of 180 µl. One µl of this dilution was used as the template for exon-specific PCR using Megamix Double (Microzone limited, Haywards Heath, UK) and specific primers[9, 11].Exon-specific amplification of exons 6, 7, 8 and 10 were performed with Megamix Double and specific primers[9, 11]. PCR products were sequenced using Big Dye Terminator v.3.1 Cycle Sequencing kit (Applied Biosystems) on an Applied Biosystems 3130XL Genetic Analyzer. The presence of the reported retrotranspositional insertion in *PMS2* intron 7 was tested in 4 patients with no identified mutation in *PMS2*, using the described PCR diagnostic test[30].

Sequences were compared to the NCBI reference sequence NG_008466.1 (NM_000535.5; NP_000526.1). Mutation nomenclature of *PMS2* is according to HGVS recommendations (version 2.0) with nucleotide 1 corresponding to the A of the ATG-translation initiation codon. The identified *PMS2* changes have been submitted to the LOVD database (www.LOVD.nl).

Analysis of frequency in controls

Screening for the identified *PMS2* variants in control population was performed by LR-PCR followed by exon specific PCR as described above.

Cosegregation analysis

DNA of relatives was screened for the familial associated *PMS2* variant by direct sequencing using LR-PCR followed by exon specific PCR. Cosegregation was considered positive if two or more affected members with LS-associated tumors were carriers of the studied variant.

Bioinformatic analysis

DNA sequences containing the identified *PMS2* variants were analyzed using bioinformatic tools addressed to evaluate their impact at RNA and protein level. Disruption or creation of SS (Splice Sites) were evaluated using NNSplice[31], Splicerport[32], NetGene2 Server[33] and SoftBerry[34]. Exonic variants were analyzed for putative ESE (exonic splicing enhancer) elements by

ESEfinder[35], RESCUE-ESE[36] and PESX[37]. The impact of missense variants at the protein level was analyzed using PolyPhen-2[38], SIFT[39] and CONDEL[40].

In addition, the putative effect of c.137G>T (p.S46I) missense variant was investigated in a structural model of the MutL*α* N-terminal heterodimer, which was based on a homology model of MLH1-NTD and a crystal structure of PMS2-NTD. Figure was created using the PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC. The 100 best hits were aligned using ClustalW and conservation was plotted in Weblogo presentation[41].

Lymphocyte culture and mRNA splicing analysis

Human lymphocytes from *PMS2* variant carriers were cultured in the absence or presence of puromycin (Sigma). Total RNA was extracted from cultured lymphocytes and cDNA was synthesized as described[23]. *PMS2* cDNA fragments were amplified by RT-PCR using LaTaq polymerase (Takara) in two overlapping fragments and sequenced as described[42]. Transcripts of carriers were compared with transcripts from 3 control lymphocyte cultures.

Allele-specific expression analysis

Allele-specific expression (ASE) was analyzed by SNuPE[23]. ASE was calculated by dividing the proportion of variant/wild-type allele in cDNA by the proportion of variant/wild-type allele in gDNA. We used ≤0.5 as a threshold value for ASE definition[43]. Experiments were performed in triplicate.

Plasmids and site-directed mutagenesis

pcDNA3.1_MLH1 and pN1_PMS2 plasmids, kindly provided by Dr. Kolodner and Dr. Nyström-Lahti, were used in MMR assays and expression analyses. pECFP_C1_MLH1 and pDSred_C1_PMS2 vectors were used in subcellular localization assays[44].

Selected *PMS2* missense variants were constructed by site-directed mutagenesis using QuikChange Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. Sequencing was used to verify the presence of the variants. In addition, two control plasmids were constructed: c.208G>A (p.D70N)[19] used as mutation control and c.1408C>T (p.P470S)[9] used as a neutral control.

HEK293T cells culture and cell transfection

Transfection of HEK293T cells (deficient for endogenous MLH1 and PMS2) cells was carried out as described[23]. In brief, HEK293T cells were transfected at 50-70% confluence with MLH1 and PMS2 expression plasmids (1µg/ml, respectively) using 10µl/ml of the cationic polymer polyethylenimine (Polysciences, Warrington, PA; stock solution 1mg/ml). After 48h, cells were prepared for confocal laser scanning microscopy or protein extraction.

MMR activity assay

MMR assays were performed as described[45].Repair efficiency was measured as the quotient of the intensities of those bands indicating repair divided by the sum of all band intensities. Relative repair efficiency was calculated by dividing the value of the tested variant protein through the value of a wild-type protein that had been expressed, processed and tested in parallel. Experiments were performed in quadruplicate.

MLH1 and PMS2 protein expression analysis

MLH1 and PMS2 expression levels were examined as described[23]. Alfa-tubulin and EGFP expression was assessed in parallel and used as loading and transfection efficiency controls. All experiments were performed in triplicate.

Confocal laser microscopy

Confocal laser microscopy was carried out as described[44]. Mock control for MLH1 and PMS2 was performed using empty pECFP-C1 and pDSRed-C1 vectors. Experiments were performed in triplicate.

RESULTS

Clinical characteristics of patients with PMS2 negative tumors

Mutational analysis of *PMS2* was performed in 13 probands (10 males and 3 females) diagnosed with LS-associated tumors showing loss of PMS2 but normal MLH1 expression (Table 1). MSI was evident in all tumors analyzed (n=10). Ten of the 13 patients (77%) were diagnosed of CRC, 2 of small bowel tumors and 1 of endometrial cancer. The median age at diagnosis was 52 years (range 42–72).

Pathogenic mutations identified in patients with PMS2 negative tumors

A total of 25 different *PMS2* DNA changes were detected in the 13 probands (Table 1 and Supp.

Figure 1). Seven *PMS2* alterations, detected in 7 patients, were classified as definitely pathogenic: two gross deletions, two splice-site mutations, one frameshift, one nonsense and one mutation affecting the first codon (Table 1).

In family C, the characterization of the *PMS2* exon 6 deletion revealed a novel *PMS2* rearrangement which results in a deletion of 1.1 Kb, g.14354_15455del (c.538-478_705+456del). Sequence analysis of the deletion breakpoint pointed to homologous recombination involving *Alu* elements as the underlying mechanism (Figure 1). This deletion was identified in one individual with CRC (carrier C2), one affected by CRC and skin cancer (carrier C1) and in one 33 years old asymptomatic carrier (Table 1 and Supp. Figure2).

The gross deletion identified in family F, c.989-?_(*175_?), encompassed exon 10 until the end of the coding region and was not further analyzed. It was identified in one individual affected by small bowel cancer (F1) and another individual affected by endometrial and ovarian cancer (F2). Both tumors diagnosed in carrier F2 also showed loss of PMS2 expression (Table 1 and Supp. Figure 2).

In families B and G, the impact at RNA level of the two mutations affecting a consensus acceptor SS was studied. In family B, the c.164-2A>G mutation leads to the creation of a new acceptor site 8 bp downstream (r.164_171del), predicted to generate a truncated protein (p.D55Afs*2) (Figure 2A). This change, identified in carrier B1 affected by CRC, was found associated with *PMS2* c.384G>A variant. Segregation analysis revealed that both alterations co-occurred *in cis* (Supp. Figure 2).

In family G, the second *PMS2* splice-site mutation, c.989-2A>G, resulted in exon 10 skipping (r.989_1144del), which is predicted to generate an in-frame deletion of 52 amino acids (p.E330_G381del) that affects the C terminal domain (IPRO13507)[46] (Figure 2B). The variant was identified in two individuals affected by CRC (carriers G1 and G2) (Supp. Figure 2). In addition, this *PMS2* alteration was identified in seven asymptomatic carriers (age 44-71 years old).

In family D, the *PMS2* frameshift mutation c.780delC (p.N261Mfs*46) was identified in one patient diagnosed by CRC displaying suggestive MSI-histology. Within this family, two other non-affected carriers of the same *PMS2* alteration (aged 32 and 27) were identified (Supp. Figure 2). In family E, fulfilling Amsterdam criteria, the *PMS2* nonsense mutation c.943C>T (p.R315*) was identified in two individuals affected by CRC and one asymptomatic carrier (age 52 years old) (Supp. Figure 2). Finally, in one individual affected by endometrial cancer of family A, the *PMS2* alteration c.1A>G (p.?), affecting the first codon, was detected. Two other carriers, one of them aged 77, were not affected (Supp. Figure 2).

PMS2 **VUS identification and frequency in healthy controls**

Mutational analysis of *PMS2* detected eighteen DNA changes with unclear functional impact (Table 1). Thirteen of the eighteen variants identified were detected in our control population series at a polymorphic frequency (>1%) (Table 2); seven of them previously reported (Supp. Table 1). The remaining five VUS (see below) were not detected in our control population. The screening in the control series lead to the identification of other *PMS2* DNA changes (Supp. Table 2).

Pathogenicity assessment of *PMS2* VUS
Clinico-pathologic characteristics

The five identified VUS (c.137G>T, c.384G>A, c.538-3C>G**,** c.1569C>G, c.1789A>T) were further analysed together with the c.59G>A variant that, albeit observed at polymorphic frequency (2.7%), was predicted as pathogenic in a multifactorial model[47] (Table 1). The six variants were identified in 5 individuals (Table 1 and Supp. Figure 2).
The c.59G>A variant (p.R20Q) was detected in one individual of family H affected by CRC (carrier H1). The c.137G>T (p.S46I) variant was identified in two individuals of family I [one harboring CRC and three skin tumors (carrier I1), and another individual affected by bladder cancer (carrier I2)]. The c.1569C>G variant was identified in one affected individual of family K showing mixed features of LS and adenomatous polyposis. He was diagnosed of two synchronous CRC at age 51, one of these with MSI and PMS2 loss and one microsatellite stable tumor. In the latter case no IHC study was performed. The association with 15 adenomatous polyps lead to *APC* and *MUTYH* testing unveiling an additional germline silent *APC* VUS (NM_000038.5 c.6525A>G). The c.538-3C>G and c.1789A>T variants coexisted *in cis* in an individual diagnosed of small bowel cancer at age 67 (carrier J1). The tumor showed MSH6 and PMS2 loss of expression. Finally, the c.384G>A was detected in family B *in cis* with the pathogenic c.164-2A>G mutation (see above).

Bioinformatic prediction, mRNA analysis and coexistence of PMS2 VUS

At least one *in silico* prediction program predicted relevant changes in splice sites (SS) for the c.538-3C>G and c.59G>A variants (Table 3 and

Supp. Table 3). Two of the three programs predicted alterations on exonic splicing enhancers for variants c.59G>A, c.137G>T and c.384G>A (Table 3 and Supp. Table 3). However, inconsistencies were observed regarding the predicted effect among programs.

Sequencing of RT-PCR products identified aberrantly expressed transcripts only in *PMS2* c.538-3C>G VUS sample (Figure 2C), which lead to the creation of two transcripts. One of them produced a deletion of 47 base pairs (r.538_586del), which is predicted to generate a truncated PMS2 protein (p.E180Qfs*5), and the other lead to exon 6 skipping (r.538_705del), which is predicted to generate an in-frame deletion of 55 amino acids (p.E180_L236del). Puromycin treatment of cultured lymphocytes increased the relative amount of the frameshift transcript (Figure 2C).

PMS2 c.538-3C>G VUS assessment is complex. It coexists *in cis* with the c.1789A>T missense variant (Table 3), being the G intronic allele at 538-3 associated with the exonic 1789 T allele. At the RNA level, an allelic imbalance – as assessed by the 1789 nucleotide - was observed in cultured cells without puromycin (Supp. Table 4). In summary, the *PMS2* VUS c.538-3C>G mutation affects the mRNA processing and stability pointing to a pathogenic effect. In the remaining cases, there was no evidence of instability of the P*MS2* VUS alleles (Supp.Table 4).

The silent c.1569C>G variant was classified as likely neutral since it did not affect mRNA processing or stability. As stated above, c.384G>A variant co-occurred *in cis* with c.164-2A>G pathogenic mutation in individual B1 (Supp. Figure 2) and cannot be definitely classified. Finally, the coexistence *in cis* of c.1789A>T variant with the c.538-3C>G mutation lead us to categorize the former as VUS.

In silico and in vitro protein analyses of PMS2 missense variants

The c.59G>A (p.R20Q) and c.137G>T (p.S46I) missense variants were analyzed at the protein level. *PMS2* c.59G>A (p.R20Q) variant was classified as likely neutral by the 3 *in silico* prediction programs used (Table Supp. 3), while c.137G>T (p.S46I) variant was predicted as pathogenic (Table Supp. 3). In fact, the S46 amino acid residue is located directly within the nucleotide binding pocket of the ATPase domain suggesting an interference with the ATPase function (Supp. Figure 3). Our results from the MMR assay showed a significantly reduced MMR activity for c.137G>T (p.S46I) variant (12.75%±4.68) while c.59G>A (p.R20Q) did not show differences in MMR activity compared to PMS2 wild-type (88.20%±9.64) (Figure 3). No differences were observed regarding expression analysis and nuclear localization when compared to the wild type protein (Figure 3). Based on the impaired MMR activity, the c.137G>T (p.S46I) mutation was classified as pathogenic.

DISCUSSION

We have analyzed the *PMS2* in a series of Spanish LS-suspected patients showing loss of PMS2 and normal MLH1 expression. Using a well established mutation detection strategy[9, 11] together with a comprehensive assessment of pathogenicity, we have been able to classify most of the variation detected. Pathogenic *PMS2* mutations have been detected in the majority of patients analyzed (69%, 9/13) and account for 6% of the LS cases identified in our series (9 of 142 LS probands including the 9 *PMS2* mutations reported herein).

By using LR-PCR strategies[9, 11] and MLPA analysis a significant number of bona fide *PMS2* DNA changes have been detected (25 in 13 patients). Analysis of polymorphisms in *PMS2* can be hampered by the use of techniques not avoiding co-amplification of pseudogenes. Our LR-PCR assay has been able to identify thirteen *PMS2* polymorphisms, six of them novel, which will help in the genetic counseling of families.

The RNA analysis of three novel *PMS2* splice-site mutations, c.538-3C>G, c.164-2A>G, and c.989-2A>G, confirmed their pathogenicity. Although there is a general agreement to classify variants localized within the consensus splice-sites as pathogenic mutations, splicing analyses confirm the presence of aberrant transcripts and better predict their consequence at protein level[48]. The c.164-2A>G mutation is predicted to produce a premature stop codon, and c.989-2A>G variant is predicted to produce an in-frame deletion of part of the PMS2 dimerization domain[46]. cDNA assays (splicing and ASE) together with segregation of a coexisting variant were necessary for the classification of the c.538-3C>G variant.

Two of the pathogenic mutations identified (22%) are gross deletions, in line with previously reported frequencies in PMS2 series (21%-37%)[10, 11]. However, genomic rearrangements may be still underreported due to the fact that only 11 of the 15 *PMS2* exons can be reliably analysed with the MLPA kit used in this study. Detection of copy number changes in the 3'end of *PMS2* (exons 12-15) is limited by the frequent gene conversion between *PMS2* and the pseudogene *PMS2CL*. Recently, deletions in this region have been described using a new version of the kit (P008-B1) that allows for reliable detection of copy number changes in exons 12-

15[28]. Moreover, genomic rearrangements that do not result in copy number changes, such as the recently described retrotranspositional insertion in *PMS2* intron 7, are also undetectable with current diagnostic testing methods[30].

Functional assays at protein level demonstrated that c.137G>T (p.S46I) is a pathogenic mutation significantly impairing MMR activity. p.S46I variant has been recurrently identified[9-11] and classified as likely pathogenic because it was detected in compound heterozygotes with constitutional mismatch repair-deficiency (CMMR-D). Recently, its founder effect of has been demonstrated in 10 unrelated American families sharing a common haplotype of 375 Kb[49]. The affected amino acid residue is within the ATP binding pocket, which is consistent with the observed MMR activity defect. We have not been able to observe a diminished expression levels which was reported by others[8]. Of note, p.S46N variant, located in the same position, was also detected in individuals with CMMR-D[10, 50].

The silent c.1569C>G and c.59G>A variants were classified as likely neutral. The pathogenicity of c.59G>A (p.R20Q) has been controversial[3, 17, 47, 51]. We have confirmed that its frequency in control population is >1%[3, 51] and that it does not influence RNA processing, MMR activity, protein expression or subcellular localization. Two variants, the novel c.384G>A and the missense c.1789A>T, are still considered VUS due to their location *in cis* with the pathogenic variants c.164-2A>G and c.538-3C>G, respectively.

PMS2 pathogenic mutations account for 6% of LS patients. This proportion is difficult to compare with other studies, due to differences in selection criteria and strategies used for mutational analysis. We consider that the strategy utilized (MLPA, LR-PCR/sequencing and functional analysis of variants) is robust as reflected by the yield of *PMS2* mutation detection (69%) of patients with tumors showing loss of PMS2 expression in the upper limit of published results (0-62%[2, 9, 11, 12]). Our results further support the use of PMS2 immunohistochemistry as a mandatory step in identifying candidates for *PMS2* mutation analysis. As we learn more about PMS2 testing and the use of immunostaining of PMS2 protein disseminates, the frequency of detection of *PMS2* positive families may increase.

Our results confirm the distinct expressivity of *PMS2* mutations compared to other MMR genes. Up to 44% (4/9) of families with detected *PMS2* pathogenic mutations did not meet any clinical LS criteria, showing their low sensitivity for identification of *PMS2* mutation carriers[3, 9, 12, 52]. The recurrent identification of elderly asymptomatic carriers is in agreement with the lower penetrance of these mutations[6, 10]. This should be taken into account when considering co-segregation information as part of algorithms of VUS classification or multifactorial likelihood models, thereby diminishing its relative weight[10, 11].

Very limited data are available on functional analysis of *PMS2* VUS (reviewed in Table 4). Our study presents, for the first time, a comprehensive assessment of *PMS2* VUS. We used the same strategy utilized for the evaluation of *MLH1* VUS[23] to evaluate*PMS2* variants, which is similar to one recently proposed by Rasmussen and collaborators[14]. While ongoing efforts aimed at reaching consensus on the interpretation of the relative weight of all clinical and experimental evidence on MMR variants finalize, here we contribute with a thorough analysis of a number of *PMS2* variants for which detailed clinical, pathological and functional assessment has been performed.

ACKNOWLEDGMENTS

This work was supported by grants from Ministerio de Ciencia e Innovación (SAF 06-06084; 09-07319, 12-33636), Fundació Gastroenterologia Dr. Francisco Vilardell [F05-01], Ministerio de Educación y Ciencia Spanish Networks RTICC [RD06/0020/1050, 1051], Acción en Cáncer (Instituto de Salud Carlos III), the Autonomous Government of Catalonia [2009-SGR-290] and Fundación Científica de la AECC and Fundación Centro Médico de Asturias. EB is a recipient of an IDIBELL fellowship and of travel grant by RTICC. Authors have no conflict of interest to declare.

REFERENCES

1. Vasen HF. Review article: The Lynch syndrome (hereditary nonpolyposis colorectal cancer). Aliment Pharmacol Ther 2007;26 Suppl 2:113-26.
2. Niessen RC, Kleibeuker JH, Westers H, et al. PMS2 involvement in patients suspected of Lynch syndrome. Genes Chromosomes Cancer 2009;48:322-9.
3. Hendriks YM, Jagmohan-Changur S, van der Klift HM, et al. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 2006;130:312-22.
4. Hampel H, Frankel W, Panescu J, et al. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. Cancer Res 2006;66:7810-7.

5. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 2005;352:1851-60.
6. Talseth-Palmer BA, McPhillips M, Groombridge C, et al. MSH6 and PMS2 mutation positive Australian Lynch syndrome families: novel mutations, cancer risk and age of diagnosis of colorectal cancer. Hered Cancer Clin Pract 2010;8:5.
7. De Vos M, Hayward BE, Picton S, et al. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am J Hum Genet 2004;74:954-64.
8. Nakagawa H, Lockman JC, Frankel WL, et al. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res 2004;64:4721-7.
9. Clendenning M, Hampel H, LaJeunesse J, et al. Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat 2006;27:490-5.
10. Senter L, Clendenning M, Sotamaa K, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 2008;135:419-28.
11. Vaughn CP, Robles J, Swensen JJ, et al. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Hum Mutat 2010;31:588-93.
12. Truninger K, Menigatti M, Luz J, et al. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160-71.
13. Peltomaki P, Vasen H. Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 2004;20:269-76.
14. Rasmussen LJ, Heinen CD, Royer-Pokora B, et al. Pathological assessment of mismatch repair gene variants in Lynch syndrome: Past, present, and future. Hum Mutat 2012.
15. Auclair J, Leroux D, Desseigne F, et al. Novel biallelic mutations in MSH6 and PMS2 genes: gene conversion as a likely cause of PMS2 gene inactivation. Hum Mutat 2007;28:1084-90.
16. van der Klift HM, Tops CM, Bik EC, et al. Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. Hum Mutat 2010;31:578-87.
17. Marinovic-Terzic I, Yoshioka-Yamashita A, Shimodaira H, et al. Apoptotic function of human PMS2 compromised by the nonsynonymous single-nucleotide polymorphic variant R20Q. Proc Natl Acad Sci U S A 2008;105:13993-8.
18. Yuan ZQ, Gottlieb B, Beitel LK, et al. Polymorphisms and HNPCC: PMS2-MLH1 protein interactions diminished by single nucleotide polymorphisms. Hum Mutat 2002;19:108-13.
19. Raschle M, Dufner P, Marra G, et al. Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. J Biol Chem 2002;277:21810-20.
20. Tomer G, Buermeyer AB, Nguyen MM, et al. Contribution of human mlh1 and pms2 ATPase activities to DNA mismatch repair. J Biol Chem 2002;277:21801-9.
21. Ji G, Long Y, Zhou Y, et al. Common variants in mismatch repair genes associated with increased risk of sperm DNA damage and male infertility. BMC Med;10:49.
22. Deschenes SM, Tomer G, Nguyen M, et al. The E705K mutation in hPMS2 exerts recessive, not dominant, effects on mismatch repair. Cancer Lett 2007;249:148-56.
23. Borras E, Pineda M, Brieger A, et al. Comprehensive functional assessment of MLH1 variants of unknown significance. Hum Mutat 2012.
24. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261-8.
25. Vasen HF, Watson P, Mecklin JP, et al. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 1999;116:1453-6.
26. Risio M, Reato G, di Celle PF, et al. Microsatellite instability is associated with the histological features of the tumor in nonfamilial colorectal cancer. Cancer Res 1996;56:5470-4.
27. Moreno V, Gemignani F, Landi S, et al. Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. Clin Cancer Res 2006;12:2101-8.
28. Vaughn CP, Hart KJ, Samowitz WS, et al. Avoidance of pseudogene interference in the detection of 3' deletions in PMS2. Hum Mutat 2011.
29. Wernstedt A, Valtorta E, Armelao F, et al. Improved multiplex ligation-dependent probe amplification analysis identifies a deleterious PMS2 allele generated by recombination with crossover between PMS2 and PMS2CL. Genes Chromosomes Cancer 2012;51:819-31.
30. van der Klift HM, Tops CM, Hes FJ, et al. Insertion of an SVA element, a nonautonomous retrotransposon, in PMS2 intron 7 as a novel cause of Lynch syndrome. Hum Mutat 2012;33:1051-5.
31. Reese MG, Eeckman FH, Kulp D, et al. Improved splice site detection in Genie. J Comput Biol 1997;4:311-23.
32. Dogan RI, Getoor L, Wilbur WJ, et al. SplicePort--an interactive splice-site analysis tool. Nucleic Acids Res 2007;35:W285-91.
33. Hebsgaard SM, Korning PG, Tolstrup N, et al. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res 1996;24:3439-52.

34. Burset M, Seledtsov IA, Solovyev VV. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 2001;29:255-9.
35. Cartegni L, Wang J, Zhu Z, et al. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 2003;31:3568-71.
36. Fairbrother WG, Yeh RF, Sharp PA, et al. Predictive identification of exonic splicing enhancers in human genes. Science 2002;297:1007-13.
37. Zhang XH, Kangsamaksin T, Chao MS, et al. Exon inclusion is dependent on predictable exonic splicing enhancers. Mol Cell Biol 2005;25:7323-32.
38. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.
39. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002;12:436-46.
40. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 2011;88:440-9.
41. Crooks GE, Hon G, Chandonia JM, et al. WebLogo: a sequence logo generator. Genome Res 2004;14:1188-90.
42. Etzler J, Peyrl A, Zatkova A, et al. RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. Hum Mutat 2008;29:299-305.
43. Perera S, Li B, Tsitsikotas S, et al. A novel and rapid method of determining the effect of unclassified MLH1 genetic variants on differential allelic expression. J Mol Diagn 2010;12:757-64.
44. Brieger A, Adam R, Passmann S, et al. A CRM1-dependent nuclear export pathway is involved in the regulation of MutLalpha subcellular localization. Genes Chromosomes Cancer 2011;50:59-70.
45. Plotz G, Welsch C, Giron-Monzon L, et al. Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic Acids Res 2006;34:6574-86.
46. Sjursen W, Bjornevoll I, Engebretsen LF, et al. A homozygote splice site PMS2 mutation as cause of Turcot syndrome gives rise to two different abnormal transcripts. Fam Cancer 2009;8:179-86.
47. Pastrello C, Pin E, Marroni F, et al. Integrated analysis of unclassified variants in mismatch repair genes. Genet Med 2011;13:115-24.
48. Spurdle AB, Couch FJ, Hogervorst FB, et al. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 2008;29:1304-13.
49. Tomsic J, Senter L, Liyanarachchi S, et al. Recurrent and founder mutations in the PMS2 gene. Clin Genet 2012.
50. Jackson CC, Holter S, Pollett A, et al. Cafe-au-lait macules and pediatric malignancy caused by biallelic mutations in the DNA mismatch repair (MMR) gene PMS2. Pediatr Blood Cancer 2008;50:1268-70.
51. Liu T, Yan H, Kuismanen S, et al. The role of hPMS1 and hPMS2 in predisposing to colorectal cancer. Cancer Res 2001;61:7798-802.
52. Worthley DL, Walsh MD, Barker M, et al. Familial mutations in PMS2 can cause autosomal dominant hereditary nonpolyposis colorectal cancer. Gastroenterology 2005;128:1431-6.

TABLES

Table 1: Clinical and molecular characteristics of probands with tumours showing loss of PMS2 expression. Abbreviations: BC, Bethesda criteria; AC, Amsterdam criteria; MSI-T, tumor with displaying suggestive MSI histology; M, male; F, female; CRC, colorectal cancer; EC, endometrial cancer; SBC, small bowel cancer; MSI, microsatellite instability: +, instable; -, stable; IHC, immunohistochemical analysis of MMR proteins in tumor tissue: + expression; - loss of expression; NP, not performed; WT, wild-type; hm, homozygous DNA change. ^Only MLPA analysis has been performed. *No *MSH6* mutations were identified. #Individual harboring an additional silent *APC* VUS (c.6525A>G).

Family ID	Proband ID	Clinical Criteria	Gender	Type of Tumor	Age of onset	MSI	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	Pathogenic *PMS2* mutation	*PMS2* VUS	*PMS2* Polymorphism
A	**A1**	MSI-T	F	EC	42	NP	+	+	+	-	c.1A>G (p.?)		c.705+17A>G hm; c.780C>G hm (p.=); c.1408C>T hm (p.=); c.1621G>A hm (p.E541K); c.*92dupA
B	**B1**	MSI-T	M	CRC	66	+	+	+	+	-	c.164-2A>G	c.384G>A (p.=)	c.*92dupA
C	**C1**	BC	M	CRC	45	+	+	+	+	-	c.538-478_705+456del		c.705+17A>G hm; c.780C>G hm (p.=); c.1408C>T hm (p.=); c.1621G>A hm (p.E541K); c.2007-4G>A
D	**D1**	BC	F	CRC	48	+	+	+	+	-	c.780delC (p.N261Mfs*46)		c.705+17A>G; c.1408C>T (p.=); c.1621A>G hm (p.E541K)
E	**E1**	AC	M	CRC	49	+	+	+	+	-	c.943C>T (p.R315*)		c.705+17A>G; c.780C>G hm (p.=); c.1408C>T (p.=); c.1621G>A hm (p.E541K); c.2007-4G>A, c.2466T>C (p.=); c.*92dup A hm
F	**F1^**	MSI-T	M	SBC	54	+	+	+	+	-	c.989-?_(*175_?)del		
G	**G1**	BC	M	CRC	72	+	+	+	+	-	c.989-2A>G		c.705+17A>G hm; c.780C>G hm (p.=); c.1408C>T hm (p.=); c.1621G>A hm (p.E541K); c.2007-4G>A
H	**H1**	BC	M	CRC	59	+	+	+	+	-		c.59G>A (p.R20Q)	c.780C>G hm (p.=); c.1621G>A hm (p.E541K); c.2007-7C>T; c.2253T>C (p.=); c.2324A>G (p.N775S); c.2340C>T (p.=); c.2466 T>C (p.=); c.*17G>C; c.92*dupA hm
I	**I1**	BC	M	CRC	66	+	+	+	+	-		c.137G>T (p.S46I)	c.705+17A>G; c.780C>G (p.=); c.1408C>T (p.=); c.1621G>A (p.E541K)
J	**J1***	MSI-T	F	SBC	67	NP	+	+	-	-		c.[538-3C>G;1789A>T] (p.[?;T597S])	c.705+17A>G; c.780C>G hm (p.=); c.1408C>T (p.=); c.1621G>A hm (p.E541K); c.*92dupA
K	**K1#**	BC+ Polyposis	M	CRC+ Polyposis	52	+	+	+	+	-		c.1569C>G (p.=)	c.705+17A>G; c.780C>G hm (p.=); c.1408C>T (p.=);c.1621G>A (p.E541K); c.*92dupA
L	**L1**	BC	M	CRC	47	+	+	+	+	-			c.705+17A>G hm; c.780C>G hm (p.=); c.1408C>T hm (p.=); c.1621G>A hm (p.E541K)
M	**M1**	BC	M	CRC	42	NP	+	+	+	-			c.705+17A>G; c.780C>G (p.=); c.1408C>T (p.=); c.1621G>A hm (p.E541K)

Table 2: Allelic frequency of *PMS2* variants in control individuals and patients. Results in bold indicate frequencies <0.01.

PMS2 variant	Exon	Allelic frequency in patients with PMS2- tumors (n=13 individuals) (95% CI)	Allelic frequency in controls (n=93 individuals) (95% CI)
c.59G>A; p.R20Q	2	0.0385 (0.002-0.2159)	0.0269 (0.0099-0.065)
c.137G>T; p.S46I	2	0.0385 (0.002-0.2159)	**0 (0-0.0252)**
c.384G>A; p.=	5	0.0385 (0.002-0.2159)	**0 (0-0.0252)**
c.538-3C>G	6	0.0385 (0.002-0.2159)	**0 (0-0.0252)**
c.705+17A>G	6	0.5385 (0.3375-0.7286)	0.4086 (0.3379-0.4831)
c.780C>G; p.=	7	0.6923 (0.481-0.8491)	0.8172 (0.7525-0.8685)
c.1408C>T; p.P470S	11	0.5385 (0.3375-0.7286)	0.4462 (0.374-0.5207)
c.1569C>G; p.=	11	0.0385 (0.002-0.2159)	**0 (0-0.0252)**
c.1621G>A; p.E541K	11	0.7692 (0.5591-0.9025)	0.1452 (0.0995-0.206)
c.1789A>T; p.T597S	11	0.0385 (0.002-0.2159)	**0 (0-0.0252)**
c.2007-4G>A	12	0.1154 (0.0303-0.3128)	0.1452 (0.0995-0.206)
c.2007-7C>T	12	0.0385 (0.002-0.2159)	0.043 (0.0201-0.086)
c.2253T>C ; p. =	13	0.0385 (0.002-0.2159)	0.1237 (0.0817-0.1818)
c.2324A>G; p.N775S	14	0.0385 (0.002-0.2159)	0.172 (0.1222-0.2357)
c.2340C>T; p.=	14	0.0385 (0.002-0.2159)	0.0323 (0.0132-0.0722)
c.2466T>C; p.=	15	0.0385 (0.002-0.2159)	0.1774 (0.1269-0.2416)
c.*17G>C	15	0.0385 (0.002-0.2159)	0.0269 (0.0099-0.065)
c.*92dupA	15	0.3077 (0.1509-0.519)	0.2849 (0.2224-0.3564)

Table 3: Summary of functional analyses at mRNA and protein levels. Bold with gray highlight major alterations. In vitro MMR activity: +++ corresponds to MMR activity >60% compared to wild type, and +, 0-30%. Expression of MLH1 and PMS2 +++ corresponds to >60% comparing to wild-type. Abbreviations: SS, splice site; ES, enhancer site; N, nuclear; NA, not available; NP, not performed; ASE, allelic specific expression. *Individual harboring an additional pathogenic mutation in *PMS2* (c.164-2A>G); #Individual harboring an additional silent *APC* VUS (c.6525A>G).

Proband ID	*PMS2* changes	Predicted Protein	mRNA *in silico* prediction (SS prediction /ES prediction)	mRNA analysis	cDNA stability analysis to the mutant allele	Classification at RNA level	Protein *in silito* prediction	MMR assay	MLH1 expression	PMS2 expression	Subcellular localization		Classification at protein level	Final classification
											MLH1	PMS2		
B1*	c.384G>A	p.=	No effect/ Inconclusive	G>A substitution only	Not decreased	Likely neutral	NA	NA	NA	NA	NA	NA	NA	**VSD***
J1	c.538-3C>G	p.?	Aberrant/NA	**r.[538_586del; 538_780del]; p.[Glu180Glnfs*5; Glu180Leu236del]**	**Decreased**	**Pathogenic**	NA	NA	NA	NA	NA	NA	NA	**Pathogenic**
	c.1789A>T	p.T597S	No effect/ Inconclusive	A>T substitution only		Likely neutral	Neutral	NP	NP	NP	NP	NP	NP	VSD"
K1#	c.1569C>G	p.=	No effect/ Inconclusive	C>G substitution only	Not decreased	Likely neutral	NA	NA	NA	NA	NA	NA	NA	Likely neutral
H1	c.59G>A	p.R20Q	Inconclusive/ Inconclusive	G>A substitution only	Not decreased	Likely neutral	Inconclusive	+++	+++	+++	N	N	Likely Neutral	Likely neutral
I1	c.137G>T	p.S46I	No effect/ Inconclusive	G>T substitution only	Not decreased	Likely neutral	Impaired	**+**	+++	+++	N	N	**Pathogenic**	**Pathogenic**

Table 4. Summary of the published evidence on the *PMS2* VUS assessed. *PMS2 compound heterozygous individual; #Variant used as a control in the present study. In MMR activity, expression and interaction assays: +++ corresponds to >60% compared to wild type, ++, 30–60% and +, 0–30%. Not available results are highlighted in grey. Abbreviations: BC, Bethesda criteria; MSI-T, tumor with displaying suggestive MSI histology; No FH, absence of family history; MSI, microsatellite instability; NP, not performed; NA, not available.

VUS	Reported functional analysis	Clinico-pathological characteristics				Control frequency (%)	mRNA analysis	Protein analysis				
		Familial criteria	MSI	PMS2 expression	Cosegregation			MMR assay	Protein expression	Subcellular localization	Protein interactions with MLH1/p73	Other assays
c.59G>A, p.R20Q	Marinovic, et al., 2008										+++ /+	Yes
	This study	BC	Yes	No	NA	2.7	G>A substitution	+++	+++	+++		No
E41A	Raschle, et al., 2002							++	+++			Yes
	Tomer, et al., 2002							++	++			No
N45A	Raschle, et al., 2002							+	+++			Yes
c.137G>T, p.S46I	Nakagawa, et al., 2004	No FH	Yes	No								No
	Auclair, et al., 2007*		Yes	No					+			No
	Van der Klift, et al., 2010		Yes	No			G>T substitution					No
	This study	BC	Yes	No	NA	0	G>T substitution	+	+++	+++		No
D70N	Raschle, et al., 2002							+	+++			Yes
	This study #							+	+++	+++		Yes
K413del	Marinovic, et al., 2008										+++/+++	Yes
T485K	Marinovic, et al., 2008										+++/+++	Yes
P511K	Yuan, et al., 2002										+/	No
K541E	Marinovic, et al., 2008										+++/+++	Yes
c.1789A>T, p.T597S	Yuan, et al., 2002										+/	No
	Marinovic, et al., 2008										+++/+++	Yes
	This study	MSI-T	NP	No	NA	0	A>T substitution					No
M622I	Yuan, et al., 2002										+/	No
	Nakagawa, et al., 2004	No FH	Yes	No		2.8			++			No
	Marinovic, et al., 2008										+++/+++	Yes
E705K	Marinovic, et al., 2008										+++/+++	Yes
	Deschenes, et al., 2007							++	+++			Yes
S815L	Van der klift, et al., 2010		Yes	No			Normal					No
c.384G>A, p.=	This study	MSI-T	Yes	No	NA	0	G>A substitution					No
c.538-3C>G	This study	MSI-T	NP	No	NA	0	r.[538_586del, 538_780del]					No
c.1569C>G, p.=	This study	BC+polyposis	Yes	No	NA	0	G>A substitution					No

Figure 1: Characterization of the *PMS2* exon 6 deletion. **A.** Agarose gel showing the LR-PCR product of one control and the carrier patient using the primers depicted in B. **B.** Schematic representation of the genomic structures of the wild-type and deleted *PMS2* alleles. F and R denote respectively the relative positions of the forward and reverse primers used for the LR-PCR analysis. **C.** Direct sequencing of the LR-PCR product in the carrier sample. **D.** Alignment of the deleted allele with intron 5 and intron 6 sequences. The red square indicates the breakpoint region.

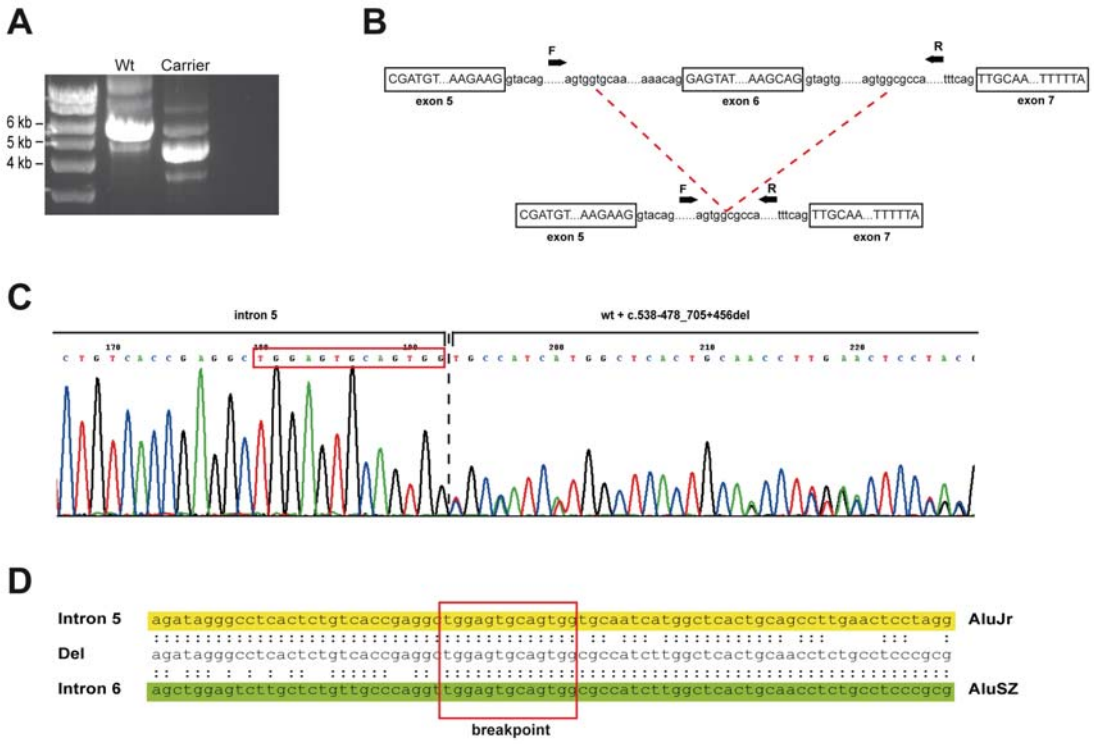


Figure 2: mRNA processing analysis. Schematic overview of *PMS2* exons and direct sequencing of the RT-PCR products in lymphocytes cultured in the absence or presence of puromycin. F and R denote respectively the relative positions of the forward and reverse primers used for direct sequencing. **A.** Characterization of the c.164-2A>G mutation. **B.** Characterization of the c.989-2A>G mutation. For both mutations, sequencing analysis showed the same pattern in the presence or absence of puromycin. **C.** Characterization of the c.538-3C>G *PMS2* variant.

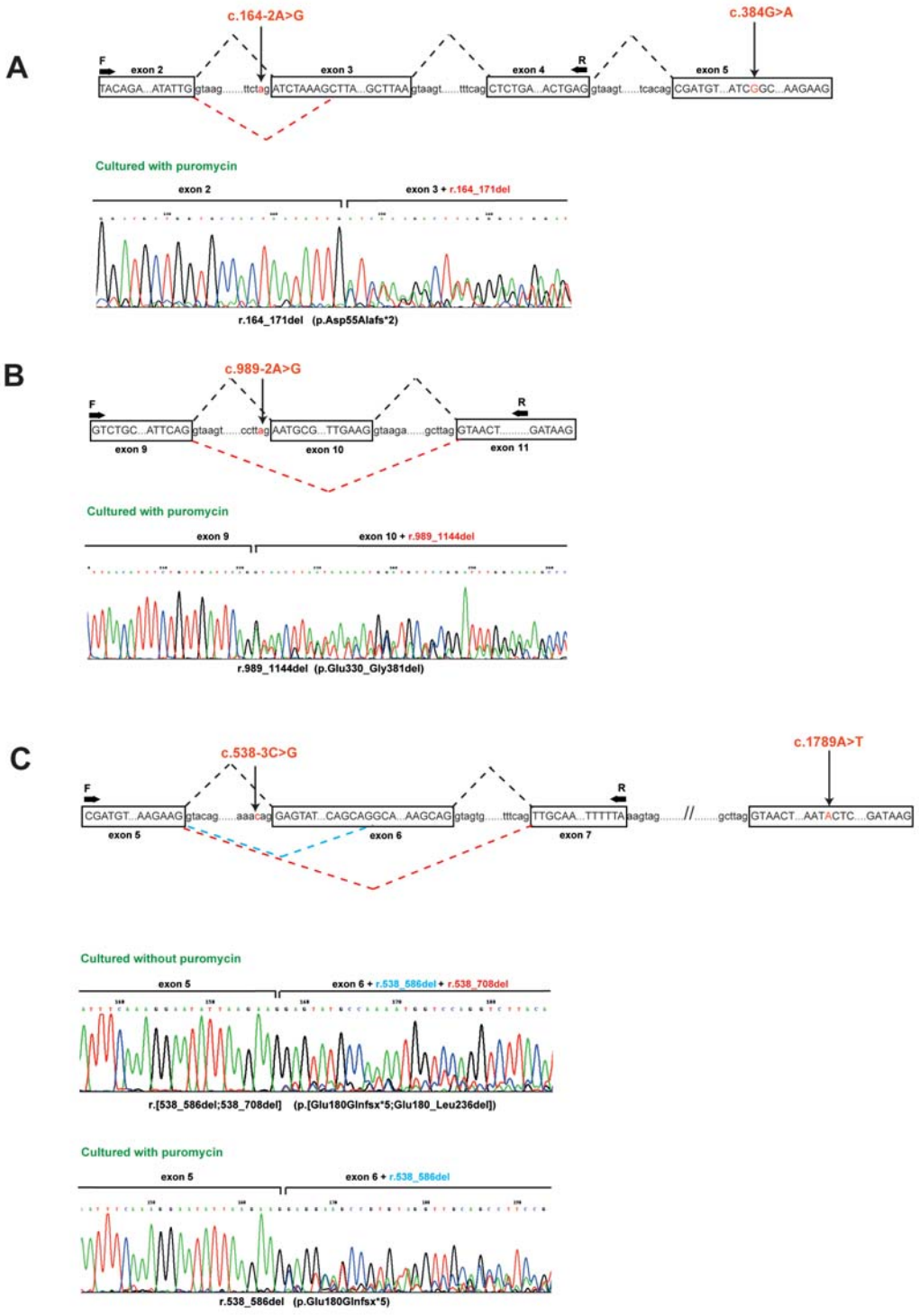
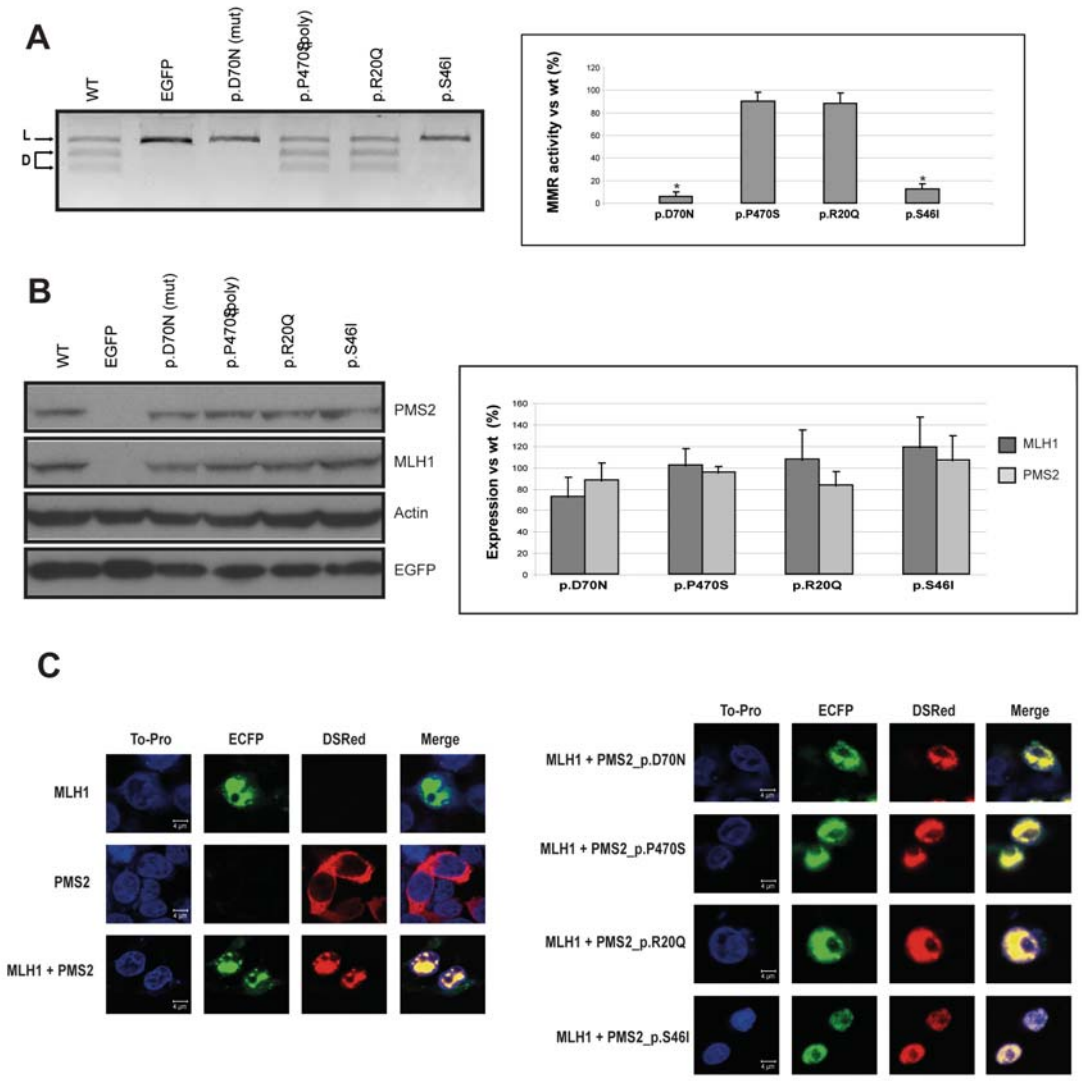
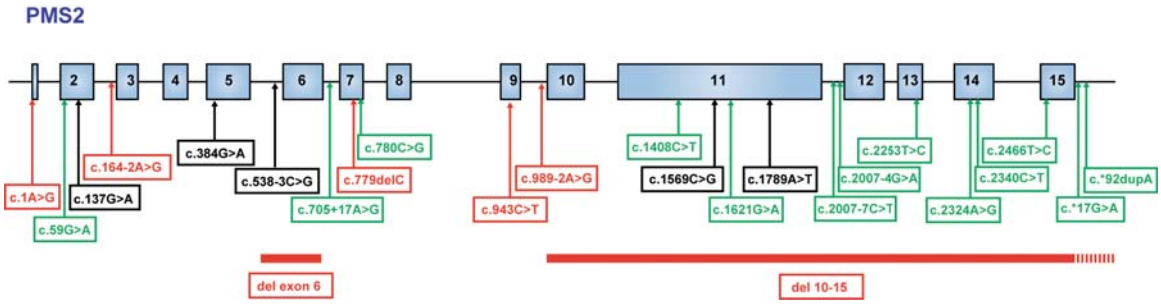


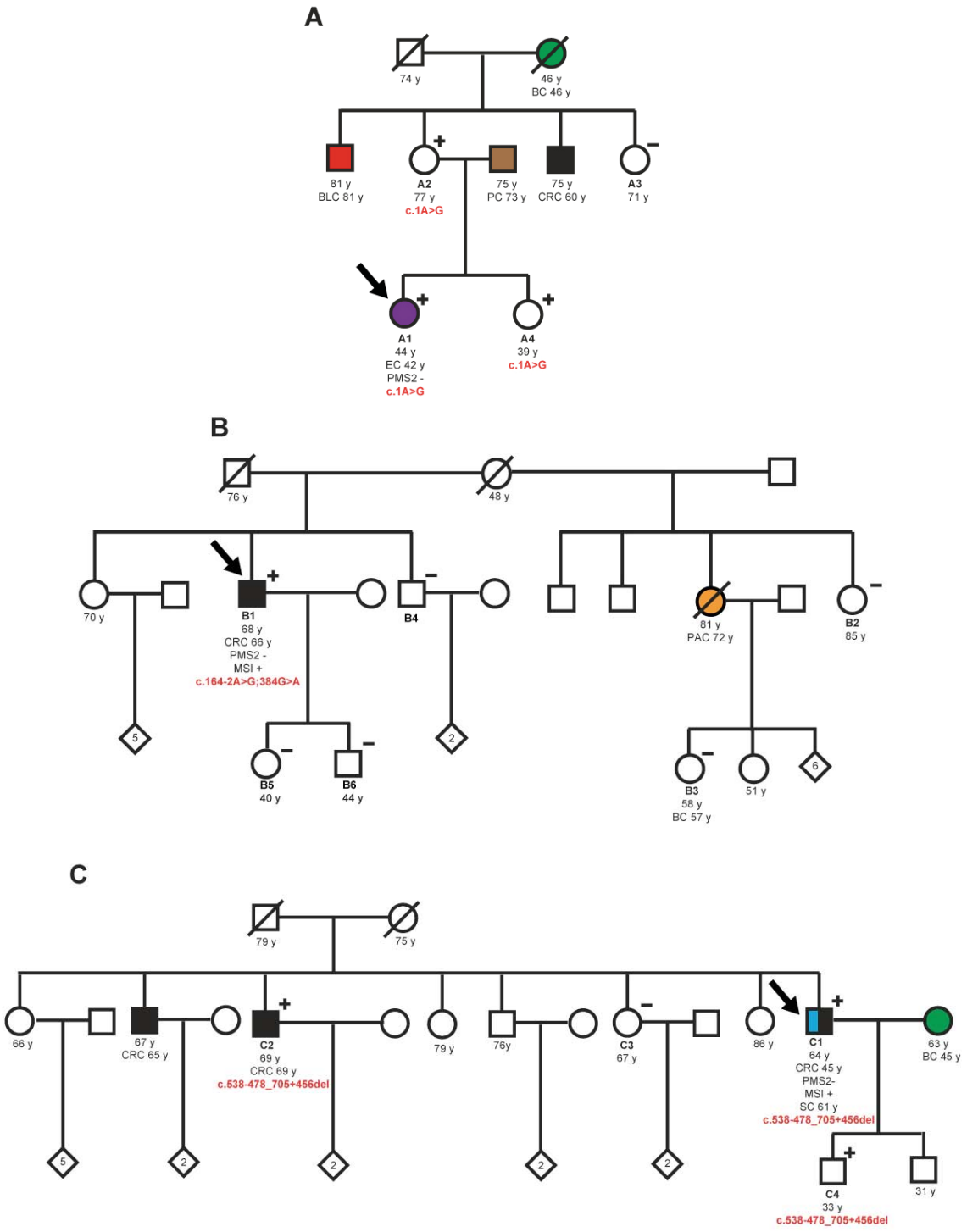
Figure 3: Mismatch repair activity, protein expression levels and subcellular localization of PMS2 variants. **A.** Agarose gel showing digestion products of MMR assay. Quantification of repair levels of *PMS2* variants in direct comparison to *PMS2* wild-type is shown. **B.** Western-blot analysis of MLH1 and PMS2. Quantification of MLH1 or PMS2 is shown in dark and light gray, respectively. **C.** Subcellular localization of PMS2 variants transfected with MLH1 in HEK293T cells. Nuclei were counterstained with TO-PRO-3 (left column). ECFP column depicts subcellular location of wildtype MLH1 protein. DSRed column depicts subcellular location of wildtype and variant PMS2 proteins. The resulting overlay is shown in the right column. Abbreviations: L, linear DNA; D, double-digested DNA.

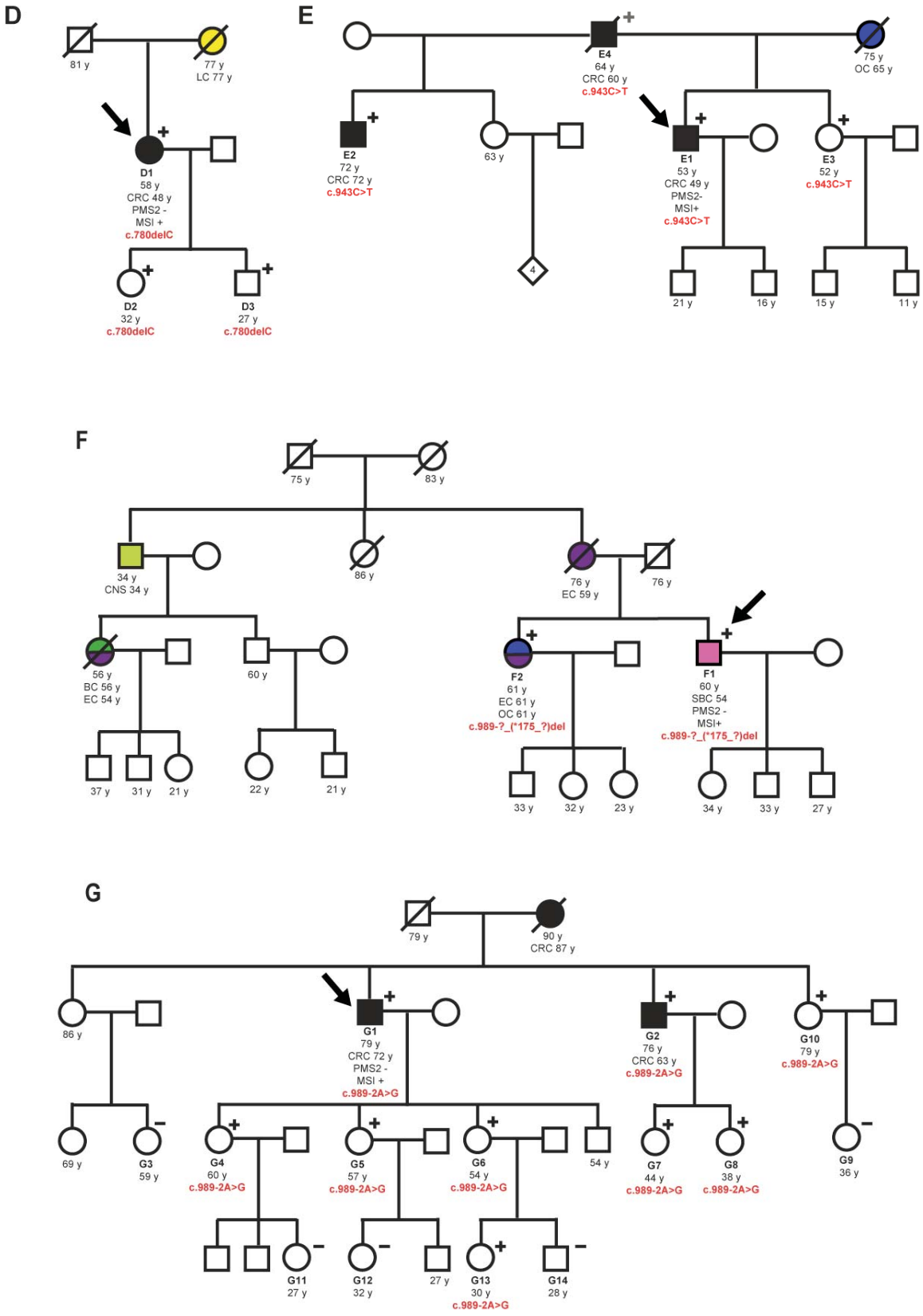


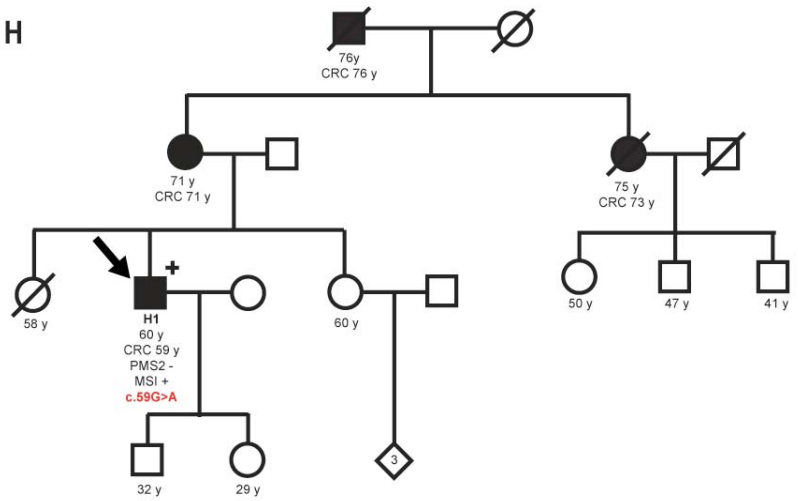
Supp. Figure 1: Graphical representation of the *PMS2* gene with DNA changes identified from this study. Red, black and green fonts represent pathogenic mutations, variants of unknown significance and polymorphisms, respectively.

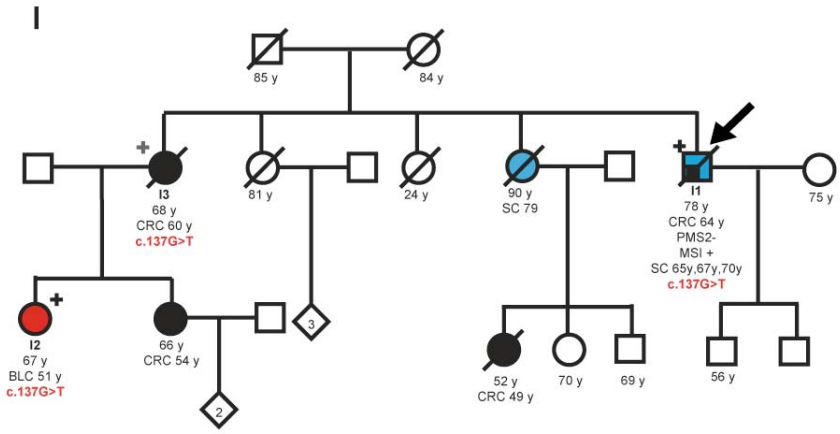


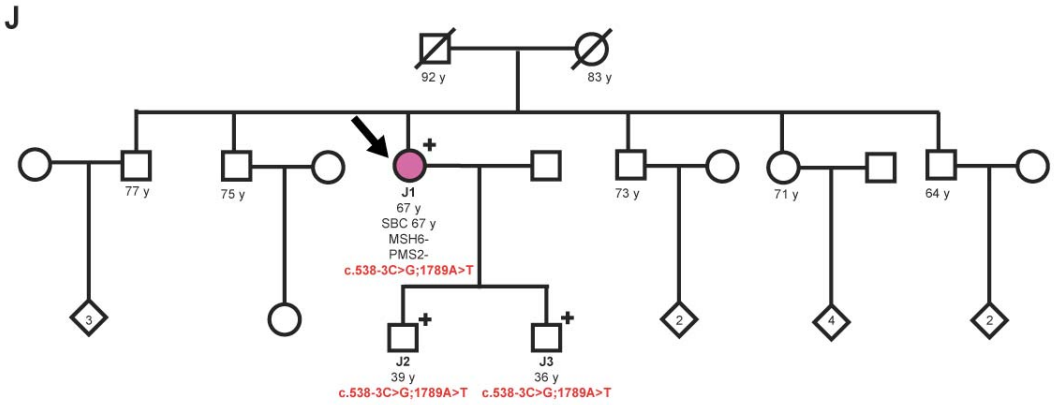
Supp. Figure 2: Pedigrees of the families included in this study. Age at diagnosis, molecular characteristics of tumors (MSI, microsatellite instability: +, instable; -, stable; IHC PMS2, immunohistochemical analysis of PMS2 protein in tumour tissue: +, expression; -, loss of expression), and mutation status are shown. Carriers are depicted by black "+", non-carriers by black "−" and obligated carries by grey "+" symbols. The probands are indicated by arrows. CRC, colorectal cancer; EC, endometrial cancer; GC, gastric cancer; BLC, bladder cancer; SC, skin cancer; LC, liver cancer; BC, breast cancer; PC, prostate cancer; OC, ovarian cancer; PAC, pancreatic cancer; SBC, small bowel cancer; CNSC, central nervous system cancer.

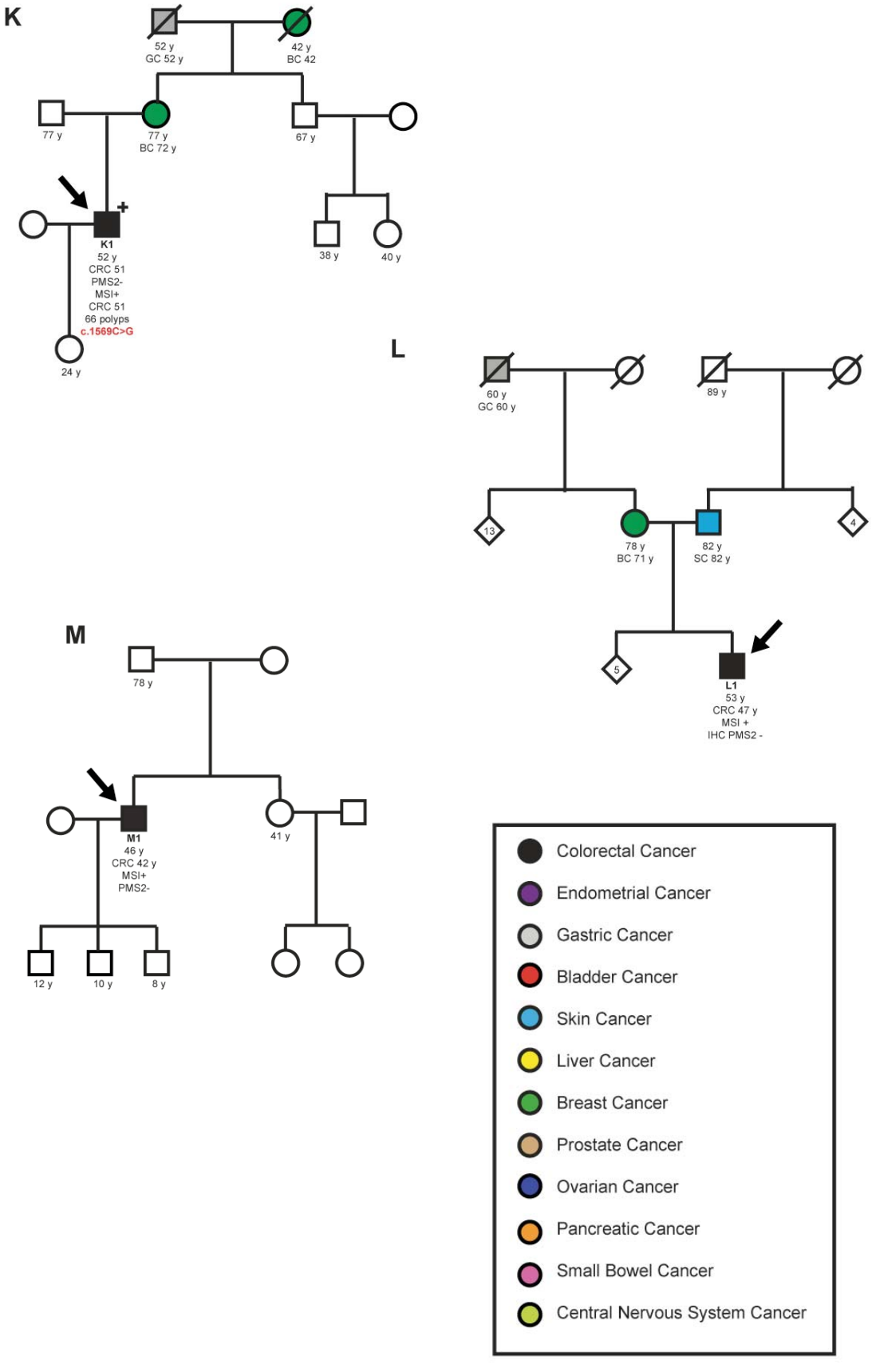



D
81 y
77 y
LC 77 y
D1
58 y
CRC 48 y
PMS2 -
MSI +
c.780delC
D2
32 y
c.780delC
D3
27 y
c.780delC
E
E4
64 y
CRC 60 y
c.943C>T
75 y
OC 65 y
E2
72 y
CRC 72 y
c.943C>T
63 y
4
E1
53 y
CRC 49 y
PMS2-
MSI+
c.943C>T
E3
52 y
c.943C>T
21 y
16 y
15 y
11 y
F
75 y
83 y
34 y
CNS 34 y
86 y
76 y
EC 59 y
56 y
BC 56 y
EC 54 y
60 y
F2
61 y
EC 61 y
OC 61 y
c.989-?_(*175_?)del
F1
60 y
SBC 54
PMS2 -
MSI+
c.989-?_(*175_?)del
37 y
31 y
21 y
22 y
21 y
33 y
32 y
23 y
34 y
33 y
27 y
G
79 y
90 y
CRC 87 y
86 y
G1
79 y
CRC 72 y
PMS2 -
MSI +
c.989-2A>G
G2
76 y
CRC 63 y
c.989-2A>G
G10
79 y
c.989-2A>G
69 y
G3
59 y
G4
60 y
c.989-2A>G
G5
57 y
c.989-2A>G
G6
54 y
c.989-2A>G
54 y
G7
44 y
c.989-2A>G
G8
38 y
c.989-2A>G
G9
36 y
G11
27 y
G12
32 y
27 y
G13
30 y
c.989-2A>G
G14
28 y

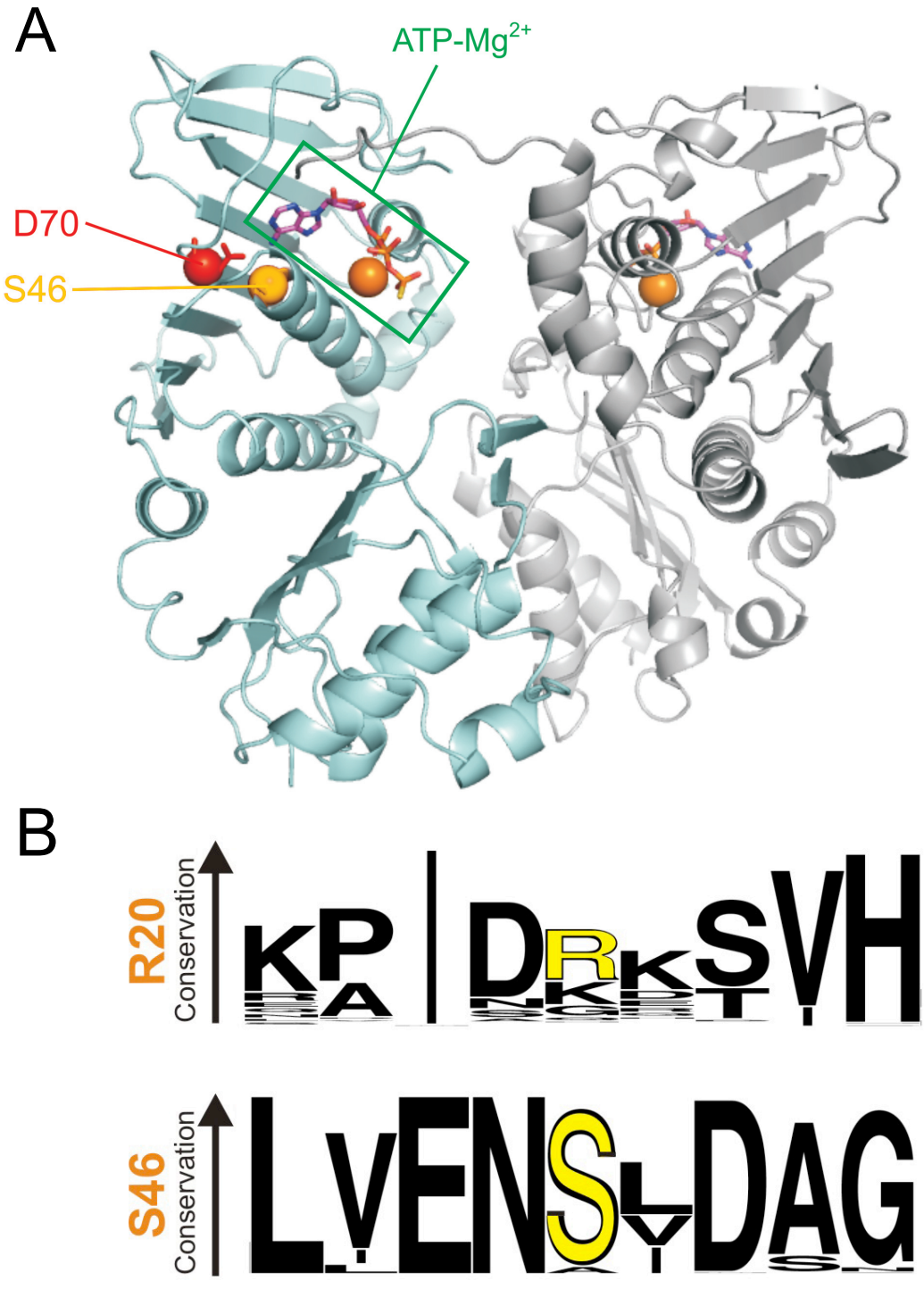

H
76y
CRC 76 y
71 y
CRC 71 y
75 y
CRC 73 y
58 y
H1
60 y
CRC 59 y
PMS2 -
MSI +
c.59G>A
60 y
50 y
47 y
41 y
32 y
29 y
3


I
85 y
84 y
I3
68 y
CRC 60 y
c.137G>T
81 y
24 y
90 y
SC 79
I1
78 y
CRC 64 y
PMS2-
MSI +
SC 65y,67y,70y
c.137G>T
75 y
I2
67 y
BLC 51 y
c.137G>T
66 y
CRC 54 y
3
2
52 y
CRC 49 y
70 y
69 y
56 y


J
92 y
83 y
77 y
75 y
J1
67 y
SBC 67 y
MSH6-
PMS2-
c.538-3C>G;1789A>T
73 y
71 y
64 y
3
2
4
2
J2
39 y
c.538-3C>G;1789A>T
J3
36 y
c.538-3C>G;1789A>T


K
52 y
GC 52 y
42 y
BC 42
77 y
77 y
BC 72 y
67 y
K1
52 y
CRC 51
PMS2-
MSI+
CRC 51
66 polyps
c.1569C>G
24 y
38 y
40 y
L
60 y
GC 60 y
89 y
13
78 y
BC 71 y
82 y
SC 82 y
4
5
L1
53 y
CRC 47 y
MSI +
IHC PMS2 -
M
78 y
M1
46 y
CRC 42 y
MSI+
PMS2-
41 y
12 y
10 y
8 y
Colorectal Cancer
Endometrial Cancer
Gastric Cancer
Bladder Cancer
Skin Cancer
Liver Cancer
Breast Cancer
Prostate Cancer
Ovarian Cancer
Pancreatic Cancer
Small Bowel Cancer
Central Nervous System Cancer

Supp. Figure 3: Localization of variants in the crystalized N-terminal domain of MLH1-PMS2 dimer. **A.** Mapping of amino acid substitutions onto the structural model of N-terminal domain of MLH1-PMS2 dimer. Substitutions occurred in the residues whose C*α* atoms are indicated as spheres, and colored according to their previous classification (red: mutation control, yellow: VUS). **B.** Conservation graphic for variants analyzed at protein level. Affected residues are indicated by yellow letters. Overall/individual font sizes indicate overall/individual conservation, respectively.



Supp. Table 1: Allelic frequency of the *PMS2* DNA changes identified in the present study in reported case and control series. *LS-suspected patients with tumors showing loss of PMS2 expression or patients with no detected germline *MLH1*, *MSH2* or *MSH6* mutations.

PMS2 VUS	Exon	Allelic frequency reported in series of LS suspected patients*	References (case series)	Allelic frequency reported in control series	References (control series)
c.59G>A; p.Arg20Gln	2	6-8%	[1-3]	7.6-11%	[1, 4], dbSNP [6, 8, 10]
c.137G>T; p.Ser46Ile	2	1-3.5%		0%	
c.384G>A; p.=	5	-		-	
c.538-3C>G	6	-	[1, 7]	-	
c.705+17A>G	6	25-43%	[1-3]	36%	dbSNP
c.780C>G; p.=	7	6.25-17%	[1, 9]	15-80%	[8], dbSNP
c.1408C>T; p.Pro470Ser	11	18.75-43%	[3]	36-48 %	[8, 9], dbSNP
c.1569C>G; p.=	11	1%	[1-3, 7, 9]		
c.1621G>A; p.Glu541Lys	11	9-16%	[1, 2]	11-80%	[8, 9], dbSNP
c.1789A>T; pThr597Ser	11	2.44-3%	[1-3]	1%	[1], dbSNP
c.2007-4G>A	12	7.3-20%	[1, 3]	-	
c.2007-7C>T	12	8-9%	[7]	-	
c.2253T>C; p. =	13	31.8%	[1, 2, 7]	39.50%	dbSNP
c.2324A>G; p.Asn775Ser	14	2.7-8.5%	[1, 2]	17%	dbSNP
c.2340C>T; p.=	14	2.4-8.5%	[1, 2]	-	
c.2466T>C; p.=	15	2.5-7.3%	[2]	2.40-31.2%	[8], dbSNP
c.*17G>C	15	6.10%	[2]	-	
c.*92dupA	15	8.50%	[1-3]	-	

Supp. Table 1 References:

1. Hendriks YM, Jagmohan-Changur S, van der Klift HM, Morreau H, van Puijenbroek M, Tops C, van Os T, Wagner A, Ausems MG, Gomez E, Breuning MH, Brocker-Vriends AH, Vasen HF, Wijnen JT. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 2006;130:312-22.
2. Clendenning M, Hampel H, LaJeunesse J, Lindblom A, Lockman J, Nilbert M, Senter L, Sotamaa K, de la Chapelle A. Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat 2006;27:490-5.
3. Niessen RC, Kleibeuker JH, Westers H, Jager PO, Rozeveld D, Bos KK, Boersma-van Ek W, Hollema H, Sijmons RH, Hofstra RM. PMS2 involvement in patients suspected of Lynch syndrome. Genes Chromosomes Cancer 2009;48:322-9.
4. Liu B, Nicolaides NC, Markowitz S, Willson JK, Parsons RE, Jen J, Papadopolous N, Peltomaki P, de la Chapelle A, Hamilton SR, et al. Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. Nat Genet 1995;9:48-55.
5. Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, Lindblom A, Lagerstedt K, Thibodeau SN, Lindor NM, Young J, Winship I, Dowty JG, White DM, Hopper JL, Baglietto L, Jenkins MA, de la Chapelle A. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 2008;135:419-28.
6. Agostini M, Tibiletti MG, Lucci-Cordisco E, Chiaravalli A, Morreau H, Furlan D, Boccuto L, Pucciarelli S, Capella C, Boiocchi M, Viel A. Two PMS2 mutations in a Turcot syndrome family with small bowel cancers. Am J Gastroenterol 2005;100:1886-91.
7. Thompson E, Meldrum CJ, Crooks R, McPhillips M, Thomas L, Spigelman AD, Scott RJ. Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations. Clin Genet 2004;65:215-25.
8. Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, Thibodeau SN, de la Chapelle A. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res 2004;64:4721-7.
9. Song H, Ramus SJ, Quaye L, DiCioccio RA, Tyrer J, Lomas E, Shadforth D, Hogdall E, Hogdall C, McGuire V, Whittemore AS, Easton DF, Ponder BA, Kjaer SK, Pharoah PD, Gayther SA. Common variants in mismatch repair genes and risk of invasive ovarian cancer. Carcinogenesis 2006;27:2235-42.
10. Auclair J, Leroux D, Desseigne F, Lasset C, Saurin JC, Joly MO, Pinson S, Xu XL, Montmain G, Ruano E, Navarro C, Puisieux A, Wang Q. Novel biallelic mutations in MSH6 and PMS2 genes: gene conversion as a likely cause of PMS2 gene inactivation. Hum Mutat 2007;28:1084-90.

Supp. Table 2. Additional variants identified in the screening of exons 2, 5, 6, 7, 9, 11, 12, 13, 14 and 15 of *PMS2* gene in our control series.

Variant	Control allelic frequency (n=93 individuals) (95% CI)
c.705+36C>T	0.0269 (0.0099-0.065)
c.706-3C>A	0.0054 (0.0003-0.0342)
c.1454C>A, p.T485K	0.0538 (0.0276-0.0995)
c.1461A>G, p.=	0.0054 (0.0003-0.0342)
c.1531A>G, p.=	0.0215 (0.0069-0.0577)
c.1561G>A, p.=	0.0054 (0.0003-0.0342)
c.1612A>G, p.K538E	0.0054 (0.0003-0.0342)
c.1688G>T, p.R563L	0.0161 (0.0042-0.0502)
c.1866G>A, p.M622I	0.0161 (0.0042-0.0502)
c.1908A>G,p.=	0.0054 (0.0003-0.0342)
c.2006+6G>A	0.0538 (0.0276-0.0995)
c.2006+19A>T	0.0054 (0.0003-0.0342)
c.2006+163_2006+164insTCA	0.129 (0.086-0.1878)
c.2149G>A, p.V717M	0.0054 (0.0003-0.0342)
c.2275+67G>A	0.0054 (0.0003-0.0342)
c.2357T>C, p.L786P	0.0054 (0.0003-0.0342)
c.2380C>T, p.P794S	0.0054 (0.0003-0.0342)
c.2445+30A>G	0.0161 (0.0042-0.0502)
c.2570G>C, p.G857A	0.0914 (0.0557-0.1446)

Supp. Table 3: Bioinformatic prediction of splicing aberrations associated with *PMS2* variants. Results highlighted in grey indicate major alterations. In splice site predictions, major alterations are destroyed or created SS or score modifications ≥45%; In enhancer site predictions, major alterations are destroyed or created ESE. Predictions are interpreted as inconclusive when the same results are not obtained by all the programs used. Abbreviations: SS, splice site; A, acceptor consensus splice site; D, donor consensus splice site; A*, new acceptor splice site; NR, consensus splice site not recognized; NA, not available; ND, not described.

			Splice Site Prediction									Enhancer site prediction				Predicted impact on protein function				
			NNSplice		Spliceport		NetGene2		Softberry											
PMS2 mutations and VUS	Exon	SS	wild-type	variant	wild-type	variant	wild-type	variant	wild-type	variant	Interpretation	Rescue ESE	ESE finder	PESX	Interpretation	Functional domain	PolyPhen-2 (score)	SIFT (score)	Condel (score)	Interpretation
c.59G>A; p.R20Q	2	A	0.45	0.45	0.47	0.05	0.44	0.34	NR	NR	Inconclusive	4 sites destroyed	2 sites created. 1 site increased	No change	Inconclusive	ATPase	Ben (0,486)	Tol (0,33)	Del (0,659)	Inconclusive
		D	1	1	0.91	0.91	0.83	0.86	0.83	0.83										
c.137G>T; p.S46I	2	A	0.45	0.45	0.45	0.45	0.44	0.44	NR	NR	No effect	1 site destroyed	1 site increased	No change	Inconclusive	ATPase	PrD (1,00)	Dam (0,00)	Del (0,998)	Impaired
		D	1	1	0.91	0.85	0.83	0.89	0.83	0.83										
c.164-2A>G	3	A	0.71	NR	2.43	NR	0.19	NR	0.89	NR	Aberrant SS	NA	NA	NA	NA	NA	NA	NA	NA	NA
		A2	NR	NR	NR	0.62	NR	NR	NR	NR										
		D	0.99	0.99	1.9	1.9	0.9	0.9	0.83	0.83										
c.384G>A; p.=	5	A	0.68	0.68	0.85	0.75	0.2	0.18	0.74	0.74	No effect	1 site destroyed	1 site destroyed	No change	Inconclusive	ATPase	NA	NA	NA	NA
		D	0.8	0.8	0.54	0.54	0.69	0.69	0.77	0.77										
c.538-3C>G	6	A	0.66	NR	1.05	NR	0.71	0.34	NR	NR	Aberrant SS	NA	NA	NA	NA	NA	NA	NA	NA	NA
		A2	NR	NR	1.08	1.29	NR	0.34	NR	NR										
		D	0.46	0.46	0.11	0.11	0.81	0.81	0.69	0.69										
c.989-2A>G	10	A	0.98	NR	1.37	NR	0.64	NR	0.72	NR	Aberrant SS	NA	NA	NA	NA	NA	NA	NA	NA	NA
		D	1	1	2.44	2.44	1	1	0.89	0.89										
c.1569C>G; p.=	11	A	0.94	0.94	1.1	1.1	0.45	0.45	0.7	0.7	No effect	No change	3 sites destroyed	No change	Inconclusive	ND	NA	NA	NA	NA
		D	0.94	0.94	1.69	1.69	0.81	0.81	0.88	0.88										
c.1789A>T p.T597S	11	A	0.94	0.94	1.1	1.1	0.45	0.45	0.7	0.7	No effect	No change	1 sites increased	No change	Inconclusive	ND	Ben (0,003)	Tol (0,74)	Neu (0,008)	Neutral
		D	0.94	0.94	1.69	1.69	0.81	0.81	0.88	0.88										

Supp. Table 4. Summary of variant allele specific expression. ASE was calculated by dividing the proportion of variant/wild-type allele in cDNA by the proportion of variant/wild-type allele in gDNA. We used ≤0.5 as a threshold value for ASE definition. Experiments were performed in triplicate. Abbreviations: varASE, variant allele specific expression; SD, standard deviation.

PMS2 variants	Puromycin Treatment	Variant allele analyzed by SNuPE	Average varASE	SD
c.59G>A	-	**c.59A**	2.00	0.08
	+		2.17	0.06
c.137G>T	-	**c.137T**	2.11	0.20
	+		2.62	0.40
c.[164-2A>G;384G>A]	-	**c.384A**	1.21	0.02
	+		1.05	0.02
c.[538-3C>G;1789A>T]	-	**c.1789T**	0.44	0.03
	+		1.99	0.01
c.1569C>G	-	**c.1569G**	1.10	0.06
	+		0.92	0.08

DISCUSSIÓ

1. ESPECTRE MUTACIONAL DELS GENS REPARADORS EN LA NOSTRA SÈRIE

Entre els anys 1998 i 2012, s'han estudiat a la Unitat de Diagnòstic Molecular els gens reparadors en línia germinal de 292 pacients amb història personal o familiar suggestiva de la síndrome de Lynch, derivats des de les Unitats de Consell Genètic de l'Institut Català d'Oncologia. Fruit d'aquest anàlisi, s'han detectat 142 famílies amb mutacions patogèniques en els gens reparadors, de les quals 74 presenten mutacions en *MLH1*, 37 en *MSH2*, 19 en *MSH6,* 9 en *PMS2* i 3 en *EPCAM* (Figura 20).

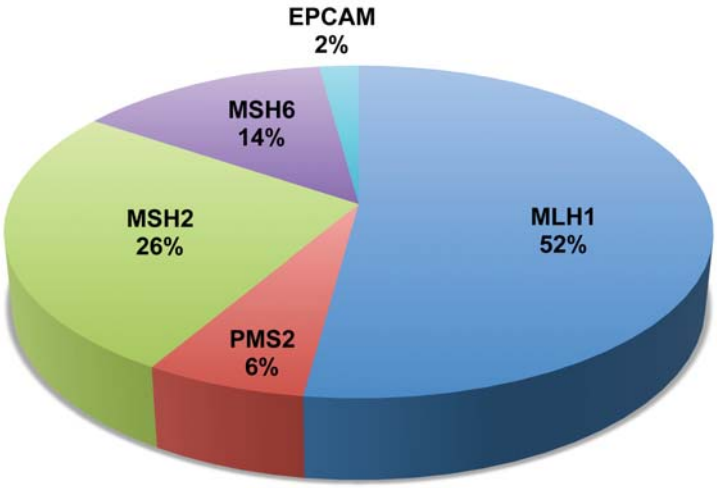


Figura 20. Distribució de mutacions en els gens reparadors identificades en pacients amb sospita de síndrome de Lynch a la unitat de diagnòstic de càncer hereditari de l'ICO.

Si restringim la nostra anàlisi a les mutacions patogèniques, un cop incorporades totes les variants classificades com a patogèniques en els articles que composen la present tesi, les mutacions en els gens *MLH1* i *MSH2* es distribueixen per tot el gen, i les grans delecions representen el 7% i el 14% de les mutacions detectades, respectivament. Per altra banda, la majoria (89%) de mutacions en el gen *MSH6* es localitzen en els exons 4 i 5 i no s'identifiquen grans reordenaments.

Pel que fa a la naturalesa molecular de les mutacions patogèniques detectades (Figura 21) observem que pels gens *MLH1, MSH2 i PMS2* n'hi ha una gran varietat. Les mutacions *nonsense* i *frameshift* són molt freqüents però hi ha una proporció important mutacions d'*splicing*. També és important remarcar l'elevat nombre de mutacions patogèniques *missense* detectades en el gen *MLH1* (algunes d'elles classificades així fruit del treball de la present tesi) i la identificació de metilació constitucional en aquest mateix gen en dos pacients (Pineda, et al., 2012). Per altra banda, en el gen *MSH6* només s'han detectat mutacions *nonsense* i *frameshift*, mentre que en el gen *EPCAM*

s'han identificat 3 portadors de grans delecions, associats a la metilació del promotor de *MSH2* (Figura 21).

Després d'analitzar les 16 VSD analitzades en aquesta tesi (10 en *MLH1,* 6 en *PMS2*) i una altra en *MSH2* en treball previ del grup (Menéndez, et al., 2010), hem aconseguit obtenir evidència experimental forta que permet classificar 13 d'elles. Si tenim en compte que les que no hem pogut classificar (n=3) i les que no hem pogut analitzar encara hi ha 27 famílies amb sospita de síndrome de Lynch que són portadores de 30 VSD: 7 en el gen *MLH1*, 12 en *MSH2,* 10 en *MSH6* i 1 en *PMS2*, algunes d'elles identificades al decurs de la present tesi doctoral. Malgrat l'esforç realitzat en caracteritzar algunes d'elles funcionalment, les VSD representen el 17,4% (30/172) de les alteracions detectades en la nostra sèrie.

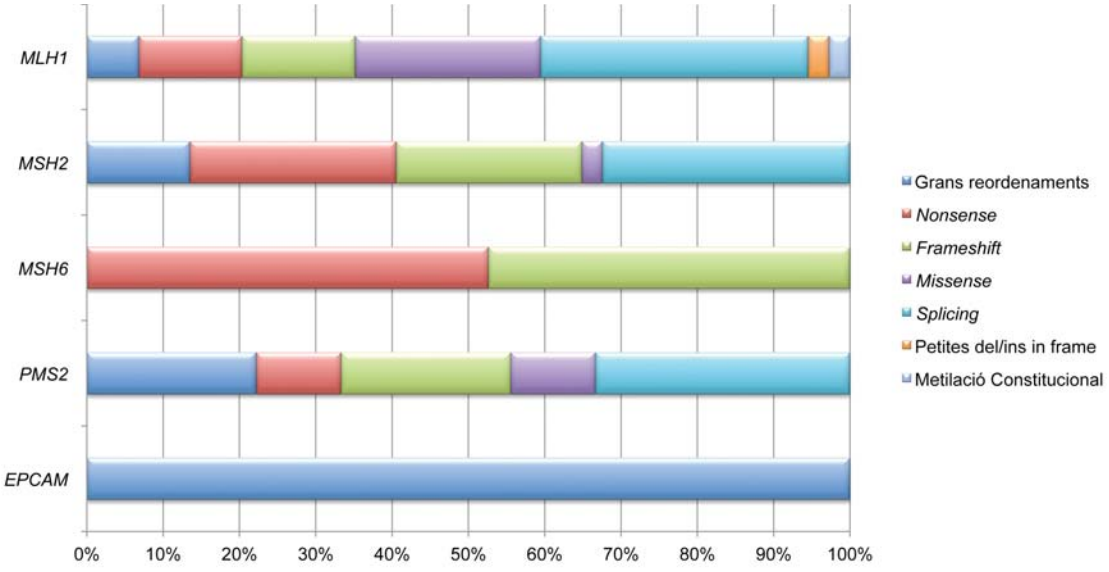


Figura 21. Espectre de mutacions patogèniques en els gens reparadors identificats a la unitat de diagnòstic de càncer hereditari de l'ICO.

L'anàlisi mutacional dels gens *MLH1, MSH2* i *MSH6* es realitza actualment mitjançant amplificació per PCR i seqüenciació automàtica. En canvi, l'anàlisi mutacional del gen *PMS2* és complex, ja que aquest presenta múltiples pseudogens que fan molt dificultós el seu anàlisi. Per aquest motiu, s'han dissenyat estratègies acurades basades en LR-PCR i encebadors específics, que ens permeten amplificar el gen però no els múltiples pseudogens (Clendenning, et al., 2006; Vaughn, et al., 2010). També s'han dissenyat sondes de MLPA capaces de distingir entre el gen i pseudogens.

Mitjançant l'ús de de LR-PCR, hem estat capaços de detectar 25 canvis identificats en pacients que presentaven tumors de l'espectre de la síndrome de Lynch i que mostraven pèrdua d'expressió de la proteïna PMS2. D'aquest canvis, 9 s'han classificats com a patogènics: 7 alteracions ho han estat per la seva naturalesa molecular i dues després de demostrar el seu defecte funcional. Dotze s'han classificat com polimorfismes, després dels estudis de freqüència, i finalment 2 canvis s'han classificat com a probablement neutrals i 2 romanen com a VSD després dels estudis funcionals (article 3, apartat de resultats).

2. MUTACIONS FUNDADORES A LA POBLACIÓ ESPANYOLA

En iniciar aquesta tesi, el primer que es va dur a terme va ser l'estudi de dues VSD identificades en el gen *MLH1,* que presentaven una elevada prevalença en la nostra sèrie. Aquestes dues VSD van ser classificades com a patogèniques en estudis funcionals i, posteriorment, es va confirmar el seu efecte fundador (article 1 de l'apartat de resultats). Així, la mutació c.306+5G>A i la c.1865T>A, identificades en el gen *MLH1*, són les dues úniques mutacions fundadores descrites en la població espanyola en aquest gen, i són originàries de la vall del Ebre i Jaén, respectivament. Cadascuna d'elles presenta un haplotip comú associat a la mutació en totes les famílies portadores, que no es detecta en la població control. A partir de l'existència d'un haplotip comú es pot estimar l'edat d'aquestes mutacions. Amb la col·laboració de professors de la Universitat de Michigan, es va estimar que l'edat de la mutació c.306+5G>A era de 75 generacions (1879 anys, assumint 25 anys per generació), mentre que per la mutació c.1865T>A era de 15 generacions (384 anys) (Figura 22).

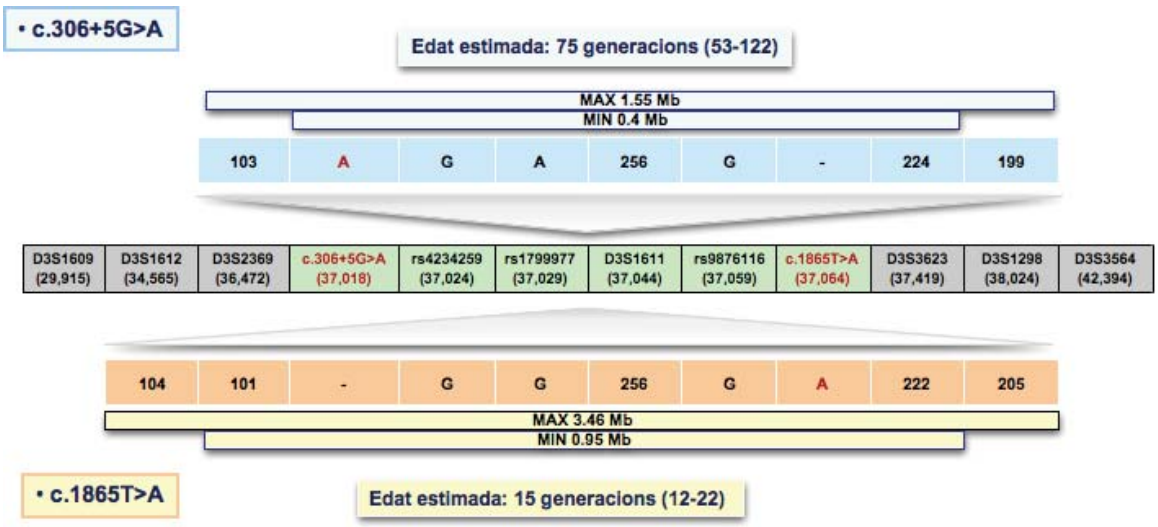


Figura 22. **Esquema resum dels resultats de l'estudi haplotip i edat de les mutacions fundadores en el gen *MLH1*.**

A la població espanyola hi ha descrites fins al moment, quatre mutacions fundadores en el gen *MSH2*, dues mutacions en el gen *MLH1* i una en *EPCAM* (Introducció, Taula 6). La mutació c.2063T>G (p.M688R) està localitzada a l'exó 13 de *MSH2*, i es tracta d'una variant *missense* que dóna lloc a un canvi d'aminoàcid no conservat en el domini ATPasa de la proteïna (Medina-Arana, et al., 2006). Aquest canvi es va detectar en cinc famílies originàries del nord de l'illa de Tenerife, i en totes elles s'ha detectat un haplotip

comú, que confirma l'efecte fundador. Aquesta va ser la primera mutació fundadora espanyola identificada en els gens reparadors.

Posteriorment al nostre grup i en col·laboració amb l'Hospital Clínic, es va detectar la mutació c.[2635-3T>C;2635-5C>T] en quatre famílies de Catalunya (Menéndez, et al., 2010). Aquesta mutació es localitza a l'intró 15 de *MSH2*, i es classificava com a variant de significat desconegut. Els estudis funcionals van permetre classificar-la com a mutació patogènica, degut a l'expressió disminuïda del mRNA de *MSH2*.

Més recentment, van ser reportades com a mutacions fundadores dues grans delecions en *MSH2*, la deleció dels exons 4 al 8 i la deleció de l'exó 7, en 4 i 3 famílies respectivament (Perez-Cabornero, et al., 2011). Els avantpassats comuns d'aquestes dues mutacions eren de la zona de Burgos, concretament de la Vall del las Navas i de Lerma. També s'ha detectat la primera mutació fundadora espanyola en el gen *EPCAM* (Mur et al; sotmès a publicació, Annex 1, Article nº7). Aquesta és una gran deleció dels exons 8-9 del gen, on les famílies portadores són originàries de l'est d'Andalusia i ha estat caracteritzada al nostre laboratori (Mur et al; sotmès a publicació, Annex 1, Article nº7).

En la Figura 23 es mostra un mapa dels orígens geogràfics de les diferents mutacions fundadores detectades a Espanya, així com les àrees on s'han identificat les famílies portadores. Tot i que les famílies van ser identificades en les províncies marcades en groc, l'origen geogràfic dels seus avantpassats, en la majoria de casos era diferent. Així, moltes de les mutacions fundadores espanyoles s'han detectat en pacients visitats al Programa de Diagnòstic de Càncer Hereditari del ICO (Borras, et al., 2010; Menéndez, et al., 2010; Perez-Cabornero, et al., 2011), Mur et al; sotmès a publicació, Annex 1, Article nº7). Això pot ser degut principalment al fet que Barcelona va ser una de les destinacions freqüents dels moviments migratoris interns del període 1960-1970.

La identificació de mutacions fundadores a Espanya recolza la idea que aquestes mutacions no només es donen en el grups comunament anomenats "poblacions fundadores", sinó que també es poden donar en subgrups geogràficament localitzats de poblacions majors. Són un clar exemple les famílies portadores de la mutació c.306+5G>A originàries de la conca del riu Ebre. El fet que la vall del riu està geogràficament aïllada per serralades, i que el riu Ebre va ser navegable fins al segle XIX, permeten hipotetitzar que la mutació va sorgir en algun lloc de la vall i es va distribuir al llarg del riu posteriorment. De forma semblant, les famílies portadores de la deleció dels exons 4-8 del gen *MSH2* també són originàries d'una vall (Vall del Navas).

Per altra banda, la distribució de la mutació c.2063C>T (p.M688R) de *MSH2* es limita a l'illa de Tenerife. Això pot ser degut al fet de que es tracta d'una població aïllada, on no hi ha hagut grans moviments migratoris cap a la península. Així, podem afirmar que l'origen d'aquestes mutacions fundadores està vinculat a àrees geogràfiques específiques, i la seva distribució actual pot explicar-se en molts casos degut als patrons de migració.

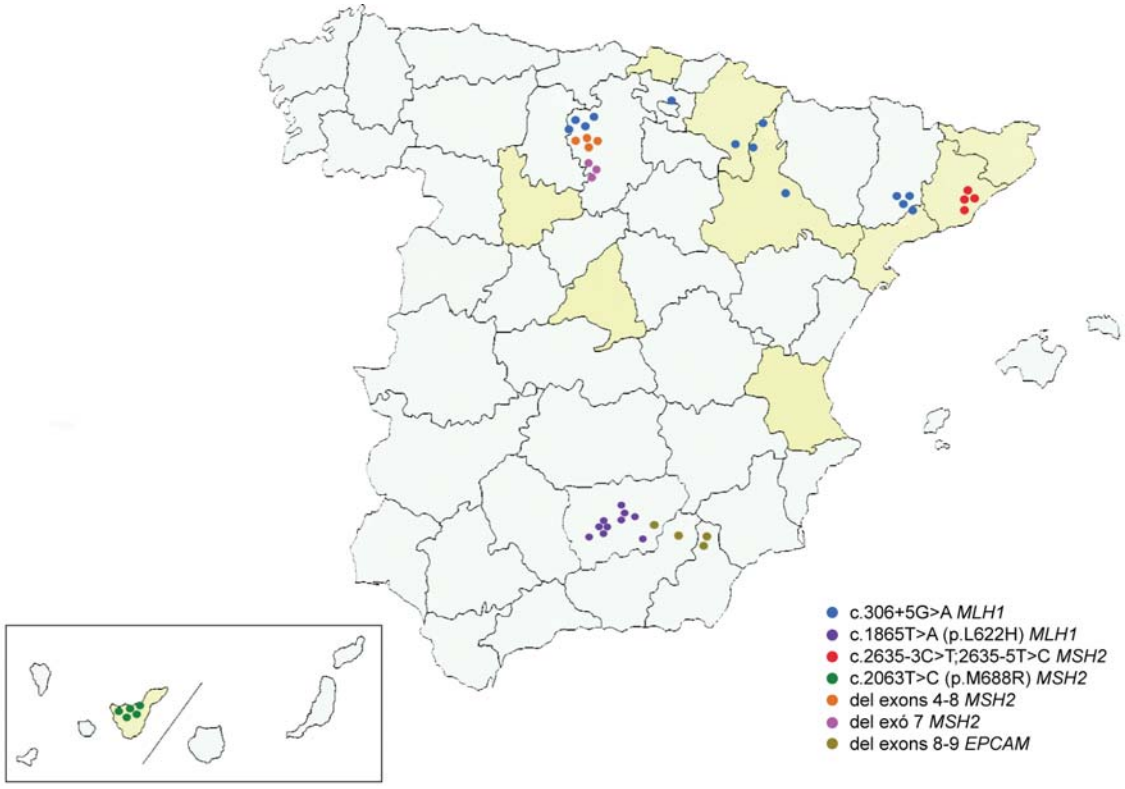


Figura 23. Localització de les famílies amb mutacions fundadores espanyoles en els gens MMR. En groc es marquen les províncies on s'han identificat els probands de les diferents famílies. Els punts representen cadascuna de les famílies identificades en la localitat de procedència de la família. Modificat de Borras, et al., 2010.

Al comparar els haplotips associats de cadascuna de les mutacions fundadores identificades a la població espanyola (Taula 9), podem observar que tenen una mida diferent per cada mutació, indicant que cadascuna d'elles es va originar en un moment diferent de l'evolució. A partir de l'existència d'un haplotip comú es pot estimar l'edat d'aquestes mutacions d'una forma més concreta. Malauradament, per la resta de mutacions fundadores espanyoles aquest anàlisi no s'ha pogut dur a terme degut al menor nombre de famílies portadores identificades. En general, com més gran és l'haplotip associat a una mutació, més recent es la formació de la mutació. Així els resultats suggereixen, que la mutació fundadora espanyola més recent seria la deleció

del exó 8-9 de *EPCAM*, mentre que la més antiga seria la mutació c.306+5G>A del gen *MLH1*.

Taula 9. Taula resum de les diferents característiques de les mutacions fundadores espanyoles. Abreviatures: ICO, Institut Català d'Oncologia; CL, Castella i Lleó; Mb, Megabases.

Gen	Mutació	Unitat Consell Genètic de Referència	Haplotip comú	Procedència	Freqüència	Reportat a
MLH1	**c.306+5G>A**	Catalunya-Madrid-Biscaia-Valladolid-Saragossa-Navarra	Max: 1,55 Mb Min: 0,44 Mb	Vall del Riu Ebre	12,1% ICO	Borràs, et al., 2010
MLH1	**c.1865T>A (p.L662H)**	Catalunya-Madrid	Max: 3,46 Mb Min: 0,95 Mb	Província Jaen	5,4% ICO	Borràs, et al., 2010
MSH2	**c.2635-3C>T;2635-5T>C**	Catalunya	Max: 4,06 Mb Min: 1,16 Mb	Catalunya	2,8% ICO	Medina-Arana, et al., 2006
MSH2	**c.2063T>C (p.M688R)**	Tenerife	1,97 Mb	Illa de Tenerife	-	Menéndez, et al., 2010
MSH2	**del exó 7**	Castella i Lleó	Max: 7,2 Mb Min: 3,6 Mb	Lerma (Burgos)	12,5% CL	Pérez-Cabornero, et al., 2011
MSH2	**del exó 4-8**	Castella i Lleó	Max: 7,2 Mb Min: 3,6 Mb	Vall del Navas (Burgos)	16% CL	Pérez-Cabornero, et al., 2011
EPCAM	**del exó 8-9**	Catalunya-València-Madrid	9,9 Mb	Andalusia	2,2 % ICO	Mur, et al., (article 7, annex1, sotmès a publ.)

2.1 Impacte de les mutacions fundadores en el diagnòstic molecular de la síndrome de Lynch

En algunes poblacions, les mutacions fundadores representen una part important del total de casos de la síndrome de Lynch. Per exemple a Finlàndia, dues mutacions en el gen *MLH1* (la deleció de l'exó 16 i la c.454-1G>A) són les responsables del 63% de totes les mutacions causants de la síndrome de Lynch (Moisio, et al., 1996; Nystrom-Lahti, et al., 1995). A la "cohort" de famílies amb síndrome de Lynch de l'ICO, les dues mutacions fundadores en el gen *MLH1* representen el 28,4% (21/74) de totes les famílies amb alteració en el gen *MLH1* i el 14,8% (21/142) de totes les famílies amb mutacions en els gens reparadors (9,9% (14/142) amb la mutació c.306+5G>A i el 4,9% (7/142) amb la c.1865T>A). A més, la mutació c.306+5G>A representa el 25% de les mutacions identificades en les Unitats de Consell Genètic localitzades dins la vall de l'Ebre (Article 1, Apartat de resultats). Si observem la resta de les mutacions fundadores aquestes també mostren una prevalença elevada, sobretot les dues grans delecions del gen *MSH2* en la sèrie de Castella i Lleó, ja que entre les dues representen un 28,5% de totes les mutacions detectades en els gens reparadors (Perez-Cabornero, et al., 2011).

Aquesta elevada prevalença ha de ser tinguda en compte en el disseny de les estratègies del diagnòstic molecular de la síndrome de Lynch en la nostra població. Així, la identificació de les mutacions fundadores a *MLH1* ha permès modificar l'estratègia en l'estudi genètic del gen *MLH1*, afegint el cribratge inicial de les dues mutacions (Figura 24).

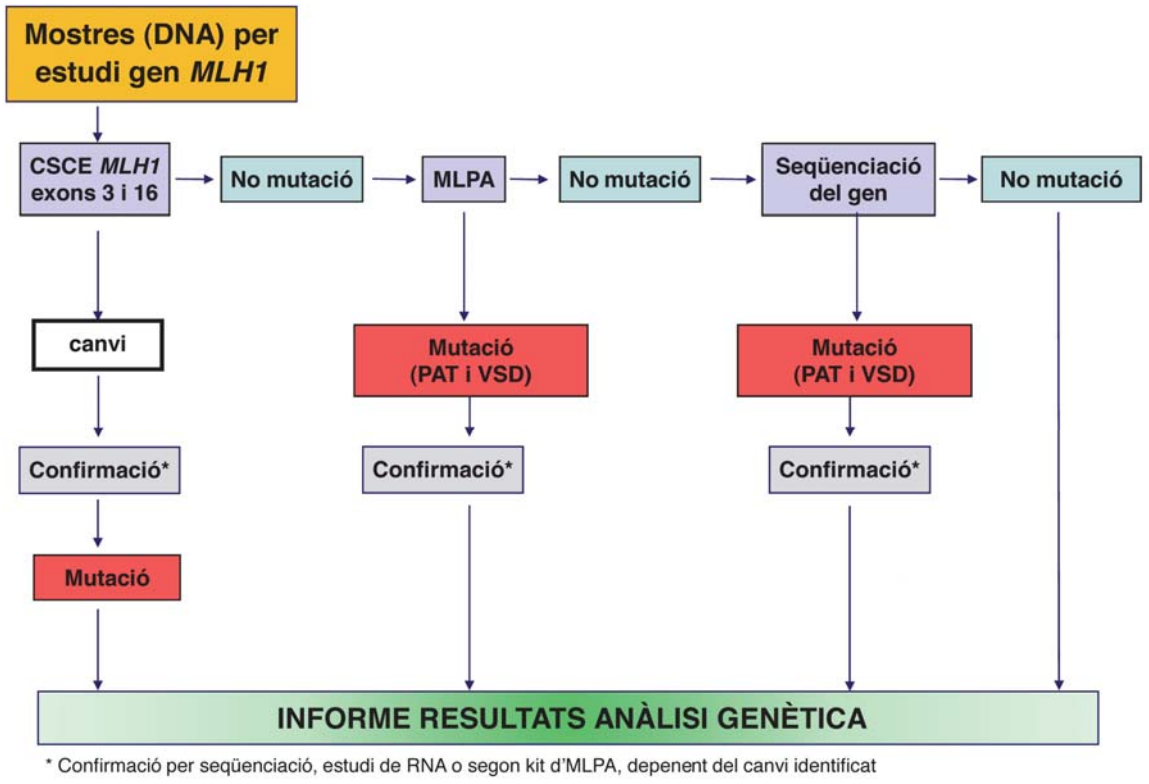


Figura 24. Algoritme per l'estudi mutacional del gen *MLH1* en el Programa de Càncer Hereditari de l'ICO. Cedit per la Unitat de Diagnòstic Molecular de l'ICO.

El cribratge de les mutacions fundadores c.1865T>A i c.306+5G>A en el gen *MLH1* es duu a terme mitjançant la tècnica del CSCE (*Conformation Sensitive Capillary Electrophoresis)* (Figura 24 i 25). Aquesta estratègia suposa un estalvi en el temps i en el cost del diagnòstic molecular de les famílies amb síndrome de Lynch. A través d'aquesta estratègia de cribratge molecular de les dues mutacions fundadores en el gen *MLH1* en el programa de diagnòstic del ICO, s'han identificat des de l'any 2010 8 noves famílies amb aquestes mutacions, 5 portadores de c.306+5G>A i 3 de c.1865T>A.

La identificació de grans reordenaments freqüents juntament amb la seva caracterització molecular, permeten dissenyar estratègies per la seva detecció que poden oferir una bona relació cost-efectivitat. Així, pel diagnòstic de la gran deleció dels exons 4-8 del gen *MSH2*, es va dissenyar una PCR diagnòstica múltiplex que ens permet la co-amplificació de l'al·lel salvatge i l'al·lel delecionat en un individu portador (Figura 25). També hem dissenyat una PCR múltiplex, pel diagnòstic de la gran deleció dels exons 8-9 del gen *EPCAM* (Mur et al; sotmès a publicació, Annex 1, Article nº7). Aquesta estratègia ens permet detectar i confirmar la presència de la gran deleció de

forma ràpida i efectiva. A més representa una millora econòmica respecte la tècnica del MLPA.

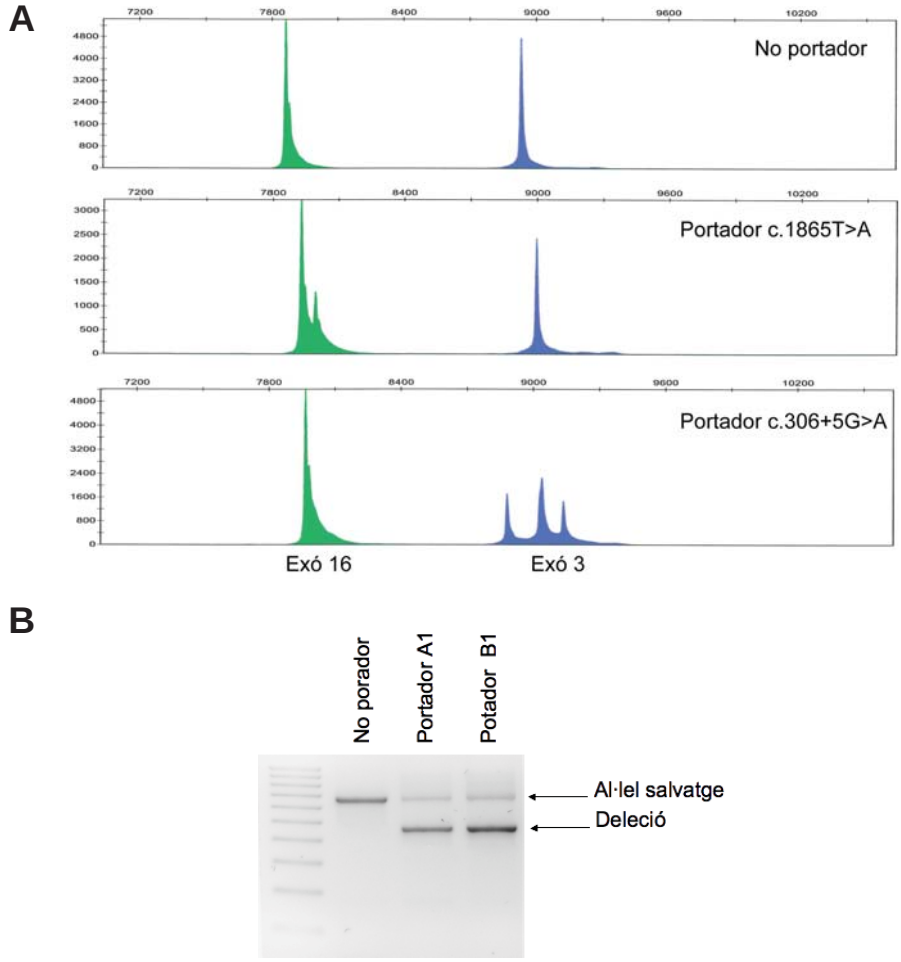


Figura 25. Cribratge de les mutacions fundadores en el gen *MLH1* i de la deleció del 4-8 en *MSH2*. A. Patrons obtinguts per CSCE per l'anàlisi de l'exó 3 (en blau) i l'exó 16 (en verd) de *MLH1* on es localitzen les mutacions c.306+5G>A i c.1865T>A, respectivament. **B.** PCR diagnòstica de la deleció fundadora que comprèn els exons 4-8 del gen *MSH2.*

2.2 Avaluació de la penetrància

En col·laboració amb la Unitat de Consell Genètic de l'ICO i membres de la Universitat de Michigan, vam explorar la penetrància de les mutacions c.306+5G>A i c.1865T>A del gen *MLH1*, ja que disposàvem d'una sèrie gran de famílies/individus portadors de la mateixa mutació i no sabíem si degut a la seva naturalesa molecular podria associar-se a una menor expressivitat. La penetrància de les mutacions c.306+5G>A i c.1865T>A va resultar moderada, dins del rang de variabilitat de la penetrància observada per les mutacions en el gen *MLH1* espanyoles (Taula 10). Si observem amb més detall aquetes dades, veurem que hi ha una tendència no significativa a una penetrància menor per la mutació *missense* c.1865T>A, un fet que es

podria atribuir a que el canvi d'un sol aminoàcid que comporta la mutació només s'associa experimentalment a una disminució clara però moderada de l'expressió de la proteïna MLH1.

Taula 10. Penetràncía observada en la síndrome de Lynch en diferents estudis. Abreviacions: ND, no disponible.

Referència	Nombre Famílies	País d'origen	Risc acumulat de CCR als 70 anys (interval de confiança del 95%)		Risc acumulat de CE als 70 anys (interval de confiança del 95%)	
			gens MMR	*MLH1*	*gens MMR*	*MLH1*
Dunlop, et al., 1997	6	Escòcia	H: 74%(ND) D: 30% (ND)		42% (ND)	
Aarino, et al., 1999	50	Finlàndia	82% (ND)		60% (ND)	
Vasen, et al., 2001	79	Països Baixos i Noruega		H: 65% (ND) D: 54% (ND)		25% (ND)
Hampel, et al., 2005	70	Finlàndia	H: 69% (59-79) D: 52% (38-67)		54% (42-66)	
Quehenberger, et al., 2005	84	Països Baixos	H: 27% (13-51) D: 22% (11-44)		32% (11-70)	
Jenkins, et al., 2006	17	Austràlia	H: 45% (29-62) D: 38% (19-51)			
Alarcon, et al., 2007	36	França	H: 47% (12-98) D: 33% (24-54)		16% (6-20)	
Barrow, et al., 2009	121	Anglaterra	H: 54% (51-58) D: 46% (43-50)		28% (25-32)	
Stoffel, et al., 2009	147	EEUU	H: 66% (59-76) D: 43% (37-53)	H: 97% (ND) D: 53% (ND)	39% (31-47)	33% (ND)
Choi, et al., 2009	32	Canada	H: 60% (35-73) D: 47% (27-60)	H: 67% (27-89) D: 35% (10-59)		
Ramsoek, et al., 2009	67	Holanda		H: 78% (ND) D: 57% (ND)		25% (ND)
Bonadona, et al., 2011	537	França	H: 38% (25-59) D: 31% (19-50)	41% (25-70)	33% (16-57)	54% (20-80)
Kempers, et al., 2011	122	Països Baixos		79% (68-90)		33% (15-51)
Borràs, et al., 2012	28	Espanya		H: 26% (8-41) D: 11% (0-17)		6% (0-13)
Borràs, et al., 2012 c.306+5G>A MLH1	17	Espanya		H: 20% (1-36) D: 14% (1-25)		7% (0-17)
Borràs, et al., 2012 c.1865T>A MLH1	12	Espanya		H: 7% (0-16) D: 7% (0-17)		3% (0-10)
Blanco, et al., (dades no publicades)	95	Espanya	H: 54% (39-65) D: 27% (18-36)	H: 62% (40-76) D: 27% (16-38)	24% (15-32)	18% (8-26)

Actualment, hi ha controvèrsia sobre la penetràncía de la síndrome de Lynch, ja que aquesta varia considerablement segons la població analitzada, el gen analitzat i el tipus d'anàlisi, entre d'altres paràmetres. A la Taula 10, es mostren els resultats de penetràncía observada en diferents estudis on s'analitza el risc de càncer associat a portadors de mutacions en els gens reparadors o en el gen *MLH1* en concret.

La penetràncía estimada de les mutacions espanyoles en el gen *MLH1* en la nostra sèrie és moderada i en consonància amb les dades reportades en famílies holandeses (Quehenberger, et al., 2005) i franceses (Bonadona, et al., 2011). Cal dir però, que la

penetràncIa estimada en les famílies espanyoles amb la síndrome de Lynch és més baixa que la reportada en la població nord-Americana, utilitzant la mateixa aproximació estadística (Choi, et al., 2009; Stoffel, et al., 2009; Blanco et al., dades no publicades) (Taula 10). Aquests resultats suggereixen possibles diferències geogràfiques en relació al risc de càncer en els portadors de la síndrome de Lynch. Això podria ser causat per diferències en factors ambientals, l'estil de vida, l'existència d'al·lels de baixa penetràncIa o degut a diferències en l'espectre mutacional de cada sèrie.

3. AVALUACIÓ DE LA PATOGENICITAT DE LES VSD DETECTADES ALS GENS *MLH1* I *PMS2*

3.1. Rendiment de la caracterització funcional de les VSD

Els resultats presentats en aquesta tesi han permès aprofundir en el coneixement de la patogenicitat de les variants de significat desconegut (VSD), identificades als gens *MLH1* i *PMS2,* en pacients amb sospita de la síndrome de Lynch estudiats a la Unitat de Diagnòstic Molecular del Programa de Càncer Hereditari de l'ICO.

L'any 2008, a l'inici de la present tesi, a la Unitat de Diagnòstic Molecular del Programa de Càncer Hereditari de l'ICO s'havien identificat 10 VSD al gen *MLH1* (Figura 26), de les quals dues (c.306+5G>A i c.1865T>A) presentaven una elevada freqüència.

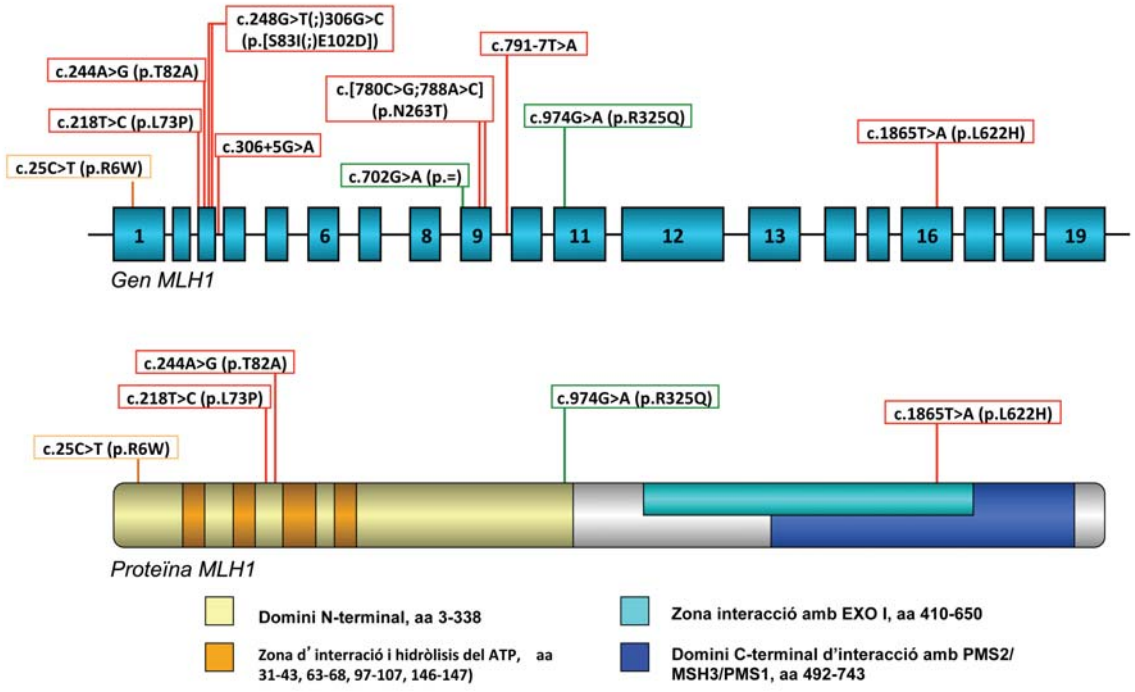


Figura 26. Localització de les VSD identificades en el gen *MLH1* en pacients amb sospita de la síndrome de Lynch. En el gen s'indiquen totes les VSD detectades, mentre que a la proteïna es localitzen les VSD analitzades funcionalment a nivell de proteïna. En vermell es marquen les VSD classificades com a patogèniques, en verd les VSD classificades com a probablement neutrals i en taronja les variants que romanen com a VSD després dels estudis funcionals.

Tal i com s'exposa en els articles 1 i 2 de l'apartat de resultats, set d'aquestes variants s'han classificat com a patogèniques: les variants c.[248G>T(;)306G>C], c.306+5G>A, c.[780C>G;788A>C] i c.791-7T>A afecten el processament correcte del mRNA, mentre que les variants c.218T>C (p.L73P), c.244A>G (p.T82A), c.1865T>A (p.L622H) mostren una activitat reparadora deficient o bé un defecte en l'expressió i/o l'estabilitat de la proteïna MLH1. Dues més s'han considerat com a probablement neutres: la variant silent c.702G>A no mostra cap efecte a nivell de processament ni estabilitat del mRNA i es localitza en *trans* amb una variant patogènica, i la variant c.974G>A (p.R325Q) no afecta ni el processament ni estabilitat del mRNA ni tampoc la funcionalitat de la proteïna. Per contra, la variant c.25C>T (p.R9W) es manté com a VSD, ja que està associada a valors intermedis d'activitat reparadora (Taula 11). Així, hem aportat evidència experimental que ha permès classificar la majoria (9/10) de les variants analitzades.

Taula 11. Classificació final de les VSD identificades en els gens *MLH1* i *PMS2* en la nostra sèrie després de la caracterització funcional. a, coexistència amb mutació c.1958+3A>G en APC (augment del *skipping* exó 14); b, coexistència amb la VSD c.1186C>G en *MSH6* (p.L396V); c, coexistència amb la mutació c.164+2A>G en *PMS2*; d, coexistència amb VSD silent c.6525A>G en *APC*.

Gen	VSD	Predicció proteïna	Anàlisi mRNA		Anàlisi proteïna		Classificació final
			predicció *in silico* a nivell de mRNA	Anàlisi del processament i estabilitat del mRNA	Predicció *in silico* a nivell de proteïna	Anàlisi proteïna	
MLH1	c.25C>T [a]	p.R9W	No efecte	**No efecte**	Inconcloent	Inconcloent	**Inconcloent**
MLH1	c.218T>C	p.L73P	Inconcloent ESE	**No efecte**	Defecte	**Defecte**	**Patogènic**
MLH1	c.244A>G [b]	p.T82A	Inconcloent ESE	**No efecte**	Defecte	**Defecte**	**Patogènic**
MLH1	c.[248G>T(;)306G>C]	p.[S83I(;)E102D]	Inconcloent SS i ESE	**Transcrit aberrant**	Defecte;Defecte	NP	**Patogènic**
MLH1	c.306+5G>A	p.?	Aberrant SS	**Transcrit aberrant**	NA	NA	**Patogènic**
MLH1	c.702G>A	p.(=)	Aberrant ESE	**No efecte**	NP	NP	**Probablement neutral**
	c.[780C>G;788A>C]	p.N263T	Inconcloent SS i ESE	**Transcrit aberrant**	Defecte	NP	**Patogènic**
MLH1	c.791-7T>A	p.?	Aberrant SS	**Transcrit aberrant**	NP	NP	**Patogènic**
MLH1	c.974G>A	p.R325Q	Inconcloent ESE	**No efecte**	Neutral	Normal	**Probablement neutral**
MLH1	c.1865T>A	p.L622H	Inconcloent ESE	**No efecte**	Defecte	**Defecte**	**Patogènic**
PMS2	c.59G>A	p.R20Q	Inconcloent SS i ESE	**No efecte**	Inconcloent	Normal	**Probablement neutral**
PMS2	c.137G>T	p.S46I	Inconcloent ESE	**No efecte**	Defecte	**Defecte**	**Patogènic**
PMS2	c.384G>A C [c]	p.=	Inconcloent ESE	**No efecte**	NA	NA	**Inconcloent**
PMS2	c.538-3C>G	p.?	Aberrant SS	**Transcrit aberrant i decreixemet de l'al·lel mutat**	NA	NA	**Patogènic**
	c.1789A>T	p.T597S	Inconcloent ESE	**No efecte**	Neutral	NP	**Inconcloent**
PMS2	c.1569G>A [d]	p.=	Inconcloent ESE	**No efecte**	NA	NA	**Probablement neutral**

Posteriorment, es va dur a terme l'anàlisi mutacional del gen *PMS2*, on es van identificar sis VSD (Figura 27). La caracterització funcional de les VSD identificades en *PMS2*, descrit en l'article 3 de l'apartat de resultats, ha permès classificar quatre de les sis variants (Taula 11). D'aquestes, dues han estat classificades com a patogèniques: la

c.538-3C>G que afecta el processament i l'estabilitat del mRNA i la c.137G>T (p.S46I) que mostra un defecte de reparació. Les dues variants restants, c.59G>A (p.R20Q) i c.1569C>G, es consideren probablement neutrals, degut a que no mostren defecte a nivell de mRNA ni proteïna. Finalment, les variants c.384G>A i c.1789A>T no s'han pogut classificar, al trobar-se en *cis* amb una mutació patogènica, fet que impedeix valorar el seu efecte individualment (Taula 11). És important destacar la moderada implicació clínica de la no classificació d'aquestes variants, ja que en aquestes famílies la mutació patogènica es considera la responsable de la malaltia.

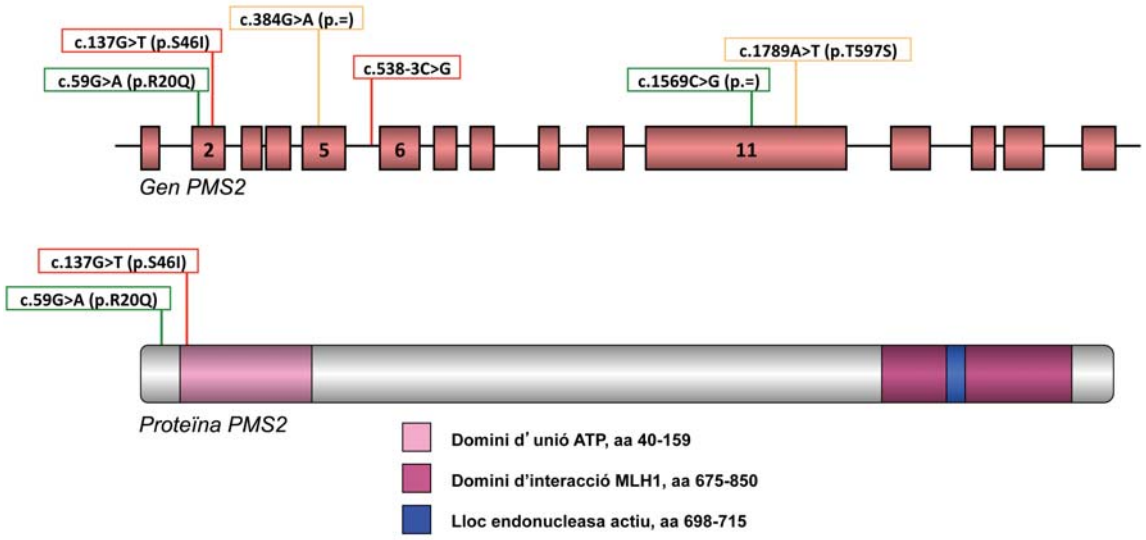


Figura 27. Localització de les VSD identificades en el *PMS2* en pacients amb sospita de la síndrome de Lynch. En el gen s'indiquen totes les VSD detectades, mentre que a la proteïna es localitzen les VSD analitzades funcionalment a nivell de proteïna. En vermell es marquen les VSD classificades com a patogèniques, en verd les VSD classificades com a probablement neutrals i en taronja les variants que romanen com a VSD després dels estudis funcionals.

Així doncs, hem pogut classificar el 81,2% (13/16) de les VSD estudiades en els gens *MLH1* i *PMS2,* després d'una caracterització funcional exhaustiva. Aquest elevat percentatge, reforça la importància dels estudis funcionals per la classificació de VSD. En el mateix sentit, resulta rellevant que el 56% (9/16) de les variants estudiades han estat classificades com a patogèniques (Figura 28). El percentatge de patogenicitat detectada es troba en línia amb d'altres sèries prèviament reportades (54-80%, 40/74-20/25)(Hardt, et al., 2011; Kansikas, et al., 2011; Kosinski, et al., 2010).

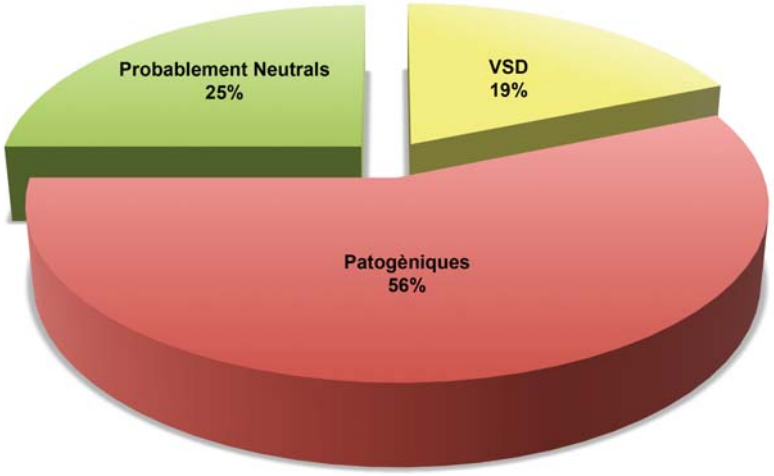


Figura 28. Classificació de les VSD dels gens *MLH1* i *PMS2* analitzades.

Al començament d'aquesta secció esmentàvem que teníem finalment identificades 74 famílies amb mutacions patogèniques en el gen *MLH1* i 9 en el gen *PMS2* (Apartat 1, discussió). D'aquestes famílies amb mutacions patogèniques, el 22% (31/142) havien estat inicialment classificades com a VSD, i després d'una exhaustiva caracterització funcional (Menéndez, et al., 2010; resultats present tesi), s'han classificat com a mutacions patogèniques (Figura 29). D'aquesta manera hem pogut augmentar el percentatge de famílies on es confirma la sospita de síndrome de Lynch, i on la informació molecular ja pot ser utilitzada pel consell genètic d'aquestes famílies.

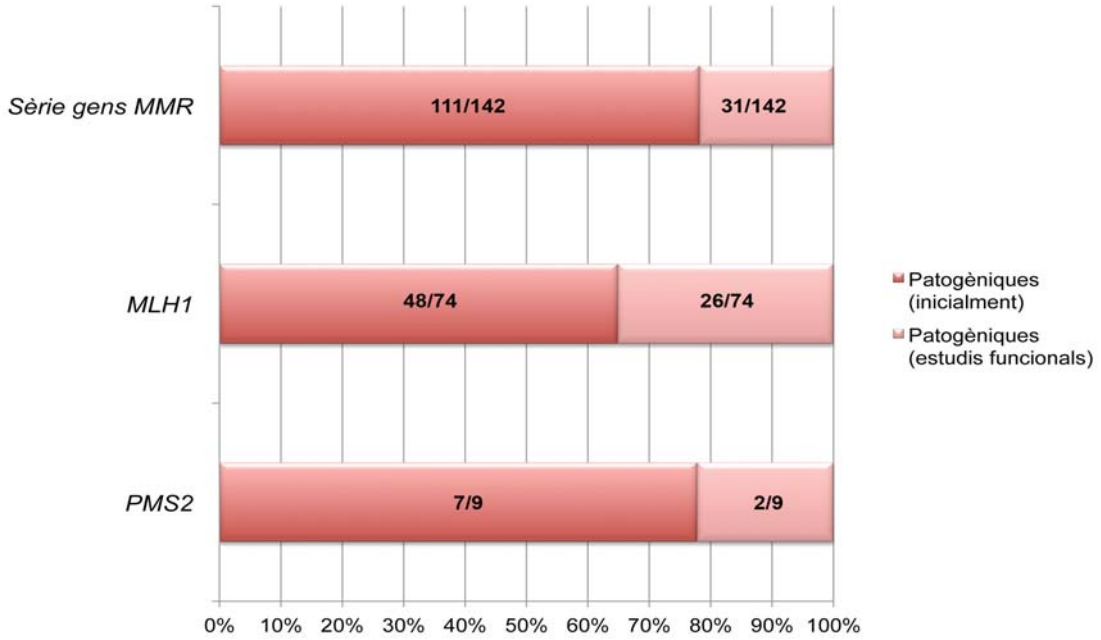


Figura 29. Gràfic on es representa el percentatge de famílies amb mutacions patogèniques classificades mitjançant estudis funcionals del total de famílies amb mutacions patogèniques detectades en els gens reparadors i en els gens *MLH1* i *PMS2*, del Programa de Càncer Hereditari de l'ICO.

3.2 Criteris de predicció de patogenicitat i assajos utilitzats

A continuació es discuteixen els diferents criteris de predicció de patogenicitat i els diferents assaigs utilitzats, que ens han permès classificar les diferents VSD estudiades en aquesta tesi.

3.2.1 Característiques clinico-patològiques

Alguns dels criteris que s'han tingut en compte en la classificació de les VSD identificades en els pacients amb sospita de síndrome de Lynch, són els criteris clínics i moleculars de selecció de pacients, la cosegregació de la variant amb la malaltia, la freqüència de la variant en una sèrie d'individus control i la coexistència amb altres variants (Couch, et al., 2008).

Criteris clínics i moleculars de selecció de pacients

Per tal de seleccionar famílies candidates de patir la síndrome de Lynch s'utilitzen els criteris clínics d'Amsterdam o Bethesda. Aquests criteris també s'utilitzen per la classificació de VSD ja que es consideren com un bon indicador de patogenicitat (Couch, et al., 2008). A la nostra sèrie, el 52% (19/37) dels individus portadors de VSD en el gen *MLH1* compleixen criteris d'Amsterdam i el 48% (18/37) restant compleixen criteris de Bethesda (Taula 3). Per contra, el 40% dels pacients amb VSD en el gen *PMS2* no compleixen criteris clínics de sospita de síndrome de Lynch. Aquests pacients havien estat diagnosticats de tumors amb característiques MSI i pèrdua d'expressió de la proteïna PMS2 (Taula 12). Cal remarcar el fet que el 44% (4/9) dels pacients amb mutacions inicialment classificades com patogèniques en el gen *PMS2*, tampoc compleixen criteris clínics de sospita de síndrome de Lynch. Sembla doncs, que aquests criteris clínics no són prou sensibles per detectar mutacions en aquest gen, com ja havia estat prèviament reportat en d'altres sèries (Clendenning, et al., 2006;Senter, et al., 2008). Aquest fet fa que aquests criteris clínics no puguin considerar-se un bon indicador de patogenicitat per les VSD localitzades en el gen *PMS2*.

Per altra banda, la inestabilitat de microsatèl·lits i la pèrdua d'expressió de les proteïnes reparadores en el tumor han estat utilitzades com a criteri de selecció per estudiar molecularment els gens reparadors en la nostra sèrie, independentment dels criteris clínics en el cas del gen *PMS2* (Taula 12). Cal destacar, que aquestes dues característiques són també uns bons criteris de classificació de VSD, ja que estan fortament relacionades amb la patogenicitat de la variant (Couch, et al., 2008).

Si observem les variants que finalment han estat classificades com a patogèniques, tots els tumors (8 de 8) dels individus portadors d'aquestes variants eren MSI-H, mentre que 8 dels 9 tumors analitzats mostraven pèrdua de la proteïna on s'havia localitzat la variant (Taula 12). Així doncs, aquests criteris semblen ser bons indicadors de patogenicitat. Cal destacar, el cas de l'individu portador de la variant c.244G>A (p.T82A) en el gen *MLH1*, el tumor del qual no mostra pèrdua d'expressió de la proteïna MLH1. Tot i que podria ser motiu de controvèrsia, nosaltres hem classificat aquesta variant com a patogènica basant-nos en la observació d'un defecte de reparació en l'assaig *in vitro.* A més, l'expressió conservada de la proteïna MLH1 en el tumor està en consonància amb els nivells normals de MLH1 observats en els estudis d'expressió en línies cel·lulars humanes transfectades amb la variant. Aquest fet, s'ha reportat també per altres variants classificades com a patogèniques en els gens reparadors, com per exemple la p.F80V en el gen *MLH1* i les p.T33P i p.E749K en el gen *MSH2* (Kansikas, et al., 2011).

Actualment, degut a la baixa sensibilitat que presenten els criteris clínics de sospita de síndrome de Lynch per la identificació d'individus portadors de mutacions en els gens *PMS2* (Clendenning, et al., 2006; Senter, et al., 2008) i *MSH6* (Sjursen, et al., 2010), s'està imposant la necessitat de realitzar un cribratge molecular mitjançant l'estudi de MSI i/o IHQ a nivell poblacional, en tots els casos diagnosticats de CCR o CE. La nostra sèrie de pacients amb mutacions identificades en *PMS2,* recolza la manca de sensibilitat dels criteris clínics per la detecció de mutacions en aquest gen. Per aquest motiu, considerem que el cribratge molecular poblacional podria ser de gran ajuda per detectar aquests pacients. A més, altres estudis han demostrat que el cribratge poblacional resulta més cost-efectiu, sobretot si es consideren els beneficis dels familiars del pacient identificat (Hampel, 2010; Hampel and de la Chapelle, 2011; Ladabaum, et al., 2011; Moreira, et al., 2012; Mvundura, et al., 2010; Perez-Carbonell, et al., 2012).

Per altra banda, és possible que en un futur no gaire llunyà, els avenços en la seqüenciació genòmica desafiïn als actuals enfocaments de proves genètiques. El desenvolupament de panells comercials de "gens del càncer" amb un cost progressivament més reduït podrien fer que l'estratègia més rendible per la detecció de la síndrome de Lynch fos l'anàlisi dels gens reparadors en línia germinal en el moment del diagnòstic del CCR (Ladabaum and Ford, 2012).

Taula 12. Característiques clinico-patològiques del individus amb VSD en els gens reparadors *MLH1* i *PMS2* estudiats en aquesta tesis. Abreviatures: CB, criteris de Betheda; CA, criteris d'Amsterdam; MSI, inestabilitat de microsatèl·lits; IHQ, inmunohistoquímica; CCR, càncer colorectal; CE, càncer d'endometri; CIP, càncer intestí prim; NA, no analitzat; NE, no avaluat; "tumor més freqüent detectat en els individus analitzats; *mitjana d'edat al diagnòstic dels individus analitzats; a, coexistènica amb mutació c.1958+3A>G en *APC* (augment del skipping exó 14); b, coexistència amb la VSD c.1186C>G en *MSH6* (p.L396V); c, coexistència amb la mutació c.164+2A>G en *PMS2*; d, coexistència amb VSD silent c.6525A>G en *APC*.

Gen	VSD	Predicció proteïna	Criteris clínics	Tipus de Tumor	Edat al diagnòstic	MSI	Expressió (IHQ)		Cosegregació	Freqüència controls (%)
							MLH1	PMS2		
MLH1	c.25C>T [a]	p.R9W	CB+poliposi	CCR	49	+	+	NA	NE	0
MLH1	c.218T>C	p.L73P	CB	CCR	36	+	NA	NA	NE	0
MLH1	c.244A>G [b]	p.T82A	CB	CCR	41	+	+	+	NE	0
MLH1	c.[248G>T(;)306G>C]	p.[S83I(;)E102D]	CB	CCR	45	+	-	NA	NE	0
MLH1	c.306+5G>A	p.?	8 CA i 9 CB	CCR"	49,5*	+	-	NA	Sí	0
MLH1	c.702G>A	p.(=)	CA	CE	54	+	-	NA	NE	0
MLH1	c.[780C>G;788A>C]	p.N263T							NE	0
MLH1	c.791-7T>A	p.?	CA	CCR	62	+	-	NA	Sí	0
MLH1	c.974G>A	p.R325Q	CB i CA	CCR / CCR	65/41	+/-	-/+	NA	Sí	0
MLH1	c.1865T>A	p.L622H	8 CA i 4 CB	CCR"	45,5*	+	-/+	NA	Sí	0
PMS2	c.59G>A	p.R20Q	CB	CCR	59	+	+	-	NE	2.7
PMS2	c.137G>T	p.S46I	CB	CCR	66	+	+	-	NE	0
PMS2	c.384G>A [c]	p.=	T-IMS	CCR	66	+	+	-	NE	0
PMS2	c.538-3C>G	p.?	T-IMS	CIP	67	NE	+	-	NE	0
PMS2	c.1789A>T	p.T597S							NE	0
PMS2	c.1569G>A [d]	p.=	CB+poliposi	CCR+poliposi	52	+	+	-	NE	0

Cosegregació de la variant amb la malaltia

La cosegregació de les VSD amb la malaltia és un criteri rellevant a tenir en compte per la seva classificació, tot i que no sempre és senzill d'aconseguir. En el nostre estudi, degut a la poca informació relacionada amb la cosegregació hem considerat que una variant cosegrega amb la malaltia si dos o més individus afectes de tumors associats a la síndrome de Lynch són portadors de la variant a estudiar. Tot i així, cal destacar que s'han desenvolupat varis mètodes i/o programes estadístics que calculen la cosegregació d'una variant amb la malaltia (Moller, et al., 2011; Thompson, et al., 2003; Zhou, et al., 2005).

De la sèrie de famílies analitzades en el present treball, en la majoria no és possible disposar d'informació sobre la cosegregació. Únicament s'ha pogut avaluar en quatre de les catorze VSD analitzades, i dues d'aquestes VSD són les mutacions fundadores identificades en el gen *MLH1*, de les que disposàvem d'un elevat nombre de famílies (Taula 3). Així doncs, en les variants identificades en una única família, la informació clínica obtinguda no ha estat suficient pels estudis de cosegregació, degut a la detecció d'aquestes variants en pedigrís petits i la dificultat d'aconseguir mostres pels estudis de

cosegregació d'una VSD en el context de recerca. A més, la penetrància incompleta de les mutacions en els gens reparadors dificulta l'avaluació de la cosegregació, ja que es poden identificar individus portadors de la variant però que no hagin desenvolupat la malaltia. Aquest fet es fa més evident en l'avaluació de les VSD en identificades en el gen *PMS2*, ja que aquests individus mostren una penetrància menor (Bonadona, et al., 2011; Kempers, et al., 2011; Senter, et al., 2008).

Freqüència de la variant en una sèrie d'individus control

L'anàlisi de freqüència en una sèrie d'individus control s'ha realitzat per tal de detectar variants polimòrfiques, que es defineixen com aquelles que presenten una freqüència superior a l'1% en la població (Goldgar, et al., 2008).Tot i així, s'ha de tenir en compte que una variant freqüent podria estar associada amb un risc moderat de desenvolupar la malaltia (Goldgar, et al., 2008).

Per les VSD detectades en el gen *MLH1*, l'estudi de freqüència es va realitzar mitjançant la tècnica de cribratge molecular CSCE, en una sèrie control d'uns 300 individus, tot i que els casos informatius finals depenien del rendiment de l'assaig concret. En canvi, pel gen *PMS2* i associat a la dificultat del seu anàlisis, aquest estudi no es va poder realitzar mitjançant una tècnica de cribratge. Aquesta dificultat i el cost de l'estudi mutacional del gen *PMS2*, van fer que la sèrie control analitzada fos relativament petita (93 individus). El tamany reduït de la sèrie del gen *PMS2* fa que els intervals de confiança no permetin descartar que algunes de variants analitzades funcionalment no siguin polimorfismes. Cal destacar que per obtenir una estimació correcta de la freqüència seria necessari utilitzar una mida mostral molt més elevada (aproximadament d'uns 1000 individus). Nogensmenys, molts cops és difícil d'aconseguir i té un cost molt elevat, tant de temps com de diners.

Els estudis de freqüència s'han dut a terme per totes les VSD detectades en el gen *MLH1* i han demostrat que es troben amb una freqüència inferior al 1% en la població control (Taula 2, article 2, apartat resultats). En canvi, l'estudi de la freqüència del gen *PMS2* es va dur a terme per totes les alteracions detectades, a excepció de les variants que van ser classificades com a patogèniques degut a la seva naturalesa molecular. D'aquesta manera es van classificar 13 dels 18 canvis com a polimorfismes, ja que presentaven freqüències superiors al 1%, mentre que 5 van seguir classificats com a VSD (Taula 2, article 3, apartat resultats). Així doncs, els estudis de freqüència ens permeten diferenciar entre els polimorfismes i les VSD. Aquest fet és de gran utilitat quan analitzes un gen que és relativament poc conegut i detectes una gran quantitat de

canvis dels quan no en coneixes el significat biològic, com seria el cas del gen *PMS2*.

Coexistència amb altres canvis

Un altre criteri a tenir en compte és la coexistència de les VSD amb altres canvis en el mateix gen o en dos gens diferents de predisposició al càncer. Aquesta coexistència representa un repte per la clínica que només pot ser abordat amb una completa caracterització dels canvis identificats.

S'ha de tenir en compte que la concurrència d'una VSD amb una mutació deletèria no és informativa si apareix en *cis*, ja que independentment de que la VSD sigui o no patogènica el risc que aporta aquesta variant en el desenvolupament del càncer pot ser camuflada per la mutació identificada com a clarament patogènica. Només es pot obtenir informació sobre la patogenicitat d'una VSD si aquesta es troba amb una mutació en *trans* en individus amb fenotip de síndrome de Lynch (no de CMMR-D), en els quals la coexistència aniria a favor de la neutralitat de la VSD.

En la nostra sèrie, s'ha identificat un cas on un individu presentava les variants c.[780C>G;788A>C] en un al·lel i la variant c.702G>A en l'altre en el gen *MLH1*. Els estudis de mRNA van demostrar patogenicitat en la combinació de les variants c.[780C>G;788A>C], i per tant es va poder classificar la variant c.702G>A com a probablement neutral, degut a la seva localització en *trans* amb les altres dues.

Per altra banda, també s'ha detectat en un individu la coexistència d'una mutació patogènica (c.164+2A>G) i una VSD (c.384G>A) en *cis*en el gen *PMS2*. El fet que estiguin en *cis* no permet classificar la VSD c.384G>A. En aquest mateix gen, s'ha detectat en un altre individu la combinació en *cis* de dues VSD, la c.1789A>T i la c.538-3C>G. Després de l'estudi funcional, la VSD c.538-3C>G s'ha classificat com a patogènica, però com que la variant c.1789A>T es localitza en *cis* no s'ha pogut classificar i roman com a VSD. En el gen *MLH1*, també s'ha detectat un pacient amb dues VSD, la c.248G>T i la c.306G>C, on es desconeix la fase. L'estudi del mRNA de limfòcits del pacient va detectar alteració en l'*splicing*, malgrat no permet concloure quina de les dues VSD és la causant de l'alteració, o si es deguda a una combinació de les dues.

Finalment, en tres ocasions s'ha detectat la presència de variants en els gens *MLH1* o *PMS2* unides a mutacions o variants en altres gens de predisposició al càncer hereditari (*APC* o *MSH6*) en un mateix individu (Taula 12). Així doncs, sembla que la

coexistència de variants en el mateix gen o en d'altres gens de predisposició al càncer es detecta freqüentment. A més, en un futur pròxim, la detecció de mutacions en la rutina diagnòstica es durà a terme mitjançant el NGS. Aquesta nova tècnica pot fomentar la detecció de múltiples VSD en diferents gens de predisposició al càncer i cal estar preparats per afrontar aquest nou repte, on els estudis funcionals ens poden ser de gran ajuda.

3.2.2 Avaluació del defecte funcional

Per tal d'avaluar el defecte a nivell funcional de les VSD identificades en individus susceptibles de patir la síndrome de Lynch, s'han utilitzat múltiples assaigs a nivell *in silico* e *in vitro*, que avaluen diferents aspectes tant a nivell de RNA com de proteïna.

Prediccions in silico

Actualment, existeixen diferents propostes sobre com utilitzar els resultats provinents de les prediccions *in silico* en la classificació de les VSD. Hi ha grups que proposen que les prediccions *in silico* són una eina vàlida per classificar les VSD (Kansikas, et al., 2011). Altres, en canvi, utilitzen aquestes mateixes prediccions com a guia pels posteriors estudis funcionals (Rasmussen, et al., 2012; Spurdle, et al., 2008; Tavtigian, et al., 2008). Ha estat feina de la present tesi establir un criteri propi en relació a les prediccions *in silico*, avaluant la correlació entre els resultats obtinguts de forma *in silico* i els experimentals, tant a nivell de RNA (*splicing* i llocs ESE) com de proteïna.

Els nostres resultats posen de manifest que dels quatre programes *in silico* utilitzats per les prediccions de *splicing*, únicament el Spliceport (Dogan, et al., 2007) ofereix una correlació fiable entre la predicció i les observacions experimentals (en 4 de les 4 variants patogèniques analitzades a nivell de mRNA) (Taula 13). Aquesta observació no avala la recomanació de la utilització simultània de varis programes de predicció *in silico* (Arnold, et al., 2009; Betz, et al., 2010). Tot i així, aquesta és una opció més conservadora, que ha de ser tinguda en compte en els algoritmes de classificació. Per contra, els resultats obtinguts mitjançant els programes de predicció dels llocs ESE no han demostrat ser útils: ni ofereixen resultats concordants entre ells ni les seves prediccions no es correlacionen amb l'evidencia experimental, confirmant les observacions fetes per altres grups (Arnold, et al., 2009; Auclair, et al., 2006; Chenevix-Trench, et al., 2006; Lastella, et al., 2006; Tournier, et al., 2008). Es desprèn que els actuals sistemes de predicció dels llocs ESE han de millorar, a fi i efecte d'obtenir una millor valor predictiu.

Taula 13. Prediccions *in silico* a nivell de RNA. En gris es marquen els canvis amb major patogenicitat. Abreviatures: SS, lloc de *splicing;* D, lloc donador; A, lloc acceptor; WT, salvatge; NR, lloc no reconegut; SD, destrucció d'un lloc ESE; SI, augment del *score* d'un lloc ESE; SC, creació d'un lloc ESE; ESS D, creació d'un lloc ESS (*exonic enhancer silencer*).

Gen	VSD	Predicció proteïna	SS	Predicció llocs *splicing*									Predicció llocs *enhancer*				Classificació final a patir dels resultas experimentals
				NNSplice		Spliceport		NetGene2		SoftBerry		Interpretació	ESEfinder	RESCUE-ESE	PESX	Interpretació	
				WT	variant	WT	variant	WT	variant	WT	variant						
MLH1	**c.25C>T**	p.R9W	D	0.93	0.93	NR	NR	0.95	0.95	7.36	7.36	No efecte	No canvi	No canvi	No canvi	No efecte	Probablement neutral
MLH1	**c.218T>C**	p.L73P	A D	0.67 1.00	0.73 1.00	0.36 0.85	0.76 0.85	0.28 0.86	0.38 0.86	6.85 13.24	6.85 13.24	No efecte	No canvi	No canvi	1SD	Inconcloent	Probablement neutral
MLH1	**c.244A>G**	p.T82A	A D	0.67 1.00	0.67 1.00	0.36 0.85	0.31 0.89	0.28 0.86	0.33 0.87	6.85 13.24	6.85 13.24	No efecte	1 SD 2 SI 1 SR	No canvi	No canvi	Inconcloent	Probablement neutral
MLH1	**c.[248G>T(;)306G>C]**	p.[S83I(;)E102D]	A D	0.67 1.00	0.67 0.93	0.36 0.85	0.55 NR	0.28 0.86	0.31 0.67	6.85 13.24	6.85 8.34	Inconcloent	1 SD 1 DI	No canvi	No canvi	Inconcloent	Patogènic
MLH1	**c.306+5G>A**		A D D2	0.67 1.00 NR	0.67 NR 0.94	0.36 0.85 NR	0.36 NR NR	0.28 0.86 0.66	0.27 0.42 0.63	6.85 13.24 10.16	6.85 10.3 10.16	SS Aberrant	NA	NA	NA	NA	Patogènic
MLH1	**c.702G>A**	p.(=)	A D	0.88 0.96	0.88 0.96	0.32 0.98	0.61 0.98	0.41 0.97	0.43 0.99	NR 12.54	NR 12.54	No efecte	3 SD	1 SD 3 SC	1 SD 3 SC	ESE Aberrant	Probablement neutral
MLH1	**c.[780C>G;788A>C]**	p.N263T	A D	0.88 0.96	0.88 0.96	0.32 0.98	0.32 0.53	0.41 0.97	0.41 0.99	nr 12.54	nr 12.40	Inconcloent	1 SC 4 SI	1 SD	No canvi	Inconcloent	Patogènic
MLH1	**c.791-7T>A**		A A* D	0.99 NR 0.93	NR 0.99 0.93	0.48 NR 0.53	NR 0.10 0.53	0.33 NR 0.44	0.16 0.16 0.41	8.75 NR 12.68	NR 9.57 12.68	SS Aberrant	NA	NA	NA	NA	Patogènic
MLH1	**c.974G>A**	p.R325Q	A D	0.86 0.98	0.86 0.98	1.13 0.67	1.13 0.67	0.95 0.94	0.95 0.94	7.47 7.64	7.47 7.64	No efecte	1 SD 2 SC	No canvi	No canvi	Inconcloent	Probablement neutral
MLH1	**c.1865T>A**	p.L622H	A D	0.99 NR	0.99 NR	0.75 0.6	0.75 0.6	0.97 0.71	0.97 0.71	0.79 NR	0.79 NR	No efecte	1 SD	1 SD	2 SC 1 ESS D	ESE Aberrant	Probablement neutral
PMS2	**c.59G>A**	p.R20Q	A D	0.45 1	0.45 1	0.47 0.91	0.05 0.91	0.44 0.83	0.34 0.86	NR 0.83	NR 0.83	Inconcloent	2 SC 1 SI	4 SD	No canvi	Inconcloent	Probablement neutral
PMS2	**c.137G>T**	p.S46I	A D	0.45 1	0.45 1	0.45 0.91	0.45 0.85	0.44 0.83	0.44 0.89	NR 0.83	NR 0.83	No efecte	1 SI	1 SD	No canvi	Inconcloent	Probablement neutral
PMS2	**c.384G>A; p.=**	p.(=)	A D	0.68 0.8	0.68 0.8	0.85 0.54	0.75 0.54	0.2 0.69	0.18 0.69	0.74 0.77	0.74 0.77	No efecte	1 SD	1 SD	No canvi	Inconcloent	Probablement neutral
PMS2	**c.538-3C>G**		A A2 D	0.66 NR 0.46	NR NR 0.46	1.05 1.08 0.11	NR 1.29 0.11	0.71 NR 0.81	0.34 0.34 0.81	NR NR 0.69	NR NR 0.69	SS Aberrant	NA	NA	NA	NA	Patogènic
PMS2	**c.1789A>T**	p.T597S	A D	0.94 0.94	0.94 0.94	1.1 1.69	1.1 1.69	0.45 0.81	0.45 0.81	0.7 0.88	0.7 0.88	No efecte	1 SD	No canvi	No canvi	Inconcloent	Probablement neutral
PMS2	**c.1569C>G**	p.(=)	A D	0.94 0.94	0.94 0.94	1.1 1.69	1.1 1.69	0.45 0.81	0.45 0.81	0.7 0.88	0.7 0.88	No efecte	3 SD	No canvi	No canvi	Inconcloent	Probablement neutral

Les prediccions *in silico* a nivell de proteïna mostren una forta correlació amb els resultats experimentals (Taula 14). Quan aquests programes han ofert prediccions concordants, els assajos funcionals també han anat en la mateixa direcció (6 de 6 de les variants analitzades). Aquestes observacions avalen la rellevància que Kaniskas i col·laboradors donen als programes de predicció *in silico* a nivell de proteïna (Kansikas, et al., 2011).

Taula 14. Prediccions *in silico* a nivell de proteïna. En gris es marquen els canvis amb major patogenicitat. Abreviatures: PrD, probablement patogènic; Ben, benigna, Aff, la proteïna es veu afectada; Tol, variant tolerant, Bor, resultat *borderline*, Del, deletèria.

Gen	VSD	Predicció proteïna	Impacte previst sobre la funció de la proteïna					Interpretació	Classificació final a partir dels resultats experimentals
			Domini Funcional	PolyPhen-2 (score)	SIFT (score)	MAPP-MMR (score)	Condel (score)		
MLH1	c.25C>T	p.R9W	ATPasa	PrD (0,97)	Aff (0,01)	Bor (4,48)	NA	Inconcloent	Inconcloent
MLH1	c.218T>C	p.L73P	ATPasa	PrD (0,94)	Aff (0,00)	Del (17,150)	NA	Patogènic	Patogènic
MLH1	c.244A>G	p.T82A	ATPasa	PrD (0,99)	Aff (0,00)	Del (36,720)	NA	Patogènic	Patogènic
MLH1	c.974G>A	p.R325Q	Transducer Domain	Ben (0,07)	Tol (0,53)	Neu (2,040)	NA	Neutral	Probablement netural
MLH1	c.1865T>A	p.L622H	Transducer Domain	PrD (0,98)	Aff (0,00)	Del (12,920)	NA	Patogènic	Patogènic
PMS2	c.59G>A	p.R20Q	ATPasa	Ben (0,486)	Tol (0,33)	NA	Del (0,659)	Inconcloent	Probablement netural
PMS2	c.137G>T	p.S46I	ATPasa	PrD (1,00)	Aff (0,00)	NA	Del (0,998)	Patogènic	Patogènic

Afegir finalment, que existeixen programes bioinformàtics comel PyMOL (http://www.pymol.org) utilitzats per localitzar la variant objecte d'estudi en l'estructura de la proteïna MLH1 o PMS2, per homologia amb MutL d'*E.Coli*, que han estat prèviament cristal·litzades (Guarne, et al., 2004; Plotz, et al., 2006). També s'utilitzen altres programes que permeten mostrar el grau de conservació dels diferents aminoàcids de la proteïna al llarg de l'evolució, com WebLogo (Crooks, et al., 2004). El coneixement de la localització i conservació d'un residu afavoreix la comprensió de les causes de la seva patogenicitat. Així per exemple, les variants p.T82A i p.L73P del gen *MLH1* i p.S46I del gen *PMS2*, mostren un defecte en el sistema de reparació degut probablement a que es localitzen en una regió altament conservada de la proteïna dins el seu domini ATPasa.

Anàlisi a nivell de RNA

En diversos estudis s'han classificat com a probablement patogèniques, una proporció significativa de VSD identificades en els gens reparadors a causa de defectes en el processament del mRNA (Arnold, et al., 2009; Sharp, et al., 2004; Tournier, et al., 2008). D'altra banda, algunes substitucions *missense* detectades en el gen *MLH1* i en altres gens implicats en el CCR, s'han associat a diferències d'expressió a nivell de mRNA (Castellsague, et al., 2010; Perera, et al., 2010; Santibanez Koref, et al., 2010).

Degut a que aquests estudis de processament i estabilitat del mRNA representen un punt clau en la classificació de variants, es va realitzar l'anàlisi del RNA de totes les variants identificades en el present estudi.

Els assaigs per avaluar el processament i l'estabilitat del mRNA normalment es solen realitzar a partir de limfòcits de pacients, ja que aquesta és la mesura més directa d'aquest anàlisi. Tot i així, per alguns laboratoris la obtenció de RNA de la sang dels pacients implica una elevada dificultat, ja que és necessari processar aquesta sang immediatament, o bé emmagatzemar-la en uns tubs adequats. Per altra banda, si no es disposa d'aquests limfòcits hi ha descrits varis assaigs *in vitro* per analitzar el processament del mRNA (Betz, et al., 2010; Lastella, et al., 2006; Naruse, et al., 2009; Tournier, et al., 2008).

L'extracció del RNA es pot dur a terme directament dels limfòcits de la sang (frescos o emmagatzemats) o bé es pot fer un cultiu curt d'aquests limfòcits, que permetrà obtenir-ne un major quantitat. A més, aquest cultiu es pot dur a terme en presència de puromicina, per tal d'evitar el mecanisme del *Nonsense Mediated Decay* (NMD). El NMD és un mecanisme cel·lular de vigilància, que assegura la degradació d'aquells mRNA que podrien donar lloc a proteïnes truncades, amb un probable efecte patogènic dominant negatiu o guany de funció (Cartegni, et al., 2002). La inactivació d'aquest mecanisme permet observar els transcrits aberrants creats, que de funcionar correctament el NMD no serien detectables. Nosaltres recomanem, sempre que sigui possible, la utilització de cultius curts de limfòcits amb i sense puromicina. Així, obtindrem una visió complerta del que provoca la variant en el mRNA del pacient, ja que veurem el seu efecte real a la cèl·lula (no puromicina) i també podrem veure quins transcrits s'estan degradant a causa del mecanisme del NMD (puromicina).

En la nostra sèrie, tots els assajos a nivell de mRNA, tant d'estabilitat com de processament, s'han dut a terme a partir del RNA de limfòcits de pacients, ja que ofereixen una mesura més fiable i nosaltres disposem d'un banc de limfòcits propi, per aquest tipus d'anàlisi. A més, aquest assajos presenten més facilitat tècnica en relació amb els assajos *in vitro*.

- **Estudi del processament del mRNA:** L'estudi del processament del mRNA es va realitzar a partir d'un disseny específic per avaluar l'efecte de cada VSD en l'*splicing* mitjançant RT-PCR en el RNA del pacient. Aquest RNA es va extreure d'un cultiu curt de limfòcits de pacients, amb i sense puromicina. Cal destacar, que moltes vegades la interpretació dels resultats dels estudis de *splicing* són complexes, ja que ens podem

trobar que en l'àrea a estudiar es produeixin varis transcrits alteratius. Per aquest motiu, considerem que és important fer un disseny acurat per tal d'evitar possibles transcrits alternatius o si fos necessari utilitzar varis dissenys per obtenir una interpretació més acurada dels resultats. D'altra banda, considerem que també és important utilitzar un nombre adequat de mostres control (entre 3-5), que ens serveixin per comparar-los amb els resultats de la variant a analitzar. En alguns casos, ens podem trobar que hi hagi un augment o disminució d'un transcrit en funció de la mostra analitzada, i llavors serà necessari ampliar aquest nombre de controls (entre 8-10), per tal de veure si aquesta modificació pot ser deguda a la variant o bé si l'expressió del transcrit és variable entre individus.

L'estudi del processament del mRNA de les variants identificades en la nostra sèrie a partir de limfòcits cultivats, ha permès la classificació com a mutacions patogèniques de cinc de les setze (31%) variants estudiades en aquesta tesi (Taula 15). Aquest alt percentatge, reforça encara més el concepte que l'anàlisi del mRNA ha de ser un dels primers passos en l'avaluació de l'impacte funcional de les VSD.

- **Anàlisi de l'expressió al·lèlica diferencial:** L'estudi d'expressió al·lèlica diferencial analitza la quantificació relativa dels dos al·lels del mRNA en comparació amb el DNA de pacient. Aquest assaig es va dur a terme a partir del RNA extret d'un cultiu curt de limfòcits de pacients, amb i sense puromicina, per les raons explicades a l'apartat anterior. Per fer aquest anàlisi s'han descrit varies tècniques diferents, com poden ser la comparació de la seqüència del RNA i DNA a partir de la utilització d'uns programes específics (Perera, et al., 2010), el *DHPLC primer extension* (Aceto, et al., 2009), o l'SNuPE (Castellsague, et al., 2010), entre d'altres. Nosaltres vam utilitzar la metodologia del SNuPE usant el kit SNaPshot per mesurar l'ASE de cada variant, ja que considerem que aquesta és una tècnica amb una elevada sensibilitat per detectar diferències d'expressió i a més té una dificultat tècnica relativament baixa. Per aquests motius és a una de les tècniques més utilitzades per aquest tipus d'anàlisi.

En la nostra sèrie, l'ASE no va mostrar diferències en 8 de les 9 VSD analitzades (Taula 15). En canvi, aquest anàlisi va permetre detectar desequilibri al·lèlic en la variant c.1789A>T del gen *PMS2*, conseqüència de la transcripció aberrant causada per la variant c.538-3C>G del gen *PMS2,* que es localitza en *cis*en el mateix individu. Així doncs, la integració dels diferents resultats obtinguts a nivell de RNA (processament i estabilitat) ens han permès la classificació de la variant c.538-3C>G com a patogènica.

Taula 15. Anàlisi a nivell de RNA de les variants estudiades en la present tesi. Els canvis que impliquen patogenicitat es marquen en gris. *Aquestes dues variants es localitzen en *cis*en el mateix individu.

Gen	VSD	Predicció proteïna	Anàlisi del processament del cDNA	Anàlisi de l'estabilitat de l'al·lel mutat en el cDNA	Classificació a nivell de RNA
MLH1	**c.25C>T**	p.R9W	canvi de C>T	NA	Probablement neutral
MLH1	**c.218T>C**	p.L73P	canvi de T>C	No decreix	Probablement neutral
MLH1	**c.244A>G**	p.T82A	canvi de G>A	No decreix	Probablement neutral
MLH1	**c.[248G>T(;)306G>C]**	p.[S83I(;)E102D]	r.208_306del; p.K70_E102del	NP	Patogènic
MLH1	**c.306+5G>A**		r.[208_306del, 208_306del]; p.[K70_E102del, K120del]	NP	Patogènic
MLH1	**c.702G>A**	p.(=)	canvi de G>A	No decreix	Probablement neutral
MLH1	**c.[780C>G;788A>C]**	p.N263T	r.678_790del; p.E227SfsX42	NP	Patogènic
MLH1	**c.791-7T>A**		r.791_884del; p.H264LfsX2	NP	Patogènic
MLH1	**c.974G>A**	p.R325Q	canvi de G>A	No decreix	Probablement neutral
MLH1	**c.1865T>A**	p.L622H	canvi de T>A	NP	Probablement neutral
PMS2	**c.59G>A**	p.R20Q	canvi de G>A	No decreix	Probablement neutral
PMS2	**c.137G>T**	p.S46I	canvi de G>T	No decreix	Probablement neutral
PMS2	**c.384G>A; p.=**	p.(=)	canvi de G>A	No decreix	Probablement neutral
PMS2	**c.538-3C>G***		r.[538_586del, 538_780del]; p.[E180Qfs*5,E180L236del]	Decreix	Patogènic
PMS2	**c.1789A>T ***	p.T597S	canvi de A>T		Probablement neutral
PMS2	**c.1569C>G**	p.(=)	canvi de C>G	No efecte	Probablement neutral

Anàlisi a nivell de proteïna

Amb l'objectiu d'analitzar l'efecte de les VSD a nivell de proteïna s'han descrit un seguit d'aproximacions que es detallen en l'apartat 6 de la introducció. Els assajos funcionals en línies cel·lulars humanes, ofereixen avantatges en front d'altres sistemes, com podrien ser els llevats. El principal avantatge és la capacitat de poder analitzar totes les VSD, independentment de la seva conservació evolutiva. Un altre avantatge rau en que podem utilitzar línies cel·lulars que presentin un *background* genètic adequat per poder testar la funcionalitat de les variants en gens reparadors. Aquests assaigs han permès classificar com a patogèniques el 25% (4/16) de les variants analitzades en aquesta tesi.

A continuació es tracten els diferents assajos utilitzats per estudiar les variants a nivell de proteïna:

- **Assaig d'activitat reparadora *in vitro*:** Els assajos funcionals d'activitat reparadora *in vitro,* es consideren els més rellevants per avaluar la funcionalitat d'una

variant (Couch, et al., 2008; Heinen and Juel Rasmussen, 2012; Kansikas, et al., 2011; Rasmussen, et al., 2012), ja que representen l'avaluació de la funció principal de la proteïna reparadora. No obstant, la preparació dels reactius és un procés tediós, l'assaig requereix d'especialització tècnica, i es necessària la construcció de mutants i la seva expressió mitjançant transfecció. També cal tenir en compte, la importància dels controls interns utilitzats, ja que ens serviran de referència per comprovar el correcte funcionament de la tècnica.

Recentment, s'ha desenvolupat una nova tècnica que en redueix parcialment el temps i la dificultat, basant-se en la creació de la proteïna mutada mitjançant PCR i posterior transcripció *in vitro,* que s'utilitzarà per a la reparació del plàsmid amb el desparellament (Drost, et al., 2010; Drost, et al., 2012). Malgrat l'agilització en el procés d'anàlisi de l'activitat reparadora, aquesta nova tècnica presenta el *handicap* de que la proteïna mutada no pot ser utilitzada posteriorment per aprofundir en l'anàlisi d'expressió, interacció o localització, per exemple. Nosaltres hem utilitzat un assaig de reparació clàssic on a partir de la proteïna amb la variant clonada en un plàsmid d'expressió transfectat en una línia cel·lular, es testa la seva activitat de reparació *in vitro* reparant un plàsmid amb un desaparellament en un lloc de restricció. L'assaig de reparació *in vitro*, ens va ajudar a classificar com a patogèniques 3 de les 6 variants analitzades (Taula 16), corroborant la utilitat de l'assaig d'activitat reparadora per la classificació de les VSD.

- **Estudis d'expressió**: Per avaluar l'expressió de les proteïnes MLH1 i PMS2, s'ha utilitzat la tècnica del *Wester-blot* en extractes proteics extrets de la transfecció transitòria d'aquestes proteïnes en línies cel·lulars deficients, que en paral·lel s'han utilitzat en assajos de reparació. Aquests estudis d'expressió mesuren l'expressió relativa de la proteïna amb la variant en relació a la proteïna salvatge. És per aquest motiu que considerem de vital importància que, en un mateix experiment, hi hagi varies rèpliques de la mostra salvatge, ja que els resultats obtinguts de cada mostra seran comparats amb aquesta. Per altra banda, també és important utilitzar un control de càrrega de les proteïnes que descarti bé problemes en la càrrega del gel o bé una mala quantificació de les proteïnes totals. A més, en l'anàlisi d'expressió per *Western-blot*, és important realitzar varies rèpliques, degut a la variabilitat interexperimental.

En aquesta tesi, tres de les set variants analitzades mostraven defectes d'expressió. D'aquestes, dues presentaven activitat reparadora reduïda, i es van classificar com a

patogèniques. La variant restant no es va poder classificar a causa dels resultats no concloents d'activitat reparadora (Taula 16).

Taula 16. Anàlisi a nivell de proteïna de les variants analitzades. Les alteracions que impliquen major patogenicitat es marquen en gris. Abreviacions: +, nivells inferiors al 30% en comparació amb la proteïna salvatge; ++, nivells entre el 30-60% en comparació amb la proteïna salvatge; +++, nivells superiors en comparació amb la proteïna salvatge; N, localització nuclear; NP, no analitzat; *, defecte d'expressió degut a la baixa estabilitat de la proteïna.

Gen	VSD	Predicció proteïna	Assaig MMR	Expressió MLH1	Expressió PMS2	Localització subcel·lular		Classificació a nivell de proteïna
						MLH1	PMS2	
MLH1	c.25C>T	p.R9W	++	+	+	N	N	Inconcloent
MLH1	c.218T>C	p.L73P	+	+	+	N	N	Patogènic
MLH1	c.244A>G	p.T82A	+	+++	+++	N	N	Patogènic
MLH1	c.974G>A	p.R325Q	+++	+++	+++	N	N	Probablement netural
MLH1	c.1865T>A	p.L622H	NP	+*	+*	NP	NP	Patogènic
PMS2	c.59G>A	p.R20Q	+++	+++	+++	N	N	Probablement netural
PMS2	c.137G>T	p.S46I	+	+++	+++	N	N	Patogènic

Per la mutació fundadora c.1865T>A (L622H) del gen *MLH1*, es va voler aprofundir en la causa de la seva expressió reduïda mitjançant l'estudi d'estabilitat de la proteïna, utilitzant cicloheximida a diferents temps. Els resultats van detectar una pèrdua de l'estabilitat de la proteïna MLH1, que probablement és la causa de la seva expressió reduïda. Aquest fet, ja havia estat reportat en la variant p.R265C del gen *MLH1* (Perera and Bapat, 2008).

Tot i que en la majoria de casos, l'expressió disminuïda d'una variant correlaciona amb baixos nivells d'activitat reparadora, s'han descrit VSD concretes, com la p.H329P i la p.L550P en el gen *MLH1*, o la p.M688P en *MSH2,* amb uns nivells baixos d'expressió de la proteïna reparadora a estudiar on, per contra de l'esperat, presentaven capacitat de reparació (Kansikas, et al., 2011). En la nostra sèrie, la variant p.R9W en el gen *MLH1*, mostra una expressió disminuïda de les proteïnes MLH1 i PMS2, però en canvi presenta una capacitat de reparació al voltant del 50%. Per aquest motiu, aquesta variant roman com a VSD en no ser capaços d'interpretar adequadament aquests resultats. Aquesta observació il·lustra les limitacions inherents a cadascun dels assajos utilitzats (funció i expressió) i ajuda a entendre la importància de ser capaços d'integrar tota la informació que es pot generar al laboratori d'una manera coherent.

D'altra banda, també s'han detectat algunes poques variants que mostren una expressió major en comparació amb l'expressió de la proteïna salvatge, com les variants

p.E663G, p.R755S i p.E102D del gen *MLH1* descrites a la literatura (Kosinski, et al., 2010; Takahashi, et al., 2007), o com seria el cas de la variant c.974G>A (p.R325Q) del gen *MLH1*, en la sèrie analitzada. Sorprenentment però, l'activitat reparadora de la p.R325Q no és superior a la de la proteïna salvatge.

En el nostre treball l'estudi de la interacció entre les proteïnes MLH1 i PMS2 no es va realitzar mitjançant els assajos clàssics d'interacció (co-inmunoprecipitació i *GST-pull down* entre d'altres). De tota manera, pensem que els defectes d'interacció podrien haver estat detectats en l'anàlisi de l'expressió, doncs esperaríem una expressió de PMS2 disminuïda respecte la de MLH1, com s'havia suggerit anteriorment (Kosinski, et al., 2010; Mohd, et al., 2006).

- **Localització subcel·lular:** En últim terme, es va analitzar la localització subcel·lular de les proteïnes reparadores en cèl·lules transfectades. El transport al nucli d'aquestes proteïnes és un prerequisit per la correcta reparació dels errors del DNA (Knudsen, et al., 2009). Aquest transport disposa d'un mecanisme regulador addicional a l'expressió de la proteïna i a la interacció proteïna-proteïna, que podria ser modificat per la VSD identificada. Valorar la implicació de la variant en aquest procés requerirà d'un estudi de localització subcel·lular. Aquests assajos, a més de tenir una dificultat tècnica notable, requereixen d'especial atenció en el moment d'escollir els controls a utilitzar, la valoració dels nivells d'expressió i la naturalesa i localització del fluorocrom a emprar, a fi i efecte de no alterar els resultats obtinguts (Brieger, et al., 2012).

L'anàlisi va demostrar que cap de les VSD analitzades en aquesta tesi, mostrava un defecte en la seva localització subcel·lular (Taula 16). Aquests resultats estan en consonància amb la localització de les variants en el gen, ja que cap d'elles es troba localitzada en els senyals de localització nuclears (NLS). Tot i així, estudis recents detecten defecte en la localització subcel·lular d'algunes variants com la p.R622L i la p.C697F del gen *MSH2* o les variants p.E37K, p.G67R i p.C233del en el gen *MLH1*, que no es troben localitzades en els NLS (Andersen, et al., 2012; Lutzen, et al., 2008). Aquest fet avala la importància dels assajos de localització en la caracterització funcional de les VSD.

4. ALGORITME PROPOSAT PER L'AVALUACIÓ DE LA PATOGENICITAT DE LES VSD IDENTIFICADES EN ELS GENS REPARADORS

Amb l'objectiu d'interpretar les dades funcionals obtingudes en l'anàlisi de les VSD s'han proposat diferents algoritmes de classificació de variants. L'any 2008, es va proposar un primer algoritme on s'integren diferents tipus d'assajos per a l'anàlisi *in vitro* de les VSD en els gens reparadors en famílies amb sospita de síndrome de Lynch (Figura 19, introducció)(Couch, et al., 2008). Recentment aquest algoritme es va validar a partir d'un conjunt de variants en els gens *MLH1*, *MSH2* i *MSH6* (Kansikas, et al., 2011). En aquesta validació de l'algoritme es va observar un bon valor predictiu dels programes *in silico* a nivell de proteïna (SIFT i MAPP-MMR) en relació als resultats dels estudis funcionals. Per aquest motiu, els autors consideren que si un dels dos programes *in silico* (SIFT i MAPP-MMR) prediu patogenicitat, aquesta variant pot classificar-se com a patogènica i no cal continuar la confirmació mitjançant un estudi funcional.

Com s'exposa a l'article 2 de resultats, a partir de les nostres observacions i de l'algoritme proposat prèviament per Couch, et al., 2008, es va dissenyar una nova proposta. El nostre algoritme inclou la informació clínica, l'estudi de la freqüència de la variant en controls, les prediccions *in silico*, els estudis a nivell de mRNA i a nivell de proteïna, així com l'assaig d'activitat reparadora, l'expressió de les proteïnes i la localització subcel·lular d'aquestes (Figura 30).

En primer lloc, creiem fonamental recopilar una sèrie de dades com són la informació clínica personal i familiar, la patologia del tumor i la cosegregació de la variant amb la malaltia, que ens podran ajudar a classificar les VSD junt amb les dades posteriorment obtingudes dels estudis funcional. En el camp de la caracterització molecular de les mutacions i la seva possible utilització clínica és fonamental tenir en compte aquesta informació. Els clínics tendeixen a desconfiar dels resultats de tècniques experimentals *in vitro* que tot sovint són poc reproduïbles. Per tant el valor relatiu que es dóna a la informació clínica és pel moment molt important.

Posteriorment, proposem que el següent anàlisi a realitzar seria l'estudi de freqüència d'aquestes VSD, ja que poden determinar si es tracta d'una variant polimòrfica. D'aquesta manera, per les variants amb una freqüència superior al 1%, no serà necessari continuar amb els estudis posteriors.

D'acord amb els arguments exposats en l'apartat anterior, els resultats de les prediccions *in silico,* permeten prioritzar els estudis a nivell de RNA en front dels de proteïna, o a la inversa (detallat més avall). Segons el nostre criteri, en absència de sèries més grans, considerem necessari l'estudi funcional per tal de confirmar aquestes prediccions, en contraposició amb altres autors (Kansikas, et al., 2011).

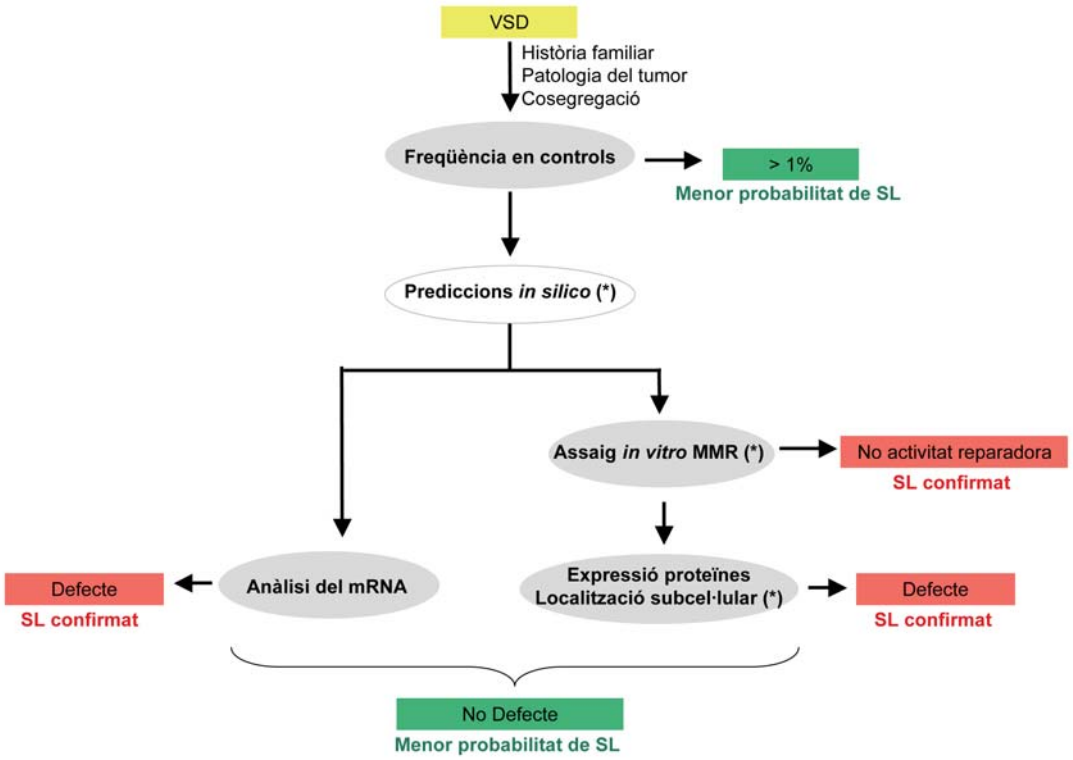


Figura 30. Algoritme proposat a Borràs et al 2012, per l'estudi de VSD en els gens reparadors. (*) Usat en l'anàlisi de VSD no intròniques ni silents.

Per altra banda, considerem que l'anàlisi del mRNA ha de ser un dels primers passos en l'avaluació de l'impacte funcional de les VSD, ja que un elevat percentatge de les VSD caracteritzades en aquesta tesi (31%) mostren defectes a nivell de processament del mRNA, corroborant així la importància que alguns autors suggerien en aquest tipus d'estudis (Arnold, et al., 2009; Sharp, et al., 2004; Tournier, et al., 2008). La integració dels resultats de l'estudi de processament junt amb els resultats de l'ASE ens van permetre classificar com a patogènica una de les variants analitzades. Considerem que els estudis d'ASE ens poden ajudar a determinar la patogenicitat d'algunes variants on els resultats de l'estudi de *splicing* no han detectat patogenicitat (Perera 2010) o bé no han sigut del tot concloents (com seria el cas de la variant c.538-3C>G), i per aquest motiu els incloem en el nostre algoritme de classificació junt amb els estudis de processament del mRNA.

A nivell de proteïna, es considera que si una variant mostra una pèrdua en l'activitat reparadora *in vitro,* aquesta pèrdua és altament predictiva de la patogenicitat de la variant (Drost, et al., 2010; Drost, et al., 2012; Kansikas, et al., 2011; Kosinski, et al., 2010; Ollila, et al., 2006), per aquest motiu proposem realitzar l'assaig d'activitat reparadora *in vitro* en primer lloc. En el cas que el resultat d'aquest assaig fos no concloent o no hi hagi defecte de reparació, a la literatura s'ha proposat la utilització d'assaigs que avaluïn altres aspectes de la proteïna per tal de classificar les VSD, com podrien ser l'expressió de les proteïnes reparadores o la localització subcel·lular (Couch, et al., 2008; Ou, et al., 2007). Des del nostre punt de vista, aquests assaigs ens poden ajudar a determinar la possible patogenicitat d'una variant degut a defectes en altres aspectes o funcions de la proteïna.

En comparació amb l'algoritme proposat per (Couch, et al., 2008), aquest nou algoritme remarca la importància de la informació clínica (antecedents familiars i la patologia tumoral) i la freqüència en controls, i proporciona un marc útil per incloure tota la informació a l'hora d'avaluar les VSD. A més, suggereix que les prediccions *in silico,* poden ajudar en la priorització dels estudis a nivell de RNA o de proteïna de manera mútuament excloent.

La utilitat de l'algoritme proposat es va validar en una sèrie de 25 variants en *MLH1* que prèviament havien estat analitzades tant a nivell de RNA com de proteïna (Taules S4 i S5 de l'article 2 de resultats). Si després d'haver avaluat les característiques clínico-patològiques haguéssim aplicat l'algoritme, 20 VSD d'aquesta sèrie de validació s'haguessin classificat, ja sigui com patogèniques (n=13) o probablement neutres (n=7). Per aquestes variants, les prediccions *in silico* de les alteracions en el processament del mRNA van mostrar una elevada correlació, ja que de les 4 VSD on almenys un dels programes *in silico* de *splicing* van predir patogenicitat, aquesta va ser confirmada experimentalment. Cal destacar, que únicament el Spliceport (Dogan , et al., 2007) ofereix una correlació fiable entre les observacions experimentals i la predicció (en 4 de les 4 variants patogèniques), igual que en la sèrie descrita en aquesta tesi. Pel que fa a les prediccions a nivell de proteïna, aquesta correlació no era tan elevada, ja que només 9 de les 13 variants predites com patogèniques van ser confirmades experimentalment. Tot i així, s'ha de tenir en compte que tots els assaigs funcionals no s'havien dut a terme per totes les variants i que a més en algunes ocasions no eren concloents. Així doncs, aquestes dades i les obtingudes en la sèrie de variants en *MLH1* i *PMS2* analitzades, donen suport a la utilització de les prediccions *in silico* per dirigir cap a un anàlisi *in vitro* a nivell de RNA o de proteïna.

En la sèrie de validació, es van classificar el 52% (13/25) de les variants com a patogèniques, el 16% (5/13) d'aquestes després dels estudis a nivell de RNA i el 61% (8/13) després dels estudis funcionals a nivell de proteïna. El 36% (9/25) van ser classificades com a probablement neutrals i el 12% (3/25) romanen com a VSD (Figura 31). Cal tenir en compte, que en algunes de les variants analitzades no es van dur a terme els estudis de localització subcel·lular. Aquest estudi de validació, reforça les observacions prèvies de l'estudi de variants en els gens *MLH1* i *PMS2*, així com suggereixen que l'algoritme proposat podria ser de gran ajuda per a la classificació de variants en el gens reparadors.

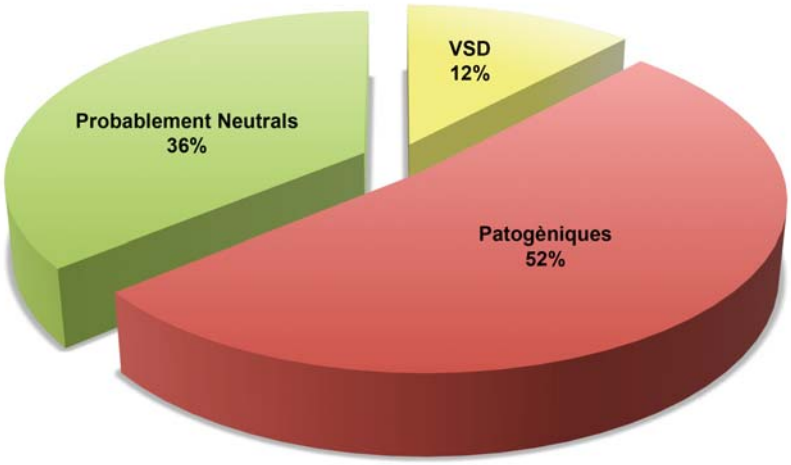


Figura 31. Classificació de les VSD analitzades en la sèrie de validació. Per aquest anàlisi s'han utilitzat 25 variant reportades a la literatura on s'havien analitzar a nivell de RNA i proteïna. Cal tenir en compte, en algunes de les variants analitzades no es va dur a terme l'assaig de localització subcel·lular (Taula S4, article 2 resultats).

Posteriorment a la publicació de l'algoritme proposat en el nostre estudi de VSD en el gen *MLH1*, s'ha publicat una nova modificació de l'algoritme de Couch i col·laboradors (Couch, et al., 2008). Aquesta nova versió, ha introduït les prediccions *in silico* i els assaigs *in vitro* a nivell de RNA, i els assajos basats en el cicle ATP-ADP (Rasmussen, et al., 2012)(Figura 32). Aquest nou algoritme també utilitza les prediccions *in silico* per prioritzar els estudis a nivell de RNA sobre els de proteïna, en el cas que aquestes prediguin un possible defecte a aquest nivell, tot considerant la importància dels estudis a nivell de RNA per classificar les VSD. A destacar que han introduït els assaigs del cicle ATP-ADP, normalment utilitzats per l'estudi de variants localitzades en els gens *MSH2* i *MSH6*. Aquestes dues proteïnes varien la seva conformació en funció de la seva unió al ATP o ADP i aquest canvi conformacional és necessari pel seu correcte funcionament. En afegir aquest nou aspecte del funcionament de les proteïnes permet tenir una visió més àmplia del possible impacte de la variant.

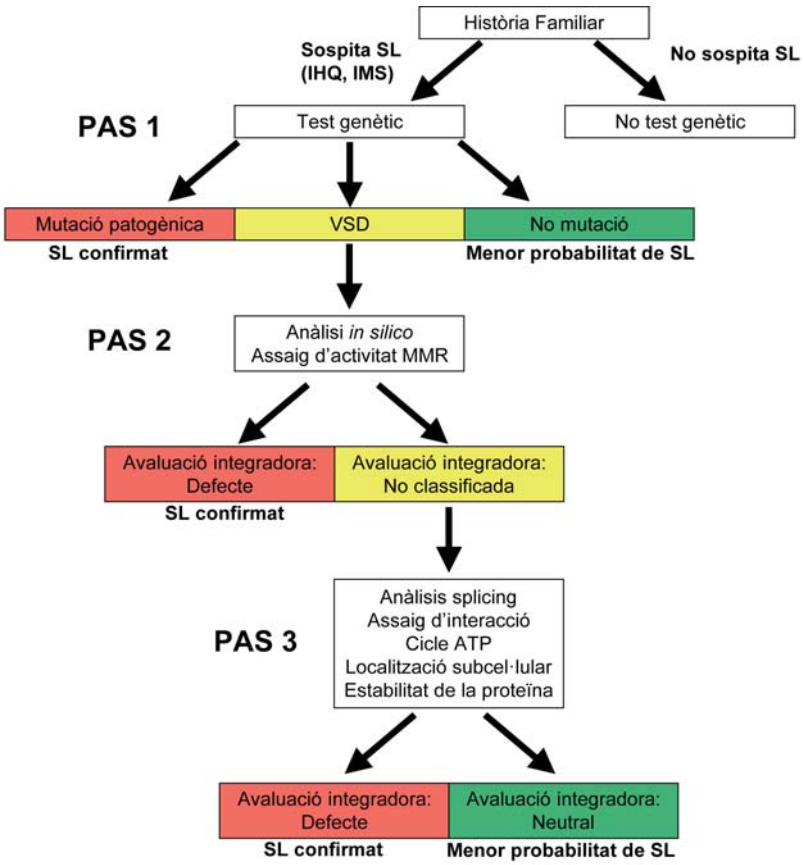


Figura 32. Algoritme proposat per l'estudi de VSD en els gens reparadors. Extret de Rasmussen, et al., 2012.

Els algoritmes de classificació funcional proposats fins al moment, són una eina que ens ajuda en el procés d'anàlisi de les VSD i la seva possible classificació. Tot i així, no està ben establert quins criteris seguir per classificar de forma definitiva una variant. Alguns autors han suggerit sistemes de classificació que es basen en la probabilitat de que una variant sigui patogènica a partir de models Bayesians (Plon, et al., 2008), i d'altres han proposat una classificació qualitativa (Thompson, et al., 2012b).

La classificació per probabilitat de patogenicitat, es porten a terme a partir de la utilització de models multifactorials Bayesians, que permeten integrar el conjunt d'evidències clinico-patològiques en un únic model (Goldgar, et al., 2008; Pastrello, et al., 2011; Plon, et al., 2008). Aquests models representen una alternativa a la classificació de les variants mitjançant estudis funcionals, i són especialment útils per analitzar variants identificades en un nombre important de famílies. Tot i així, és necessari tenir un grau d'evidència clínica molt sòlid per poder classificar aquestes variants. Alguns autors han intentat integrar les característiques tumorals amb altres evidències funcionals per classificar variants (Arnold, et al., 2009; Barnetson, et al.,

2008; Pastrello, et al., 2011). Així, recentment s'ha publicat un model multifactorial que incorpora les probabilitats dels estudis *in silico* junt amb les característiques del tumor i sembla oferir una bona correlació amb la clínica d'aquestes variants (Thompson, et al., 2012a).

Per altra banda, existeixen models qualitatius de classificació de VSD originalment desenvolupats en el CCFR (*Colon Cancer Family Registry*) i que segueixen els suggeriments del *InSiGHT Mutation Interpetation Committee*. Aquest sistema de classificació es basa en les 5 classes proposades per Plon, et al., 2008 i segueix els criteris descrits a la taula 17.

Taula 17. Criteris qualitatius de classificació proposats per l'estudi de VSD. Abreviatures: AMS, Amsterdam; MSI-H, inestabilitat de microsatèl·lits elevada. Extreta de Thompson, et al., 2012b.

Classe	Criteri utiiltzat per la classificació quantitativa
Classe 5-Patogènic	S'han de complir **totes** les característiques: (I) Funció deficient de la proteïna en assajos funcionals *in vivo/ex vivo* en un sistema de mamífers (no pot ser només en llevats). (II) Cosegregació amb la malaltia en almenys una família AMS amb ≥4 portadors afectats, o ≥2 families amb ≥3 portadors afectats. (III) No està present en la població general (>160 individus=320 al·lels). (IV) MSI-H en ≥2 tumors independents, sense resultats contradictòris de IHQ o pèrdua de l'expressió de les proteïnes reparadores en relació amb la localització de la variant en ≥2 tumors independents per MLH1 o ≥1 tumor per MSH2, MSH6 i PMS2.
Classe 4-Probablement Patogènic	Funció de la proteïna deficient **en un o més** assaigs *in vitro / ex vivo* en qualsevol sistema eucariota, **més una** de les següents característiques: (I) Cosegregació amb la malaltia en almenys una família AMS amb ≥3 portadors afectats. (II) MSI-H en ≥2 tumors independents, sense resultats contradictòris IHC o pérdua d'expressió de les proteïnes MMR (s) en relació amb la localització de la variant en ≥ 2 tumors independents per MLH1 o ≥ 1 tumor per MSH2, MSH6 i PMS2.
Classe 3-Variant de significat incert	Insuficients evidències per classificar-la en cap de les altres classes.
Classe 2-Probablment neutral	Variant reportada amb una freqüència ≥1% en un grup ètnic específic, i que no ha estat exclosa de ser una mutació fundadora coneguda. **O** una variant reportada amb una freqüència <1% en la població en general, amb una funció normal de la proteïna en els assajos *in vitro / ex vivo* en qualsevol línia cel·lular eucariota i sense defecte de *splicing*.
Classe 1-No Patogènica	Variant reportada en la població general amb una freqüència ≥ 1% **O** variant present en la població en general amb una freqüència del 0,1% - 1% i determinda en série gran de cas-control i associada amb un risc estimat <1.5.

Aquest criteris són altament restrictius per poder classificar las variants en les classes 5, 4, 2 i 1 (les clínicament rellevants) i per tant molts cops després dels estudis funcionals les variants continuen classificades com a classe 3 (VSD). Això és degut al fet que per exemple, per poder classificar una variant en les classes 5 i 4, no només són necessaris els estudis funcionals sinó que també és de vital importància tenir una informació clínica rellevant, informació de les característiques del tumor però també de la cosegregació de la variant amb la malaltia, com s'ha comentat anteriorment, l'obtenció d'aquesta informació molts cops presenta dificultats.

Des del nostre punt de vista, aquesta classificació és de gran utilitat si es té tant la informació clínica com la informació dels estudis funcionals, però en el cas que no es disposi de la informació clínica, els algoritmes funcionals de classificació de VSD proposats són els que actualment ens poden ajudar a determinar la patogenicitat de la variant.

Tot i així, és necessari seguir treballant tant en els models bayesians com en els criteris de classificació qualitativa a nivell col·lectiu, per poder arribar a un consens de classificació de les VSD a nivell mundial. Cal remarcar, que en paral·lel al desenvolupament d'aquests models i dels criteris de classificació de les VSD, la validació dels estudis funcionals utilitzats, és fonamental per poder introduir aquesta informació de forma generalitzada a la clínica.

SUMARI

• Les mutacions c.306+5G>A i c.1865T>A (p.L622H) són les primeres mutacions fundadores en el gen *MLH1* identificades a Espanya. Les mutacions c.306+5G>A i c.1865T>A presenten elevada prevalença i penetrància moderada

• La caracterització funcional de les VSD en els gens reparadors *MLH1* i *PMS2* en la sèrie descrita en la present tesi, ha permès classificar el 56% (9/16) de les variants analitzades com a patogèniques i el 25% (4/16) com a probablement neutrals. El 19% (3/16) romanen com a VSD.

• Les prediccions *in silico* sobre l'impacte d'una variant en el processament del mRNA i en la funcionalitat de la proteïna ofereixen una bona correlació entre predicció i resultats experimentals. Aquestes prediccions *in silico* són útils per prioritzar els estudis funcionals a nivell de RNA en front dels de proteïna, o a la inversa.

• Els estudis a nivell de mRNA han permès classificar com a patogèniques el 31%(5/16) de les variants analitzades. Aquest fet confirma l'alta prevalença d'aquestes alteracions i la importància d'aquests estudis per la classificació de VSD.

• Els assaigs *in vitro* de reparació i l'avaluació de l'expressió de proteïnes han demostrat ser útils en la classificació de les VSD identificades en gens reparadors, permetent classificar com a patogèniques el 25% (4/16) de les variants analitzades.

• S'ha proposat un algoritme de classificació per les VSD en els gens reparadors, que estableix una estratègia raonada per a l'avaluació de la seva patogenicitat, integrant la informació clínico-patològica amb estudis a nivell de RNA i proteïna, *in silico* i *in vitro.*

CONCLUSIONS

• L'elevada prevalença i la penetrància moderada de les mutacions c.306+5G>A i c.1865T>A, s'ha de tenir en consideració quan es dissenyin estratègies diagnòstiques moleculars en les famílies amb síndrome de Lynch espanyoles.

• L'anàlisi exhaustiu de les mutacions que inclou l'avaluació de la freqüència en població control, l'anàlisi de cosegregació, dades clinico-patològiques i el resultat d'assaigs funcionals a diferents nivells ha permès classificar un 81.2% de totes les VSD detectades en els gens *MLH1* i *PMS2* en la nostra sèrie. La classificació d'aquestes variants permet millorar el diagnòstic molecular i alhora el consell genètic d'aquests pacients i les seves famílies.

BIBLIOGRAFIA

A

Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR and others. 1993. Clues to the pathogenesis of familial colorectal cancer. Science (New York, N.Y 260:812-6.

Aarnio M, Mecklin JP, Aaltonen LA, Nystrom-Lahti M, Jarvinen HJ. 1995. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 64:430-3.

Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomaki P, Mecklin JP, Jarvinen HJ. 1999. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 81:214-8.

Abeloff M. 1995. Clinical Oncology. Sutton, UK: Churchill Livingstone.

Aceto GM, De Lellis L, Catalano T, Veschi S, Radice P, Di Iorio A, Mariani-Costantini R, Cama A, Curia MC. 2009. Nonfluorescent denaturing HPLC-based primer-extension method for allele-specific expression: application to analysis of mismatch repair genes. Clinical chemistry 55:1711-8.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nature methods 7:248-9.

Ahnen DJ. 2011. The American College of Gastroenterology Emily Couric Lecture--the adenoma-carcinoma sequence revisited: has the era of genetic tailoring finally arrived? Am J Gastroenterol 106:190-8.

Alarcon F, Lasset C, Carayol J, Bonadona V, Perdry H, Desseigne F, Wang Q, Bonaiti-Pellie C. 2007. Estimating cancer risk in HNPCC by the GRL method. Eur J Hum Genet 15:831-6.

Alemayehu A, Sebova K, Fridrichova I. 2008. Redundant DNA methylation in colorectal cancers of Lynch-syndrome patients. Genes Chromosomes Cancer 47:906-14.

Altieri F, Grillo C, Maceroni M, Chichiarelli S. 2008. DNA damage and repair: from molecular mechanisms to health implications. Antioxid Redox Signal 10:891-937.

Andersen SD, Liberti SE, Lutzen A, Drost M, Bernstein I, Nilbert M, Dominguez M, Nystrom M, Hansen TV, Christoffersen JW and others. 2012. Functional characterization of MLH1 missense variants identified in lynch syndrome patients. Hum Mutat

Armaghany T, Wilson JD, Chu Q, Mills G. 2012. Genetic alterations in colorectal cancer. Gastrointest Cancer Res 5:19-27.

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA and others. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30:757-70.

Astler VB, Coller FA. 1954. The prognostic significance of direct extension of carcinoma of the colon and rectum. Ann Surg 139:846-52.

Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S and others. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. Hum Mutat 27:145-54.

B

Balmana J, Stockwell DH, Steyerberg EW, Stoffel EM, Deffenbaugh AM, Reid JE, Ward B, Scholl T, Hendrickson B, Tazelaar J and others. 2006. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. JAMA 296:1469-78.

Barnetson RA, Cartwright N, van Vliet A, Haq N, Drew K, Farrington S, Williams N, Warner J, Campbell H, Porteous ME and others. 2008. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. Hum Mutat 29:367-74.

Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. 2006. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 354:2751-63.

Bellizzi AM, Frankel WL. 2009. Colorectal cancer due to deficiency in DNA mismatch repair function: a review. Adv Anat Pathol 16:405-17.

Belvederesi L, Bianchi F, Loretelli C, Gagliardini D, Galizia E, Bracci R, Rosati S, Bearzi I, Viel A, Cellerino R and others. 2006. Assessing the pathogenicity of MLH1 missense mutations in patients with suspected hereditary nonpolyposis colorectal cancer: correlation with clinical, genetic and functional features. Eur J Hum Genet 14:853-9.

Betz B, Theiss S, Aktas M, Konermann C, Goecke TO, Moslein G, Schaal H, Royer-Pokora B. 2010. Comparative in silico analyses and experimental validation of novel splice site and missense mutations in the genes MLH1 and MSH2. Journal of cancer research and clinical oncology 136:123-34.

Blasi MF, Ventura I, Aquilina G, Degan P, Bertario L, Bassi C, Radice P, Bignami M. 2006. A human cell-based assay to evaluate the effects of alterations in the MLH1 mismatch repair gene. Cancer Res 66:9036-44.

Boland CR, Goel A. 2010. Microsatellite instability in colorectal cancer. Gastroenterology 138:2073-2087 e3.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN and others. 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58:5248-57.

Bonadona V, Bonaiti B, Olschwang S, Grandjouan S, Huiart L, Longy M, Guimbaud R, Buecher B, Bignon YJ, Caron O and others. 2011. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA 305:2304-10.

Borelli I, Barberis MA, Spina F, Cavalchini GC, Vivanet C, Balestrino L, Micheletti M, Allavena A, Sala P, Carcassi C and others. 2012. A unique MSH2 exon 8 deletion accounts for a major portion of all mismatch repair gene mutations in Lynch syndrome families of Sardinian origin. Eur J Hum Genet

Borras E, Pineda M, Blanco I, Jewett EM, Wang F, Teule A, Caldes T, Urioste M, Martinez-Bouzas C, Brunet J and others. 2010. MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. Cancer Res 70:7379-91.

Boyle P, Levin B, eds. 2008. World Cancer Report 2008. World Health Organization. International Agency for Research on Cancer. Lyon, France.

Bozzao C, Lastella P, Stella A. 2011. Anticipation in lynch syndrome: where we are where we go. Curr Genomics 12:451-65.

Bray F, Ren JS, Masuyer E, Ferlay J. 2012. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer

Brieger A, Plotz G, Hinrichsen I, Passmann S, Adam R, Zeuzem S. 2012. C-terminal fluorescent labeling impairs functionality of DNA mismatch repair proteins. PLoS One 7:e31863.

Brieger A, Plotz G, Raedle J, Weber N, Baum W, Caspary WF, Zeuzem S, Trojan J. 2005. Characterization of the nuclear import of human MutLalpha. Mol Carcinog 43:51-8.

Brieger A, Trojan J, Raedle J, Plotz G, Zeuzem S. 2002. Transient mismatch repair gene transfection for functional analysis of genetic hMLH1 and hMSH2 variants. Gut 51:677-84.

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A and others. 1994. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 368:258-61.

Brunak S, Engelbrecht J, Knudsen S. 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 220:49-65.

Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. Curr Opin Struct Biol 8:346-54.

Burset M, Seledtsov IA, Solovyev VV. 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 29:255-9.

Burt RW. 2000. Colon cancer screening. Gastroenterology 119:837-53.

C

Caluseriu O, Di Gregorio C, Lucci-Cordisco E, Santarosa M, Trojan J, Brieger A, Benatti P, Pedroni M, Colibazzi T, Bellacosa A and others. 2004. A founder MLH1 mutation in families from the districts of Modena and Reggio-Emilia in northern Italy with hereditary non-polyposis colorectal cancer associated with protein elongation and instability. J Med Genet 41:e34.

Cancer Genome Atlas N. 2012. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330-7.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nature reviews 3:285-98.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31:3568-71.

Casey G, Lindor NM, Papadopoulos N, Thibodeau SN, Moskow J, Steelman S, Buzin CH, Sommer SS, Collins CE, Butz M and others. 2005. Conversion analysis for mutation detection in MLH1 and MSH2 in patients with colorectal cancer. Jama 293:799-809.

Castellsague E, Gonzalez S, Guino E, Stevens KN, Borras E, Raymond VM, Lazaro C, Blanco I, Gruber SB, Capella G. 2010. Allele-specific expression of APC in adenomatous polyposis families. Gastroenterology 139:439-47, 447 e1.

Cederquist K, Emanuelsson M, Wiklund F, Golovleva I, Palmqvist R, Gronberg H. 2005. Two Swedish founder MSH6 mutations, one nonsense and one missense, conferring high cumulative risk of Lynch syndrome. Clin Genet 68:533-41.

Chan TL, Chan YW, Ho JW, Chan C, Chan AS, Chan E, Lam PW, Tse CW, Lee KC, Lau CW and others. 2004. MSH2 c.1452-1455delAATG is a founder mutation and an important cause of hereditary nonpolyposis colorectal cancer in the southern Chinese population. Am J Hum Genet 74:1035-42.

Chan TL, Yuen ST, Ho JW, Chan AS, Kwan K, Chung LP, Lam PW, Tse CW, Leung SY. 2001. A novel germline 1.8-kb deletion of hMLH1 mimicking alternative splicing: a founder mutation in the Chinese population. Oncogene 20:2976-81.

Chang DK, Ricciardiello L, Goel A, Chang CL, Boland CR. 2000. Steady-state regulation of the human DNA mismatch repair system. J Biol Chem 275:29178.

Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, Ng P, Gruber SB, Watson P, Rennert G, Anton-Culver H and others. 2008. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). Hum Mutat 29:852-60.

Chen S, Wang W, Lee S, Nafa K, Lee J, Romans K, Watson P, Gruber SB, Euhus D, Kinzler KW and others. 2006. Prediction of germline mutations and cancer risk in the Lynch syndrome. JAMA 296:1479-87.

Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T and others. 2006. Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. Cancer Res 66:2019-27.

Choi YH, Cotterchio M, McKeown-Eyssen G, Neerav M, Bapat B, Boyd K, Gallinger S, McLaughlin J, Aronson M, Briollais L. 2009. Penetrance of colorectal cancer among MLH1/MSH2 carriers participating in the colorectal cancer familial registry in Ontario. Hered Cancer Clin Pract 7:14.

Christmann M, Tomicic MT, Kaina B. 2002. Phosphorylation of mismatch repair proteins MSH2 and MSH6 affecting MutSalpha mismatch-binding activity. Nucleic Acids Res 30:1959-66.

Chung DC, Rustgi AK. 2003. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. Ann Intern Med 138:560-70.

Clark AB, Cook ME, Tran HT, Gordenin DA, Resnick MA, Kunkel TA. 1999. Functional analysis of human MutSalpha and MutSbeta complexes in yeast. Nucleic Acids Res 27:736-42.

Clendenning M, Baze ME, Sun S, Walsh K, Liyanarachchi S, Fix D, Schunemann V, Comeras I, Deacon M, Lynch JF and others. 2008a. Origins and prevalence of the American Founder Mutation of MSH2. Cancer Res 68:2145-53.

Clendenning M, Hampel H, LaJeunesse J, Lindblom A, Lockman J, Nilbert M, Senter L, Sotamaa K, de la Chapelle A. 2006. Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat 27:490-5.

Clendenning M, Senter L, Hampel H, Robinson KL, Sun S, Buchanan D, Walsh MD, Nilbert M, Green J, Potter J and others. 2008b. A frame-shift mutation of PMS2 is a widespread cause of Lynch syndrome. J Med Genet 45:340-5.

Cotton RG, Scriver CR. 1998. Proof of "disease causing" mutation. Hum Mutat 12:1-3.

Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de Wind N. 2008. Assessment of functional effects of unclassified genetic variants. Hum Mutat 29:1314-26.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome research 14:1188-90.

Cyr JL, Brown GD, Stroop J, Heinen CD. 2012. The predicted truncation from a cancer-associated variant of the MSH2 initiation codon alters activity of the MSH2-MSH6 mismatch repair complex. Mol Carcinog 51:647-58.

D

Davies H, Dicks E, Stephens P, Cox C, Teague J, Greenman C, Bignell G, O'Meara S, Edkins S, Parker A and others. 2006. High throughput DNA sequence variant detection by conformation sensitive capillary electrophoresis and automated peak comparison. Genomics 87:427-32.

de la Chapelle A. 2004. Genetic predisposition to colorectal cancer. Nat Rev Cancer 4:769-80.

De Vos M, Hayward BE, Picton S, Sheridan E, Bonthron DT. 2004. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am J Hum Genet 74:954-64.

Debniak T, Kurzawski G, Gorski B, Kladny J, Domagala W, Lubinski J. 2000. Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining hMLH1 and hMSH2 gene mutations in patients with colorectal cancer. Eur J Cancer 36:49-54.

den Bakker MA, Seynaeve C, Kliffen M, Dinjens WN. 2003. Microsatellite instability in a pleomorphic rhabdomyosarcoma in a patient with hereditary non-polyposis colorectal cancer. Histopathology 43:297-9.

Deng G, Bell I, Crawley S, Gum J, Terdiman JP, Allen BA, Truta B, Sleisenger MH, Kim YS. 2004. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. Clin Cancer Res 10:191-5.

Deschenes SM, Tomer G, Nguyen M, Erdeniz N, Juba NC, Sepulveda N, Pisani JE, Liskay RM. 2007. The E705K mutation in hPMS2 exerts recessive, not dominant, effects on mismatch repair. Cancer letters 249:148-56.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:e67.

Dietmaier W, Wallinger S, Bocker T, Kullmann F, Fishel R, Ruschoff J. 1997. Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. Cancer Res 57:4749-56.

Dogan RI, Getoor L, Wilbur WJ, Mount SM. 2007. SplicePort--an interactive splice-site analysis tool. Nucleic Acids Res 35:W285-91.

Donehower LA, Creighton CJ, Schultz N, Shinbrot E, Chang K, Gunaratne PH, Muzny D, Sander C, Hamilton SR, Gibbs RA and others. 2012. MLH1-Silenced and Non-Silenced Subgroups of Hypermutated Colorectal Carcinomas Have Distinct Mutational Landscapes. J Pathol

Drost M, Zonneveld JB, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind N. 2010. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum Mutat 31:247-53.

Drost M, Zonneveld JB, van Hees S, Rasmussen LJ, Hofstra RM, de Wind N. 2012. A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. Hum Mutat 33:488-94.

Drotschmann K, Clark AB, Tran HT, Resnick MA, Gordenin DA, Kunkel TA. 1999. Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. Proc Natl Acad Sci U S A 96:2970-5.

Dukes. 1932. The classification of cancer of the rectum. The Journal of patology 35:323.

Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, Liu B, Kinzler KW, Vogelstein B. 1997. Cancer risk associated with germline DNA mismatch repair gene mutations. Human molecular genetics 6:105-10.

Dzantiev L, Constantin N, Genschel J, Iyer RR, Burgers PM, Modrich P. 2004. A defined human system that supports bidirectional mismatch-provoked excision. Mol Cell 15:31-41.

E

Eng L, Coutinho G, Nahas S, Yeo G, Tanouye R, Babaei M, Dork T, Burge C, Gatti RA. 2004. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. Hum Mutat 23:67-76.

F

Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. Science (New York, N.Y 297:1007-13.

Fazakerley GV, Quignard E, Woisard A, Guschlbauer W, van der Marel GA, van Boom JH, Jones M, Radman M. 1986. Structures of mismatched base pairs in DNA and their recognition by the Escherichia coli mismatch repair system. EMBO J 5:3697-703.

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759-67.

Feliubadalo L, A. L-D, E. C, J. dV, M. M, E. T, E. M, R. C, C. G, O. C and others. (In press). Next-Generation Sequencing meets Genetic Diagnostics: development of a comprehensive workflow for the analysis of BRCA1 and BRCA2 genes. European Journal of Human Genetics

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127:2893-917.

Ferrari P, Jenab M, Norat T, Moskal A, Slimani N, Olsen A, Tjonneland A, Overvad K, Jensen MK, Boutron-Ruault MC and others. 2007. Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition (EPIC). Int J Cancer 121:2065-72.

Fischer SG, Lerman LS. 1979. Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. Cell 16:191-200.

Fleming ID, Phillips JL, Menck HR, Murphy GP, Winchester DP. 1997. The National Cancer Data Base report on recent hospital cancer program progress toward complete American Joint Committee on Cancer/TNM staging. Cancer 80:2305-10.

Foulkes WD, Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C, Shia J, Markowitz A, Figer A, Friedman E and others. 2002. The founder mutation MSH2*1906G-->C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. Am J Hum Genet 71:1395-412.

Froggatt NJ, Green J, Brassett C, Evans DG, Bishop DT, Kolodner R, Maher ER. 1999. A common MSH2 mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer. J Med Genet 36:97-102.

Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids 2010: 260512.

G

Gammie AE, Erdeniz N, Beaver J, Devlin B, Nanji A, Rose MD. 2007. Functional characterization of pathogenic human MSH2 missense mutations in Saccharomyces cerevisiae. Genetics 177:707-21.

Gassman NR, Clodfelter JE, McCauley AK, Bonin K, Salsbury FR, Jr., Scarpinato KD. 2011. Cooperative nuclear localization sequences lend a novel role to the N-terminal region of MSH6. PLoS One 6:e17907.

Gausachs M, Mur P, Corral J, Pineda M, Gonzalez S, Benito L, Menendez M, Espinas JA, Brunet J, Iniesta MD and others. 2012. MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. Eur J Hum Genet

Geary J, Sasieni P, Houlston R, Izatt L, Eeles R, Payne SJ, Fisher S, Hodgson SV. 2008. Gene-related cancer spectrum in families with hereditary non-polyposis colorectal cancer (HNPCC). Fam Cancer 7:163-72.

Giardiello FM, Brensinger JD, Petersen GM. 2001. AGA technical review on hereditary colorectal cancer and genetic testing. Gastroenterology 121:198-213.

Gille JJ, Hogervorst FB, Pals G, Wijnen JT, van Schooten RJ, Dommering CJ, Meijer GA, Craanen ME, Nederlof PM, de Jong D and others. 2002. Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach. British journal of cancer 87:892-7.

Giovannucci E. 2006. The epidemiology of vitamin D and colorectal cancer: recent findings. Curr Opin Gastroenterol 22:24-9.

Goldberg Y, Porat RM, Kedar I, Shochat C, Galinsky D, Hamburger T, Hubert A, Strul H, Kariiv R, Ben-Avi L and others. 2010. An Ashkenazi founder mutation in the MSH6 gene leading to HNPCC. Fam Cancer 9:141-50.

Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS. 2008. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat 29:1265-72.

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ, Breast Cancer Information Core Steering C. 2004. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 75:535-44.

Gonzalez-Garcia I, Moreno V, Navarro M, Marti-Rague J, Marcuello E, Benasco C, Campos O, Capella G, Peinado MA. 2000. Standardized approach for microsatellite instability detection in colorectal carcinomas. J Natl Cancer Inst 92:544-9.

Gonzalez-Perez A, Lopez-Bigas N. 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88:440-9.

Gradia S, Subramanian D, Wilson T, Acharya S, Makhov A, Griffith J, Fishel R. 1999. hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. Mol Cell 3:255-61.

Grady WM, Carethers JM. 2008. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 135:1079-99.

Gruber SB. 2006. New developments in Lynch syndrome (hereditary nonpolyposis colorectal cancer) and mismatch repair gene testing. Gastroenterology 130:577-87.

Gruber SB, Mukherjee B. 2009. Anticipation in lynch syndrome: still waiting for the answer. J Clin Oncol 27:326-7.

Guarne A, Ramon-Maiques S, Wolff EM, Ghirlando R, Hu X, Miller JH, Yang W. 2004. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. Embo J 23:4134-45.

Guerrette S, Acharya S, Fishel R. 1999. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. J Biol Chem 274:6336-41.

Guerrette S, Wilson T, Gradia S, Fishel R. 1998. Interactions of human hMSH2 with hMSH3 and hMSH2 with hMSH6: examination of mutations found in hereditary nonpolyposis colorectal cancer. Molecular and cellular biology 18:6616-23.

Gylling A, Ridanpaa M, Vierimaa O, Aittomaki K, Avela K, Kaariainen H, Laivuori H, Poyhonen M, Sallinen SL, Wallgren-Pettersson C and others. 2009. Large genomic rearrangements and germline epimutations in Lynch syndrome. Int J Cancer 124:2333-40.

H

Hampel H. 2010. Point: justification for Lynch syndrome screening among all patients with newly diagnosed colorectal cancer. J Natl Compr Canc Netw 8:597-601.

Hampel H, de la Chapelle A. 2011. The search for unaffected individuals with Lynch syndrome: do the ends justify the means? Cancer Prev Res (Phila) 4:1-5.

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Clendenning M, Sotamaa K, Prior T, Westman JA and others. 2008. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J Clin Oncol 26:5783-8.

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J and others. 2005a. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 352:1851-60.

Hampel H, Stephens JA, Pukkala E, Sankila R, Aaltonen LA, Mecklin JP, de la Chapelle A. 2005b. Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. Gastroenterology 129:415-21.

Han HJ, Maruyama M, Baba S, Park JG, Nakamura Y. 1995. Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). Human molecular genetics 4:237-42.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell 100:57-70.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. Cell 144:646-74.

Hardt K, Heick SB, Betz B, Goecke T, Yazdanparast H, Kuppers R, Servan K, Steinke V, Rahner N, Morak M and others. 2011. Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. Fam Cancer 10:273-84.

Hayes AP, Sevi LA, Feldt MC, Rose MD, Gammie AE. 2009. Reciprocal regulation of nuclear import of the yeast MutSalpha DNA mismatch repair proteins Msh2 and Msh6. DNA Repair (Amst) 8:739-51.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. 1996. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24:3439-52.

Heinen CD, Juel Rasmussen L. 2012. Determining the functional significance of mismatch repair gene missense variants using biochemical and cellular assays. Hered Cancer Clin Pract 10:9.

Heinen CD, Schmutte C, Fishel R. 2002a. DNA repair and tumorigenesis: lessons from hereditary cancer syndromes. Cancer Biol Ther 1:477-85.

Heinen CD, Wilson T, Mazurek A, Berardini M, Butz C, Fishel R. 2002b. HNPCC mutations in hMSH2 result in reduced hMSH2-hMSH6 molecular switch functions. Cancer Cell 1:469-78.

Hemminki A, Peltomaki P, Mecklin JP, Jarvinen H, Salovaara R, Nystrom-Lahti M, de la Chapelle A, Aaltonen LA. 1994. Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. Nat Genet 8:405-10.

Hendriks YM, de Jong AE, Morreau H, Tops CM, Vasen HF, Wijnen JT, Breuning MH, Brocker-Vriends AH. 2006a. Diagnostic approach and management of Lynch syndrome (hereditary nonpolyposis colorectal carcinoma): a guide for clinicians. CA Cancer J Clin 56:213-25.

Hendriks YM, Jagmohan-Changur S, van der Klift HM, Morreau H, van Puijenbroek M, Tops C, van Os T, Wagner A, Ausems MG, Gomez E and others. 2006b. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 130:312-22.

Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW and others. 1998. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A 95:6870-5.

Hernan I, Borras E, de Sousa Dias M, Gamundi MJ, Mane B, Llort G, Agundez JA, Blanca M, Carballo M. 2012. Detection of genomic variations in BRCA1 and BRCA2 genes by long-range PCR and next-generation sequencing. J Mol Diagn 14:286-93.

Hirata K, Kanemitsu S, Nakayama Y, Nagata N, Itoh H, Ohnishi H, Ishikawa H, Furukawa Y, registry H, genetic testing project of the Japanese Society for Cancer of the C and others. 2006. A novel germline mutation of MSH2 in a hereditary nonpolyposis colorectal cancer patient with liposarcoma. Am J Gastroenterol 101:193-6.

Hitchins MP, Ward RL. 2009. Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. J Med Genet 46:793-802.

Hofstra RM, Osinga J, Buys CH. 1997. Mutations in Hirschsprung disease: when does a mutation contribute to the phenotype. Eur J Hum Genet 5:180-5.

Hofstra RM, Spurdle AB, Eccles D, Foulkes WD, de Wind N, Hoogerbrugge N, Hogervorst FB, Group IUGVW. 2008. Tumor characteristics as an analytic tool for classifying genetic variants of uncertain clinical significance. Hum Mutat 29:1292-303.

Holmes J, Jr., Clark S, Modrich P. 1990. Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. Proc Natl Acad Sci U S A 87:5837-41.

Hsieh P. 2001. Molecular mechanisms of DNA mismatch repair. Mutat Res 486:71-87.

Hutter P, Couturier A, Scott RJ, Alday P, Delozier-Blanchet C, Cachat F, Antonarakis SE, Joris F, Gaudin M, D'Amato L and others. 1996. Complex genetic predisposition to cancer in an extended HNPCC family with an ancestral hMLH1 mutation. J Med Genet 33:636-40.

I

Iacopetta B, Grieu F, Amanuel B. 2010. Microsatellite instability in colorectal cancer. Asia Pac J Clin Oncol 6:260-9.

Imai K, Yamamoto H. 2008. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. Carcinogenesis 29:673-80.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. 1993. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363:558-61.

Jager AC, Bisgaard ML, Myrhoj T, Bernstein I, Rehfeld JF, Nielsen FC. 1997. Reduced frequency of extracolonic cancers in hereditary nonpolyposis colorectal cancer families with monoallelic hMLH1 expression. Am J Hum Genet 61:129-38.

J

Jager AC, Rasmussen M, Bisgaard HC, Singh KK, Nielsen FC, Rasmussen LJ. 2001. HNPCC mutations in the human DNA mismatch repair gene hMLH1 influence assembly of hMutLalpha and hMLH1-hEXO1 complexes. Oncogene 20:3590-5.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. 2011. Global cancer statistics. CA Cancer J Clin 61:69-90.

Jenkins MA, Baglietto L, Dowty JG, Van Vliet CM, Smith L, Mead LJ, Macrae FA, St John DJ, Jass JR, Giles GG and others. 2006. Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study. Clin Gastroenterol Hepatol 4:489-98.

Joseph N, Duppatla V, Rao DN. 2006. Prokaryotic DNA mismatch repair. Prog Nucleic Acid Res Mol Biol 81:1-49.

Jun SH, Kim TG, Ban C. 2006. DNA mismatch repair system. Classical and fresh roles. FEBS J 273:1609-19.

Jung B, Doctolero RT, Tajima A, Nguyen AK, Keku T, Sandler RS, Carethers JM. 2004. Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. Gastroenterology 126:654-9.

K

Kääriäinen H HM, Kristoffersson U, Nippert I, Rantanen E, Sequeiros J, Schmidtke J. 2008. Recommendations for genetic counselling related to genetic testing.

Kadyrov FA, Dzantiev L, Constantin N, Modrich P. 2006. Endonucleolytic function of MutLalpha in human mismatch repair. Cell 126:297-308.

Kaneda A, Yagi K. 2011. Two groups of DNA methylation markers to classify colorectal cancer into three epigenotypes. Cancer Sci 102:18-24.

Kang GH. 2011. Four molecular subtypes of colorectal cancer and their precursor lesions. Arch Pathol Lab Med 135:698-703.

Kansikas M, Kariola R, Nystrom M. 2011. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. Hum Mutat 32:107-15.

Kempers MJ, Kuiper RP, Ockeloen CW, Chappuis PO, Hutter P, Rahner N, Schackert HK, Steinke V, Holinski-Feder E, Morak M and others. 2011. Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. Lancet Oncol 12:49-55.

Knudsen NO, Andersen SD, Lutzen A, Nielsen FC, Rasmussen LJ. 2009. Nuclear translocation contributes to regulation of DNA excision repair activities. DNA Repair (Amst) 8:682-9.

Knudsen NO, Nielsen FC, Vinther L, Bertelsen R, Holten-Andersen S, Liberti SE, Hofstra R, Kooi K, Rasmussen LJ. 2007. Nuclear localization of human DNA mismatch repair protein exonuclease 1 (hEXO1). Nucleic Acids Res 35:2609-19.

Knudson AG, Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 68:820-3.

Kolas NK, Cohen PE. 2004. Novel and diverse functions of the DNA mismatch repair family in mammalian meiosis and recombination. Cytogenet Genome Res 107:216-31.

Kolodner RD, Marsischky GT. 1999. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev 9:89-96.

Kondo E, Horii A, Fukushige S. 2001. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. Nucleic Acids Res 29:1695-702.

Kondo E, Suzuki H, Horii A, Fukushige S. 2003. A yeast two-hybrid assay provides a simple way to evaluate the vast majority of hMLH1 germ-line mutations. Cancer Res 63:3302-8.

Kosinski J, Hinrichsen I, Bujnicki JM, Friedhoff P, Plotz G. 2010. Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. Hum Mutat 31:975-82.

Kunkel TA, Erie DA. 2005. DNA mismatch repair. Annu Rev Biochem 74:681-710.

Kunz C, Saito Y, Schar P. 2009. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. Cell Mol Life Sci 66:1021-38.

L

Ladabaum U, Ford JM. 2012. Lynch syndrome in patients with colorectal cancer: finding the needle in the haystack. JAMA 308:1581-3.

Ladabaum U, Wang G, Terdiman J, Blanco A, Kuppermann M, Boland CR, Ford J, Elkin E, Phillips KA. 2011. Strategies to identify the Lynch syndrome among patients with colorectal cancer: a cost-effectiveness analysis. Ann Intern Med 155:69-79.

Laguri C, Duband-Goulet I, Friedrich N, Axt M, Belin P, Callebaut I, Gilquin B, Zinn-Justin S, Couprie J. 2008. Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. Biochemistry 47:6199-207.

Larsen K, Petersen J, Bernstein I, Nilbert M. 2009. A parametric model for analyzing anticipation in genetically predisposed families. Stat Appl Genet Mol Biol 8:Article26.

Lastella P, Patruno M, Forte G, Montanaro A, Di Gregorio C, Sabba C, Suppressa P, Piepoli A, Panza A, Andriulli A and others. 2011. Identification and surveillance of 19 Lynch syndrome families in southern Italy: report of six novel germline mutations and a common founder mutation. Fam Cancer 10:285-95.

Lastella P, Surdo NC, Resta N, Guanti G, Stella A. 2006. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. BMC genomics 7:243.

Leong V, Lorenowicz J, Kozij N, Guarne A. 2009. Nuclear import of human MLH1, PMS2, and MutLalpha: redundancy is the key. Mol Carcinog 48:742-50.

Li GM. 2008. Mechanisms and functions of DNA mismatch repair. Cell research 18:85-98.

Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer C. 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clinical chemistry 50:1156-64.

Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ and others. 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 41:112-7.

Lutzen A, de Wind N, Georgijevic D, Nielsen FC, Rasmussen LJ. 2008. Functional analysis of HNPCC-related missense mutations in MSH2. Mutat Res 645:44-55.

Lynch HT, de la Chapelle A. 2003. Hereditary colorectal cancer. N Engl J Med 348:919-32.

Lynch HT, Lynch JF, Lynch PM. 2007. Toward a consensus in molecular diagnosis of hereditary nonpolyposis colorectal cancer (Lynch syndrome). J Natl Cancer Inst 99:261-3.

Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet 76:1-18.

M

Marinovic-Terzic I, Yoshioka-Yamashita A, Shimodaira H, Avdievich E, Hunton IC, Kolodner RD, Edelmann W, Wang JY. 2008. Apoptotic function of human PMS2 compromised by the nonsynonymous single-nucleotide polymorphic variant R20Q. Proc Natl Acad Sci U S A 105:13993-8.

Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B and others. 1995. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science (New York, N.Y 268:1336-8.

Marsischky GT, Kolodner RD. 1999. Biochemical characterization of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 complex and mispaired bases in DNA. J Biol Chem 274:26668-82.

Mastrocola AS, Heinen CD. 2010a. Lynch syndrome-associated mutations in MSH2 alter DNA repair and checkpoint response functions in vivo. Hum Mutat 31:E1699-708.

Mastrocola AS, Heinen CD. 2010b. Nuclear reorganization of DNA mismatch repair proteins in response to DNA damage. DNA Repair (Amst) 9:120-33.

Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, Tavtigian SV. 2006. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic Acids Res 34:1317-25.

Medina-Arana V, Barrios Y, Fernandez-Peralta A, Herrera M, Chinea N, Lorenzo N, Jimenez A, Martin-Lopez JV, Gonzalez-Hermoso F, Salido E and others. 2006. New founding mutation in MSH2 associated with hereditary nonpolyposis colorectal cancer syndrome on the Island of Tenerife. Cancer letters 244:268-73.

Menéndez M, Castellví-Bel S, Pineda M, de Cid R, Muñoz J, González S, Teulé A, Balaguer F, Ramón y Cajal T, Reñé JM and others. 2010. Founder effect of a pathogenic MSH2 mutation identified in Lynch syndrome Spanish families. Clin Genet

Modrich P, Lahue R. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 65:101-33.

Mohd AB, Palama B, Nelson SE, Tomer G, Nguyen M, Huo X, Buermeyer AB. 2006. Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. DNA Repair (Amst) 5:347-61.

Moller P, Clark N, Maehle L. 2011. A SImplified method for Segregation Analysis (SISA) to determine penetrance and expression of a genetic variant in a family. Hum Mutat 32:568-71.

Moreira L, Balaguer F, Lindor N, de la Chapelle A, Hampel H, Aaltonen LA, Hopper JL, Le Marchand L, Gallinger S, Newcomb PA and others. 2012. Identification of Lynch syndrome among patients with colorectal cancer. JAMA 308:1555-65.

Mvundura M, Grosse SD, Hampel H, Palomaki GE. 2010. The cost-effectiveness of genetic testing strategies for Lynch syndrome among newly diagnosed patients with colorectal cancer. Genet Med 12:93-104.

N

Nagasaka T, Sasamoto H, Notohara K, Cullings HM, Takeda M, Kimura K, Kambara T, MacPhee DG, Young J, Leggett BA and others. 2004. Colorectal cancer with

mutation in BRAF, KRAS, and wild-type with respect to both oncogenes showing different patterns of DNA methylation. J Clin Oncol 22:4584-94.

Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, Thibodeau SN, de la Chapelle A. 2004. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res 64:4721-7.

Nakagawa H, Yan H, Lockman J, Hampel H, Kinzler KW, Vogelstein B, De La Chapelle A. 2002. Allele separation facilitates interpretation of potential splicing alterations and genomic rearrangements. Cancer Res 62:4579-82.

Nalla VK, Rogan PK. 2005. Automated splicing mutation analysis by information theory. Hum Mutat 25:334-42.

Naruse H, Ikawa N, Yamaguchi K, Nakamura Y, Arai M, Ishioka C, Sugano K, Tamura K, Tomita N, Matsubara N and others. 2009. Determination of splice-site mutations in Lynch syndrome (hereditary non-polyposis colorectal cancer) patients using functional splicing assay. Fam Cancer 8:509-17.

Nathke IS. 2004. The adenomatous polyposis coli protein: the Achilles heel of the gut epithelium. Annu Rev Cell Dev Biol 20:337-66.

Ng PC, Henikoff S. 2002. Accounting for human polymorphisms predicted to affect protein function. Genome research 12:436-46.

Nilbert M, Timshel S, Bernstein I, Larsen K. 2009. Role for genetic anticipation in Lynch syndrome. J Clin Oncol 27:360-4.

Nystrom-Lahti M, Perrera C, Raschle M, Panyushkina-Seiler E, Marra G, Curci A, Quaresima B, Costanzo F, D'Urso M, Venuta S and others. 2002. Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. Genes Chromosomes Cancer 33:160-7.

O

Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. 2007. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. J Mol Diagn 9:305-14.

Olatubosun A, Valiaho J, Harkonen J, Thusberg J, Vihinen M. 2012. PON-P: integrated predictor for pathogenicity of missense variants. Hum Mutat 33:1166-74.

Ollila S, Dermadi Bebek D, Jiricny J, Nystrom M. 2008. Mechanisms of pathogenicity in human MSH2 missense mutants. Hum Mutat 29:1355-63.

Ollila S, Sarantaus L, Kariola R, Chan P, Hampel H, Holinski-Feder E, Macrae F, Kohonen-Corish M, Gerdes AM, Peltomaki P and others. 2006. Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. Gastroenterology 131:1408-17.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci U S A 86:2766-70.

Ou J, Niessen RC, Lutzen A, Sijmons RH, Kleibeuker JH, de Wind N, Rasmussen LJ, Hofstra RM. 2007. Functional analysis helps to clarify the clinical importance of unclassified variants in DNA mismatch repair genes. Hum Mutat 28:1047-54.

P

Pang Q, Prolla TA, Liskay RM. 1997. Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. Molecular and cellular biology 17:4465-73.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD and others. 1994. Mutation of a mutL homolog in hereditary colon cancer. Science (New York, N.Y 263:1625-9.

Parsons R, Li GM, Longley MJ, Fang WH, Papadopoulos N, Jen J, de la Chapelle A, Kinzler KW, Vogelstein B, Modrich P. 1993. Hypermutability and mismatch repair deficiency in RER+ tumor cells. Cell 75:1227-36.

Pastrello C, Pin E, Marroni F, Bedin C, Fornasarig M, Tibiletti MG, Oliani C, Ponz de Leon M, Urso ED, Della Puppa L and others. 2011. Integrated analysis of unclassified variants in mismatch repair genes. Genet Med 13:115-24.

Pedroni M, Tamassia MG, Percesepe A, Roncucci L, Benatti P, Lanza G, Jr., Gafa R, Di Gregorio C, Fante R, Losi L and others. 1999. Microsatellite instability in multiple colorectal tumors. Int J Cancer 81:1-5.

Peltomaki P. 2005. Lynch syndrome genes. Fam Cancer 4:227-32.

Peltomaki P, Vasen H. 2004. Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. Disease markers 20:269-76.

Perera S, Bapat B. 2008. The MLH1 variants p.Arg265Cys and p.Lys618Ala affect protein stability while p.Leu749Gln affects heterodimer formation. Hum Mutat 29:332.

Perera S, Li B, Tsitsikotas S, Ramyar L, Pollett A, Semotiuk K, Bapat B. 2010. A novel and rapid method of determining the effect of unclassified MLH1 genetic variants on differential allelic expression. J Mol Diagn 12:757-64.

Perez-Cabornero L, Borras Flores E, Infante Sanz M, Velasco Sampedro E, Acedo Becares A, Lastra Aras E, Cuevas Gonzalez J, Pineda Riu M, Ramon y Cajal Asensio T, Capella Munar G and others. 2011. Characterization of new founder Alu-mediated rearrangements in MSH2 gene associated with a Lynch syndrome phenotype. Cancer Prev Res (Phila) 4:1546-55.

Perez-Carbonell L, Alenda C, Paya A, Castillejo A, Barbera VM, Guillen C, Rojas E, Acame N, Gutierrez-Avino FJ, Castells A and others. Methylation analysis of MLH1 improves the selection of patients for genetic testing in Lynch syndrome. J Mol Diagn 12:498-504.

Perez-Carbonell L, Ruiz-Ponte C, Guarinos C, Alenda C, Paya A, Brea A, Egoavil CM, Castillejo A, Barbera VM, Bessa X and others. 2012. Comparison between universal molecular screening for Lynch syndrome and revised Bethesda guidelines in a large population-based cohort of patients with colorectal cancer. Gut 61:865-72.

Pertea M, Lin X, Salzberg SL. 2001. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res 29:1185-90.

Pineda M, Gonzalez S, Lazaro C, Blanco I, Capella G. 2010. Detection of genetic alterations in hereditary colorectal cancer screening. Mutat Res 693:19-31.

Pineda M, Mur P, Iniesta MD, Borras E, Campos O, Vargas G, Iglesias S, Fernandez A, Gruber SB, Lazaro C and others. 2012. MLH1 methylation screening is effective in identifying epimutation carriers. Eur J Hum Genet

Pinheiro M, Pinto C, Peixoto A, Veiga I, Mesquita B, Henrique R, Baptista M, Fragoso M, Sousa O, Pereira H and others. 2011. A novel exonic rearrangement affecting MLH1 and the contiguous LRRFIP2 is a founder mutation in Portuguese Lynch syndrome families. Genet Med 13:895-902.

Pinol V, Castells A, Andreu M, Castellvi-Bel S, Alenda C, Llor X, Xicola RM, Rodriguez-Moranta F, Paya A, Jover R and others. 2005. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. Jama 293:1986-94.

Plon SE, Cooper HP, Parks B, Dhar SU, Kelly PA, Weinberg AD, Staggs S, Wang T, Hilsenbeck S. 2008. Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. Genet Med 13:148-54.

Plotz G, Welsch C, Giron-Monzon L, Friedhoff P, Albrecht M, Piiper A, Biondi RM, Lengauer T, Zeuzem S, Raedle J. 2006. Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic Acids Res 34:6574-86.

Polaczek P, Putzke AP, Leong K, Bitter GA. 1998. Functional genetic tests of DNA mismatch repair protein activity in Saccharomyces cerevisiae. Gene 213:159-67.

Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, Gulden C, Kupfer SS, Swisher EM, Bennett RL and others. 2012. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J Mol Diagn 14:357-66.

Q

Quehenberger F, Vasen HF, van Houwelingen HC. 2005. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. J Med Genet 42:491-6.

R

Radman M, Wagner R. 1986. Mismatch repair in Escherichia coli. Annu Rev Genet 20:523-38.

Radtke F, Clevers H. 2005. Self-renewal and cancer of the gut: two sides of a coin. Science (New York, N.Y 307:1904-9.

Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lonnqvist KE, Holinski-Feder E, Sutter C, McKinnon W, Duraisamy S and others. 2005. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology 129:537-49.

Rahner N, Friedrichs N, Steinke V, Aretz S, Friedl W, Buettner R, Mangold E, Propping P, Walldorf C. 2008. Coexisting somatic promoter hypermethylation and pathogenic MLH1 germline mutation in Lynch syndrome. J Pathol 214:10-6.

Ramsoekh D, Wagner A, van Leerdam ME, Dooijes D, Tops CM, Steyerberg EW, Kuipers EJ. 2009. Cancer risk in MLH1, MSH2 and MSH6 mutation carriers; different risk profiles may influence clinical management. Hered Cancer Clin Pract 7:17.

Raschle M, Dufner P, Marra G, Jiricny J. 2002. Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. J Biol Chem 277:21810-20.

Raskin L, Schwenter F, Freytsis M, Tischkowitz M, Wong N, Chong G, Narod SA, Levine DA, Bogomolniy F, Aronson M and others. 2011. Characterization of two Ashkenazi Jewish founder mutations in MSH6 gene causing Lynch syndrome. Clin Genet 79:512-22.

Rasmussen LJ, Heinen CD, Royer-Pokora B, Drost M, Tavtigian S, Hofstra RM, de Wind N. 2012. Pathological assessment of mismatch repair gene variants in Lynch syndrome: Past, present, and future. Hum Mutat

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4:311-23.

Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118.

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE, Molecular Subcommittee of the ALQAC. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med 10:294-300.

Ripa RS, Katballe N, Wikman FP, Jager AC, Bernstein I, Orntoft T, Schwartz M, Nielsen FC, Bisgaard ML. 2005. Presymptomatic diagnosis using a deletion of a single

codon in families with hereditary non-polyposis colorectal cancer. Mutat Res 570:89-96.

Risio M, Reato G, di Celle PF, Fizzotti M, Rossini FP, Foa R. 1996. Microsatellite instability is associated with the histological features of the tumor in nonfamilial colorectal cancer. Cancer Res 56:5470-4.

Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L and others. 1997. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 89:1758-62.

Rodriguez-Bigas MA, Lee PH, O'Malley L, Weber TK, Suh O, Anderson GR, Petrelli NJ. 1996. Establishment of a hereditary nonpolyposis colorectal cancer registry. Dis Colon Rectum 39:649-53.

Rogan PK, Faux BM, Schneider TD. 1998. Information analysis of human splice site mutations. Hum Mutat 12:153-71.

Rolnick SJ, Rahm AK, Jackson JM, Nekhlyudov L, Goddard KA, Field T, McCarty C, Nakasato C, Roblin D, Anderson CP and others. 2011. Barriers in identification and referral to genetic counseling for familial cancer risk: the perspective of genetic service providers. J Genet Couns 20:314-22.

S

Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML. 2005. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. Gastroenterology 129:837-45.

Samowitz WS, Albertsen H, Sweeney C, Herrick J, Caan BJ, Anderson KE, Wolff RK, Slattery ML. 2006. Association of smoking, CpG island methylator phenotype, and V600E BRAF mutations in colon cancer. J Natl Cancer Inst 98:1731-8.

Santibanez Koref M, Wilson V, Cartwright N, Cunnington MS, Mathers JC, Bishop DT, Curtis A, Dunlop MG, Burn J. 2010. MLH1 Differential allelic expression in mutation carriers and controls. Annals of human genetics 74:479-88.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57.

Senapathy P, Shapiro MB, Harris NL. 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol 183:252-78.

Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, Lindblom A, Lagerstedt K, Thibodeau SN, Lindor NM and others. 2008. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 135:419-28.

Sharp A, Pichert G, Lucassen A, Eccles D. 2004. RNA analysis reveals splicing mutations and loss of expression defects in MLH1 and BRCA1. Hum Mutat 24:272.

Shcherbakova PV, Kunkel TA. 1999. Mutator phenotypes conferred by MLH1 overexpression and by heterozygosity for mlh1 mutations. Molecular and cellular biology 19:3177-83.

Shimodaira H, Filosi N, Shibata H, Suzuki T, Radice P, Kanamaru R, Friend SH, Kolodner RD, Ishioka C. 1998. Functional analysis of human MLH1 mutations in Saccharomyces cerevisiae. Nat Genet 19:384-9.

Shin YK, Heo SC, Shin JH, Hong SH, Ku JL, Yoo BC, Kim IJ, Park JG. 2004. Germline mutations in MLH1, MSH2 and MSH6 in Korean hereditary non-polyposis colorectal cancer families. Hum Mutat 24:351.

Sinicrope FA, Sargent DJ. 2012. Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. Clin Cancer Res 18:1506-12.

Sjursen W, Haukanes BI, Grindedal EM, Aarset H, Stormorken A, Engebretsen LF, Jonsrud C, Bjornevoll I, Andresen PA, Ariansen S and others. 2010. Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers. J Med Genet 47:579-85.

Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Human molecular genetics 15:2490-508.

Soravia C, van der Klift H, Brundler MA, Blouin JL, Wijnen J, Hutter P, Fodde R, Delozier-Blanchet C. 2003. Prostate cancer is part of the hereditary non-polyposis colorectal cancer (HNPCC) tumor spectrum. Am J Med Genet A 121A:159-62.

Spurdle AB. 2010. Clinical relevance of rare germline sequence variants in cancer genes: evolution and application of classification models. Curr Opin Genet Dev 20:315-23.

Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM. 2008. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 29:1304-13.

Stella A, Surdo NC, Lastella P, Barana D, Oliani C, Tibiletti MG, Viel A, Natale C, Piepoli A, Marra G and others. 2007. Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. Clin Genet 71:130-9.

Stella A, Wagner A, Shito K, Lipkin SM, Watson P, Guanti G, Lynch HT, Fodde R, Liu B. 2001. A nonsense mutation in MLH1 causes exon skipping in three unrelated HNPCC families. Cancer Res 61:7020-4.

Stoffel E, Mukherjee B, Raymond VM, Tayob N, Kastrinos F, Sparr J, Wang F, Bandipalliam P, Syngal S, Gruber SB and others. 2009. Calculation of Risk of Colorectal and Endometrial Cancer Among Patients with Lynch Syndrome. Gastroenterology

Sun S, Greenwood CM, Thiffault I, Hamel N, Chong G, Foulkes WD. 2005. The HNPCC associated MSH2*1906G-->C founder mutation probably originated between 1440 CE and 1715 CE in the Ashkenazi Jewish population. J Med Genet 42:766-8.

Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, Seruca R, Iacopetta B, Hamelin R. 2002. Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology 123:1804-11.

Syngal S, Fox EA, Li C, Dovidio M, Eng C, Kolodner RD, Garber JE. 1999. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing. Jama 282:247-53.

T

Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. 2007. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. Cancer Res 67:4595-604.

Tang R, Hsiung C, Wang JY, Lai CH, Chien HT, Chiu LL, Liu CT, Chen HH, Wang HM, Chen SX and others. 2009. Germ line MLH1 and MSH2 mutations in Taiwanese Lynch syndrome families: characterization of a founder genomic mutation in the MLH1 gene. Clin Genet 75:334-45.

Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 43:295-305.

Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB, Group IUGVW. 2008. In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat 29:1327-36.

Thibodeau SN, French AJ, Cunningham JM, Tester D, Burgart LJ, Roche PC, McDonnell SK, Schaid DJ, Vockley CW, Michels VV and others. 1998. Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Res 58:1713-8.

Thiffault I, Foulkes WD, Marcus VA, Farber D, Kasprzak L, MacNamara E, Wong N, Hutter P, Radice P, Bertario L and others. 2004. Putative common origin of two MLH1 mutations in Italian-Quebec hereditary non-polyposis colorectal cancer families. Clin Genet 66:137-43.

Thompson BA, Goldgar DE, Paterson C, Clendenning M, Walters R, Arnold S, Parsons MT, Michael D, Gallinger S, Haile RW and others. 2012a. A Multifactorial Likelihood Model for MMR Gene Variant Classification Incorporating Probabilities Based on Sequence Bioinformatics and Tumor Characteristics: A Report from the Colon Cancer Family Registry. Hum Mutat

Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, Adzhubey IA, Li B, Bell R, Feng B and others. 2012b. Calibration of Multiple In Silico Tools for Predicting Pathogenicity of Mismatch Repair Gene Missense Substitutions. Hum Mutat

Thompson D, Easton DF, Goldgar DE. 2003. A full-likelihood method for the evaluation of causality of sequence variants from family data. Am J Hum Genet 73:652-5.

Tomsic J, Liyanarachchi S, Hampel H, Morak M, Thomas BC, Raymond VM, Chittenden A, Schackert HK, Gruber SB, Syngal S and others. 2012a. An American founder mutation in MLH1. Int J Cancer 130:2088-95.

Tomsic J, Senter L, Liyanarachchi S, Clendenning M, Vaughn CP, Jenkins MA, Hopper JL, Young J, Samowitz W, de la Chapelle A. 2012b. Recurrent and founder mutations in the PMS2 gene. Clin Genet

Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine MP, Soret J, Tazi J and others. 2008. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat 29:1412-24.

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. 1999. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 96:8681-6.

Toyota M, Issa JP. 1999. CpG island methylator phenotypes in aging and cancer. Semin Cancer Biol 9:349-57.

Trachan T. 1999. Human Molecular Genetics. New York.

Tran H, Degtyareva N, Gordenin D, Resnick MA. 1997. Altered replication and inverted repeats induce mismatch repair-independent recombination between highly diverged DNAs in yeast. Molecular and cellular biology 17:1027-36.

Tran PT, Simon JA, Liskay RM. 2001. Interactions of Exo1p with components of MutLalpha in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 98:9760-5.

U

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R and others. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 96:261-8.

V

Vahteristo P, Ojala S, Tamminen A, Tommiska J, Sammalkorpi H, Kiuru-Kuhlefelt S, Eerola H, Aaltonen LA, Aittomaki K, Nevanlinna H. 2005. No MSH6 germline

mutations in breast cancer families with colorectal and/or endometrial cancer. J Med Genet 42:e22.

van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP and others. 2002. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 111:241-50.

van der Klift H, Wijnen J, Wagner A, Verkuilen P, Tops C, Otway R, Kohonen-Corish M, Vasen H, Oliani C, Barana D and others. 2005. Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). Genes Chromosomes Cancer 44:123-38.

Vasen HF. 2007. Review article: The Lynch syndrome (hereditary nonpolyposis colorectal cancer). Alimentary pharmacology & therapeutics 26 Suppl 2:113-26.

Vasen HF, Mecklin JP, Khan PM, Lynch HT. 1991. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 34:424-5.

Vasen HF, Moslein G, Alonso A, Bernstein I, Bertario L, Blanco I, Burn J, Capella G, Engel C, Frayling I and others. 2007. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). J Med Genet 44:353-62.

Vasen HF, Taal BG, Griffioen G, Nagengast FM, Cats A, Menko FH, Oskam W, Kleibeuker JH, Offerhaus GJ, Khan PM. 1994. Clinical heterogeneity of familial colorectal cancer and its influence on screening protocols. Gut 35:1262-6.

Vasen HF, Watson P, Mecklin JP, Lynch HT. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 116:1453-6.

Vasen HF, Wijnen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, Nagengast FM, Meijers-Heijboer EH, Bertario L, Varesco L and others. 1996. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 110:1020-7.

Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, Bayrak-Toydemir P, Samowitz WS. 2010. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Hum Mutat 31:588-93.

W

Wagner A, Barrows A, Wijnen JT, van der Klift H, Franken PF, Verkuijlen P, Nakagawa H, Geugien M, Jaghmohan-Changur S, Breukel C and others. 2003. Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the MSH2 gene. Am J Hum Genet 72:1088-100.

Walsh MD, Buchanan DD, Walters R, Roberts A, Arnold S, McKeone D, Clendenning M, Ruszkiewicz AR, Jenkins MA, Hopper JL and others. 2009. Analysis of families with Lynch syndrome complicated by advanced serrated neoplasia: the importance of pathology review and pedigree analysis. Fam Cancer 8:313-23.

Wanat JJ, Singh N, Alani E. 2007. The effect of genetic background on the function of Saccharomyces cerevisiae mlh1 alleles that correspond to HNPCC missense mutations. Human molecular genetics 16:445-52.

Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. 2007. Structure of the human MutSalpha DNA lesion recognition complex. Mol Cell 26:579-92.

Warthin A. 1925. The further study of a cancer family. J. Cancer Research 9:279-286.

Watson P, Lynch HT. 2001. Cancer risk in mismatch repair gene mutation carriers. Fam Cancer 1:57-60.

Weinberg R. 2006. The Biology of Cancer. Baltimore, MD: Garland Science.

Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D and others. 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 38:787-93.

Westenend PJ, Schutte R, Hoogmans MM, Wagner A, Dinjens WN. 2005. Breast cancer in an MSH2 gene mutation carrier. Hum Pathol 36:1322-6.

Westphalen AA, Russell AM, Buser M, Berthod CR, Hutter P, Plasilova M, Mueller H, Heinimann K. 2005. Evidence for genetic anticipation in hereditary non-polyposis colorectal cancer. Hum Genet 116:461-5.

Wimmer K, Etzler J. 2008. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? Hum Genet 124:105-22.

Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, Ganiats T, Levin T, Woolf S, Johnson D and others. 2003. Colorectal cancer screening and surveillance: clinical guidelines and rationale-Update based on new evidence. Gastroenterology 124:544-60.

Winawer SJ, Fletcher RH, Miller L, Godlee F, Stolar MH, Mulrow CD, Woolf SH, Glick SN, Ganiats TG, Bond JH and others. 1997. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology 112:594-642.

Wong JJ, Hawkins NJ, Ward RL. 2007. Colorectal cancer: a model for epigenetic tumorigenesis. Gut 56:140-8.

Woods MO, Williams P, Careen A, Edwards L, Bartlett S, McLaughlin JR, Younghusband HB. 2007. A new variant database for mismatch repair genes associated with Lynch syndrome. Hum Mutat 28:669-73.

X

Xiao W, Oefner PJ. 2001. Denaturing high-performance liquid chromatography: A review. Hum Mutat 17:439-74.

Y

Yan H, Papadopoulos N, Marra G, Perrera C, Jiricny J, Boland CR, Lynch HT, Chadwick RB, de la Chapelle A, Berg K and others. 2000. Conversion of diploidy to haploidy. Nature 403:723-4.

Yuan ZQ, Gottlieb B, Beitel LK, Wong N, Gordon PH, Wang Q, Puisieux A, Foulkes WD, Trifiro M. 2002. Polymorphisms and HNPCC: PMS2-MLH1 protein interactions diminished by single nucleotide polymorphisms. Hum Mutat 19:108-13.

Z

Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. Genes & development 18:1241-50.

Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA. 2005a. Exon inclusion is dependent on predictable exonic splicing enhancers. Molecular and cellular biology 25:7323-32.

Zhang Y, Yuan F, Presnell SR, Tian K, Gao Y, Tomkinson AE, Gu L, Li GM. 2005b. Reconstitution of 5'-directed human mismatch repair in a purified system. Cell 122:693-705.

Zhou X, Iversen ES, Jr., Parmigiani G. 2005. Classification of Missense Mutations of Disease Genes. J Am Stat Assoc 100:51-60.

ANNEXES

ANNEX I: Altres publicacions

ALLELE-SPECIFIC EXPRESSION OF *APC* IN ADENOMATOUS POLYPOSIS FAMILIES

Ester Castellsagué, Sara González, Elisabet Guinó, Kristen N. Stevens, **Ester Borràs**, Victoria M. Raymond, Conxi Lázaro, Ignacio Blanco, Stephen B. Gruber Gabriel Capellá.

Gastroenterology:139(2):439-47, 447.e1 (2010)

Resum del treball: Les mutacions germinals en el gen *APC* són la principal causa de poliposi adenomatosa familiar (FAP) i, en una menor proporció, de poliposis adenomatosa familiar atenuada (AFAP). L'anàlisi sistemàtic del gen *APC* a nivell de RNA pot proporcionar informació sobre la patogenicitat de les mutacions trobades i identificar la base molecular en la famílies on no s'han identificat mutacions. En aquest treball, es va analitzar la prevalença dels desequilibris en l'expressió al·lèlica del gen APC en famílies amb poliposi amb mutacions en la línia germinal en el gen APC així com en famílies sense mutacions detectades en *APC* i/o *MUTYH*.

Per realitzar aquest anàlisis d'expressió es va utilitzar la tècnica de *single nucleotide primer extension* (SNuPE) en un polimorfisme del exó 11 del gen *APC*.

De les famílies amb mutació en el gen *APC*, la majoria de les que eren portadores d'una mutació situada abans de l'últim exó del gen, presentaven desequilibri en l'expressió al·lèlica (ASE), això és consistent amb el mecanisme de *nonsense-mediated decay*. De les famílies que no presentaven mutacions en *APC/MUTYH*, el 9% presentaven desequilibri en ASE, que podrien ser responsables de la malaltia. L'expressió normal de l'al·lel era restaurada després de cultivar el limfòcits en puromicina, recolzant la hipòtesis del *nonsense-mediated decay.*

Així doncs, l'anàlisi de ASE pot ser utilitzat per determinar la patogenicitat d'alguns casos de FAP i AFAP on no s'han identificat mutacions en el gen APC. A més, l'ASE també pot ser utilitzat per prioritzar l'ordre en que s'analitzin les diferents zones del gen en la cerca de mutacions. En conclusió, els estudis a nivell de RNA són importants en el diagnòstic molecular de la FAP.

GASTROENTEROLOGY 2010;139:439–447

Allele-Specific Expression of *APC* in Adenomatous Polyposis Families

ESTER CASTELLSAGUÉ,* SARA GONZÁLEZ,* ELISABET GUINÓ,‡ KRISTEN N. STEVENS,§ ESTER BORRÀS,* VICTORIA M. RAYMOND,‖ CONXI LÁZARO,* IGNACIO BLANCO,¶ STEPHEN B. GRUBER,# and GABRIEL CAPELLÁ*

*Translational Research Laboratory, ‡Bioinformatics and Biostatistics Unit, Department of Epidemiology, ¶Cancer Genetic Counseling Program, Institut d'Investigacions Biomédiques de Bellvitge–Institut Català d'Oncologia, Barcelona, Spain; §Department of Epidemiology, #Department of Internal Medicine, Epidemiology, and Human Genetics, Division of Molecular Medicine and Genetics, University of Michigan, Ann Arbor, Michigan; and ‖Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan

See related articles, Rozas DA et al, on page e85; and Ruys AT et al, on page 731 in *CGH*.

BACKGROUND & AIMS: Germline mutations in the *APC* gene cause of most cases of familial adenomatous polyposis (FAP) and a lesser proportion of attenuated FAP (AFAP). Systematic analysis of *APC* at the RNA level could provide insight into the pathogenicity of identified mutations and the molecular basis of FAP/AFAP in families without identifiable mutations. Here, we analyzed the prevalence of imbalances in the allelic expression of *APC* in polyposis families with germline mutations in the gene and without detectable mutations in *APC* and/or *MUTYH*. **METHODS:** Allele-specific expression (ASE) was determined by single nucleotide primer extension using an exon 11 polymorphism as an allele-specific marker. In total, 52 *APC*-mutation–positive (36 families) and 24 *APC/MUTYH*-mutation–negative (23 families) informative patients were analyzed. Seventy-six controls also were included. **RESULTS:** Of the *APC*-mutation–positive families, most of those in whom the mutation was located before the last exon of the gene (12 of 14) had ASE imbalance, which is consistent with a mechanism of nonsense-mediated decay. Of the *APC/MUTYH*-mutation–negative families, 2 (9%) had ASE imbalance, which might cause the disease. Normal allele expression was restored shortly after lymphocytes were cultured with puromycin, supporting a 'nonsense-mediated' hypothesis. **CONCLUSIONS: ASE analysis might be used to determine the pathogenesis of some cases of FAP and AFAP in which *APC* mutations are not found. ASE also might be used to prioritize the order in which different areas of *APC* are tested. RNA-level studies are important for the molecular diagnosis of FAP.**

Keywords: Allele-Specific Expression; Nonsense-Mediated Decay.

Familial adenomatous polyposis (FAP), an autosomal-dominant disease predisposing to colorectal cancer, is caused mainly by truncating germline mutations in the *APC* gene (5q21–22).[1] Widespread use of sequencing techniques has led to the identification of an increased number of missense mutations,[2] variants with a putative impact on messenger RNA (mRNA) splicing,[3] and cases of somatic and germline mosaicisms.[4] Exon dose analysis has detected gross rearrangements in a minority of cases.[5] However, in a lesser proportion of cases, attenuated FAP (AFAP) also is associated with *APC* mutations and with biallelic germline mutations in the *MUTYH* gene, showing an autosomal-recessive pattern.[6]

In other cancer predisposition genes such as *MSH2*, *MLH1*, *BRCA1*, *BRCA2*, and *NF1*, studies at the RNA level have shown that mutations causing a premature termination codon (PTC) usually trigger nonsense-mediated decay (NMD) of the mRNA.[7–11] This mRNA surveillance mechanism reduces the abundance of premature stop-codon–harboring mRNA and of the corresponding truncated proteins. NMD of the affected transcript occurs if the PTC is located approximately 55 base pairs upstream of the last intron-exon boundary.[12] Some RNA studies of the *APC* gene have been published, focused mainly on the presence of aberrant splicing caused by missense, silent, or unclassified variants.[3,13]

The proportion of FAP families with identifiable *APC* mutations has shown a slight increase with the incorporation of new techniques into diagnostic algorithms. However, a large subset has undetectable pathogenic changes (designated here as *APC*[−]/*MUTYH*[−]). Small decreases in *APC* mRNA have been detected in *APC*(−)[14,15] and *APC*(−)/*MUTYH*(−) families.[16–18] In contrast, high germline levels of an *APC* mRNA isoform resulting from an exon 10–15 connection have been observed in a case of *APC*(−)/*MUTYH*(−) AFAP.[19] Germline imbalances in allele-specific expression (ASE) of the *APC* gene have been detected in FAP and AFAP

Abbreviations used in this paper: AFAP, attenuated familial adenomatous polyposis; ASE, allele-specific expression; cDNA, complementary DNA; dNTP, deoxynucleoside triphosphate; F-UM, family code of the University of Michigan; FAP, familial adenomatous polyposis; ICO, Catalan Institute of Oncology; NMD, nonsense-mediated decay; PCR, polymerase chain reaction; PTC, premature termination codon; SD, standard deviation; SNP, single nucleotide polymorphism; UM, University of Michigan.

0016-5085/$36.00
doi:10.1053/j.gastro.2010.04.047

families,[14,15,18,20] but their potential contribution to diagnostic yield is unknown.

The prevalence of imbalances in *APC* allelic expression was analyzed in 2 sets of polyposis families: *APC* mutation-positive, designated here as *APC*(+), and *APC*(−)/*MUTYH*(−). Here we show that mutations outside exon 15 are associated with a clear allelic imbalance at the RNA level that is secondary to NMD, and that these imbalances also are detected in a low proportion of *APC*(−)/*MUTYH*(−) families.

Patients and Methods

Patients

A total of 127 FAP and AFAP families from the Catalan Institute of Oncology (ICO) and the University of Michigan (UM) initially were tested for germline *APC* and *MUTYH* mutations by comprehensive diagnostic methods (sequencing or single-strand conformation polymorphism-analysis of all exons and exon-intron boundaries, gross deletion analysis by quantitative multiplex polymerase chain reaction [PCR] of short fluorescent fragments, and/or multiplex ligation-dependent probe amplification, and/or Southern blot for the *APC* gene, presence of germline epimutations in *APC* promoter in conjunction with sequencing of the 2 most prevalent variants of the *MUTYH* gene—Y165C and G382D). Mutations were found in 76 families, although the molecular pathogenic cause of the disease remained unknown in 51 families. The mutation detection rates were 83% for FAP and 25% for AFAP.

Individuals from 59 of the families tested heterozygous for the rs2229992 *APC* coding single nucleotide polymorphism (SNP) in exon 11 of the *APC* gene and were the informative population of the present study. These families were divided according to mutational status as follows: (1) 36 *APC*(+) families (24 classic FAP, 34 carrier individuals; 12 AFAP, 18 carrier individuals); and (2) 23 *APC*(−)/*MUTYH*(−) families (5 FAP, 6 individuals; 18 AFAP, 18 individuals).

A total of 76 heterozygous controls were included: 29 noncarriers from *APC* mutation families, 6 carriers and 24 noncarriers from *MSH2* and *MLH1* mutation families, and 17 Ashkenazi Jewish individuals diagnosed with sporadic microsatellite-unstable colorectal cancer that formed part of the Molecular Epidemiology of Colorectal Cancer study. This was a population-based, case-control study of incident colorectal cancer, including histopathologically confirmed cases from all incident colorectal cancer cases diagnosed in northern Israel from March 31, 1998, onward. Informed consent was obtained from all of the subjects who participated in the study.

Genotyping: Analysis of Genomic DNA

Peripheral blood lymphocyte DNA was obtained using the FlexiGene DNA kit (QIAGEN, Hilden, Germany). A total of 25 ng of genomic DNA was amplified in a final volume of 25 μL containing 0.13 μmol/L primers, 0.2 mmol/L deoxynucleoside triphosphate (dNTP), 2.5 mmol/L $MgCl_2$, and 1.25 U *Taq* polymerase (Thermoprime Plus DNA Polymerase; ABgene, Rochester, NY). After an initial denaturing step, samples underwent 5 initial cycles (1 min at 94°C, 30 s at 65°C, and 30 s at 72°C), 20 cycles (1 min at 94°C, 30 s at 65°C decreasing 0.5°C/cycle, and 30 s at 72°C), and 10 final cycles (1 min at 94°C, 30 s at 55°C, and 30 s at 72°C). Primers in exon 11 (forward: 5′-GGGACTACAGGCCATTGCA-3′, and reverse: 5′-CAAGTTTGTCAAAGCCATTCCAGC-3′) were used to amplify the rs2229992 SNP. To remove unincorporated primers and dNTPs, PCR fragments were purified using illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Boston, MA). For the single nucleotide primer extension reaction, primer extension was performed with the SNaPshot Multiplex Kit (Applied Biosystems, Carlsbad, CA) with 5′-ATTGCAAGTGGACTGTGAAATGTA-3′ according to the manufacturer's instructions. Briefly, reactions were performed in a total volume of 10 μL containing 1.5 μL treated PCR product, 4.5 μL SNaPshot Ready Reaction Mix, and 0.2 μmol/L extension primer. Primer extension thermocycling conditions consisted of 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 30 seconds. SNaPshot reaction products were treated with 1 U shrimp alkaline phosphatase (USB Corporation, Cleveland, OH) for 60 minutes at 37°C and then 15 minutes at 75°C. Products were run in an ABI Prism 3130 DNA sequencer and analyzed by GeneMapper v4.0 (Applied Biosystems). Heterozygous samples showed a profile with 2 peaks (black and red peaks represent C and T alleles, respectively), whereas only one peak was observed for homozygous samples.

Measuring ASE: Analysis of Complementary DNA

Total RNA was isolated from frozen lymphocytes using Trizol Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. A total of 250 ng of RNA were reverse-transcribed into complementary DNA (cDNA) using random hexa-nucleotide mixture primers and MMLV reverse transcriptase (Invitrogen). To specifically amplify rs2229992 SNP in cDNA, we used the same exon 11 forward primer as for DNA amplification and a reverse primer targeting the exon 11–12 junction (5′-ATAGAGCATAGCGTAGCCTTGTTG -3′). PCR reactions were performed in a final volume of 25 μL containing 2 μL of cDNA, 0.2 μmol/L primers, 0.2 mmol/L dNTPs, 2.5 mmol/L $MgCl_2$, and 1.25 U *Taq* polymerase (Thermoprime Plus DNA Polymerase; ABgene). After a denaturing step, 35 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C were performed. The remaining steps were the same as described for genotyping, including purification, SNaP-

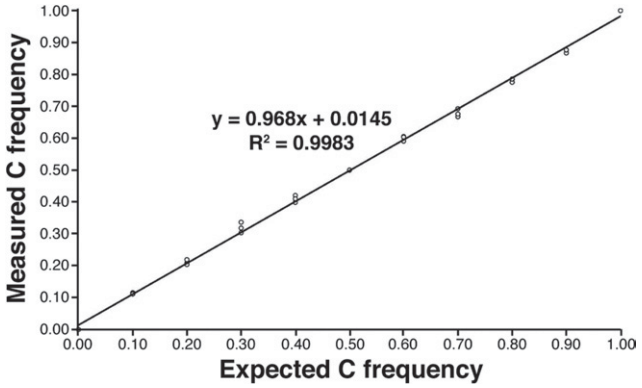


Figure 1. SNaPshot analysis of rs2229992 SNP as a quantitative test. Regression analysis of the amplification products of 2 homozygous control cDNAs mixed in different proportions and analyzed in triplicate by SNaPshot. Allele frequencies were calculated by the peak heights: freq. C = C/(C + *k*T), where the correction factor *k* is determined from the mix simulating an actual allele frequency of 0.5 (5:5). Measured allele frequencies were plotted against the expected values. A near-linear relationship was observed across all data points (R2 = 0.9983).

shot reaction, phosphatase treatment, and capillary electrophoresis.

ASE was measured using peak intensities in heterozygous samples. Allelic frequencies were calculated as frequency C = C/(C + *k*T) or freq. T = T/(T + *k*'C), where *k* and *k*' are constants given by the mean of C/T (*k*) and T/C (*k*') proportions in control samples. These constants were used to correct for unequal representation of alleles in known control heterozygotes caused by differential PCR amplification of alleles, differential efficiencies of dideoxynucleoside triphosphates incorporation in the extension reaction, unequal emission energies of fluorescent dyes,[21] or putative differences in physiologic RNA levels owing to alternative splicing affecting exons 1, 7, 9, 10A, and 14 of the *APC* gene.[22–27] ASE values are expressed as the proportion of frequencies of the 2 alleles (freq. C:freq. T). Three independent replicates of all experiments were obtained, and a set of controls was included in every experiment. A Mann–Whitney test was used to evaluate ASE differences among groups.

To validate the SNaPshot analysis as a quantitative technique, 2 homozygous samples representing the 2 alleles were selected. After reverse transcription, cDNAs were mixed in different proportions (0:10, 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1, and 10:0) and analyzed in triplicate. As seen in Figure 1, a SNaPshot assay of rs2229992 SNP provides a quantitative measurement of ASE of the *APC* gene with an analytic sensitivity of at least 10% (Figure 1). To set the cut-off points for equal and unequal expression we used 14 controls from the ICO subset of samples and obtained a mean value of 1.002 ± 0.055 (Figure 2*A*). From these results we established a conservative range for normal ASE values of 0.836–1.168 with a confidence of 99% (mean values ± 3 standard deviations). These values were validated independently in an additional set of 14 controls with a range for normal values of 0.732–1.198 (mean values ± 3 standard deviations), which has been replicated consistently in independent experiments using different control sets. In the presence of modest interexperiment variability, a set of control samples was always run in the same experiment. Pyrosequencing was used to confirm the range of

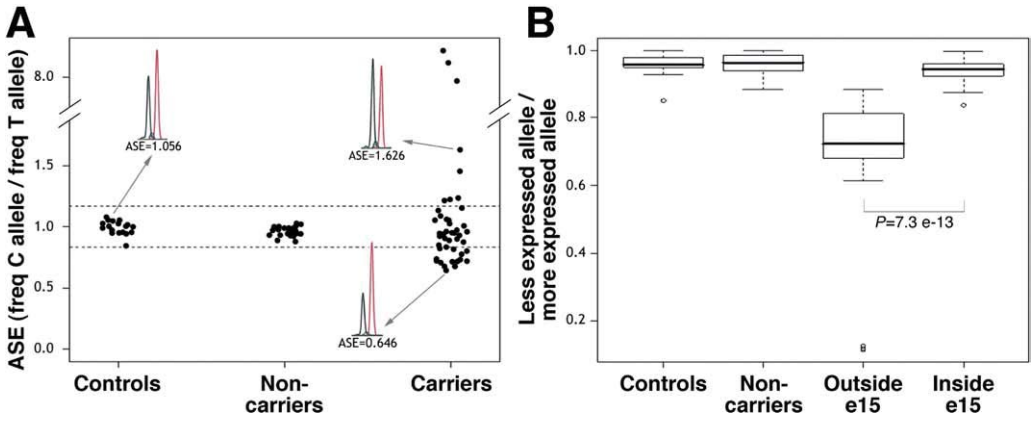


Figure 2. Quantification of ASE in *APC*(+) polyposis patients. (*A*) Distribution of *APC* ASE in controls, noncarriers, and carriers of a pathogenic *APC* mutation from the ICO and UM subsets. ASE values are the proportion between the C and T allele frequencies of the *APC* coding SNP rs2229992. Each point represents the mean of the triplicate measurements, and each replicate was performed for all samples in each subset. Allelic ratios ranged from 0.646 to 8.706. The normal ASE cut-off range is indicated by dashed horizontal lines (at 1.168 and 0.836) and determined by control ASE values, as indicated in the Patients and Methods section. Three sample electropherograms obtained from the SNaPshot reaction, and their corresponding ASE values, also are shown. (*B*) Box plots of ASE values for the location of the mutation in the *APC* gene. Patients were stratified into 2 groups by mutation site (OUTSIDE e15, mutation located from exons 1-14; INSIDE e15, mutation located in exon 15). To give an overview of our results, ASE values are represented as the proportion of allelic frequencies between the underexpressed and the overexpressed allele in each sample, independently of whether it is C or T. The interquartile range includes 50% of the samples and is shown by white boxes. The interdecile range includes 90% of the patients and is shown as whiskers. Outliers are shown by empty dots. Mutated groups were compared using the nonparametric Mann–Whitney test.

normal ASE values in an independent set of 18 controls. Briefly, germline cDNA and germline DNA were PCR-amplified using the forward primer (GAATTATTGCAAGTGGACTGTGAA) and the reverse biotinylated primer (GCCATTCCAGCATATCGTCTTA) to yield an 83–base pair amplicon containing rs2229992. The PCR reaction mixture (25 μL) contained 5 ng of genomic DNA or 1 μL of cDNA, 2.5 μL of 10× PCR buffer (Applied Biosystems), 2 μL of 25 mmol/L $MgCl_2$ (Applied Biosystems), 1.25 μL of 2.5 mmol/L dNTP (New England Biolabs, Ipswich, MA), 0.5 μL of each 10 μmol/L primer, and 0.75 U of AmpliTaq Gold DNA polymerase (Applied Biosystems). Cycling conditions were as follows: initial denaturation at 95°C for 3 minutes, 50 cycles of 95°C for 34 seconds, 60°C for 45 seconds, 72°C for 45 seconds, and a final extension at 72°C for 10 minutes. A total of 5 μL of the resulting PCR product was used for pyrosequencing with the sequencing primer (GCAAGTGGACTGTGAAAT) according to the standard Qiagen protocol. Normal ASE values were similar using this technique, ranging from 0.732 to 1.198 (mean values ± 3 standard deviations).

Puromycin Analysis

Frozen lymphocytes from a subset of patients and controls were short-cultured in RPMI 1640 + GlutaMAX medium (10% fetal bovine serum, 1% penicillin-streptomycin) (Invitrogen) and phytohemagglutinin (Sigma-Aldrich, St. Louis, MO). After 6–9 days, 2 subcultures were derived: one was treated with 250 μg/mL puromycin (Sigma-Aldrich) for 5 hours, and the other was left untreated. RNA was extracted and ASE was calculated as described earlier.

APC cDNA Molecular Analysis and Promoter Analyses

APC cDNA molecular analysis was performed in all *APC*(−)/*MUTYH*(−) families showing ASE imbalance. Six overlapping primer pairs covering all exon-exon boundaries of the *APC* gene were designed to test for abnormal cDNA products in these families. Differential running patterns between test samples and controls were assessed visually using agarose gel electrophoresis and observed abnormal bands were sequenced. Sequencing of the 1A and 1B promoter regions of the *APC* gene also was performed. Finally, *APC* gene expression levels were analyzed in all ICO samples by quantitative real-time PCR using a primer pair targeting exons 2 and 3. The primer sequences and experimental conditions for these analyses are detailed in Supplementary Table 1.

Results

ASE Analysis in APC(+) *Polyposis Families*

ASE was assessed in 52 individuals (34 FAP, 18 AFAP) who were *APC*(+) from 36 families (24 FAP, 12 AFAP) that tested heterozygous at rs2229992, in parallel with noncarrier individuals and controls. To categorize ASE, a cut-off range was calculated as described in the Patients and Methods section. ASE values less than 0.836 designated C-allele underexpression and values greater than 1.168 designated C-allele overexpression, in both cases depicting allelic imbalance. The average and standard deviation ASE values for noncarriers (n = 24) were 0.964 ± 0.038 (Figure 2*A*), with none outside the normal range. Twenty-two families (24 individuals) showed normal ASE values (range, 0.876–1.155). The remaining 14 (39%) showed ASE imbalances: 5 families (8 individuals) showed C-allele overexpression (range, 1.214–8.706), 8 families (18 individuals) showed C-allele underexpression (range, 0.677–0.830), and 1 family showed the borderline value of 0.838. Another family, harboring a deletion from the promoter to exon 4, showed complete loss of T-allele expression (family code of the University of Michigan [F-UM]-8) and was used as an internal control (Figure 2*A*; Table 1).

In accordance with the canonical model of NMD, a strong correlation was observed between the location of the mutation outside exon 15 (leading to a premature termination codon before the last exon of the gene) and the presence of ASE imbalance. Twelve of 14 (86%) families with mutations outside exon 15 showed abnormal ASE, whereas the proportion decreased to 2 of 22 (9%) for families with exon 15 mutations (P = 7.3e-13, obtained from ASE values) (Table 1, Figure 2*B*). Interestingly, the 4 cases that did not fit with the model showed borderline ASE values. Thus, the ASE values for the 2 families with mutations outside exon 15 and with no imbalance (F-UM-2 and F-UM-9) were 1.155 and 1.133 (upper cut-off value, 1.168). The 2 cases with exon 15 mutations and classified as harboring putative imbalances (F-8 and F-9) showed ASE values of 0.830 and 0.838 (lower cut-off value, 0.836). F-8 harbors a mutation at intron 14 that is predicted to alter intron-exon processing, leading to a truncated protein at exon 15. F-9 has a recurrent mutation, p.Gln1062X, that was not associated with ASE in other cases in our series sharing the same alteration (F-13 and F-22). However, F-9 was classified as putatively abnormal because the imbalance essentially coincided with the diagnostic threshold value and was confirmed in cultured lymphocytes (Table 1).

Variations in ASE imbalance depending on the location of the mutation associate with different *APC* expression levels. Values of *APC* germline expression in 43 controls were 3.793 ± 2.933. Expression levels in polyposis with exon 15 mutations were similar (3.647 ± 0.911), whereas expression levels in polyposis with mutations outside exon 15 were lower (2.887 ± 0.867; P = .014).

More than one member (range, 2–5) was analyzed in 8 FAP and AFAP families, and all carriers showed concordant ASE values in 7 of these families (Table 1). In 1 family (F-2), 2 of the 3 carriers analyzed showed abnor-

Table 1. Characteristics of the *APC*(+) Polyposis Families Analyzed for ASE

Individual	Family	Phenotype	Mutation[a]	Exon or intron	Type of mutation	Reference	ASE ± SD (freq C/freq T)	ASE (−)Puromycin (freqC/freqT)	ASE (+)Puromycin (freqC/freqT)
41	8	FAP	c.1958+3A>G	i14[b]	Splicing	2,3	**0.830 ± 0.036**		
UM-6	UM-6	AFAP	c.2004del; p.Leu669X	e15	Frameshift	This study	0.911 ± 0.250		
78	19	FAP	c.2116del; p.Met706CysfsX12	e15	Frameshift	This study	0.964 ± 0.010		
61	15	FAP	c.2397T>A; p.Tyr799X	e15	Nonsense	3	0.876 ± 0.018		
38	14	FAP	c.2701C>T; pGln901X	e15	Nonsense	This study	1.062 ± 0.025		
39	14	FAP	c.2701C>T; pGln901X	e15	Nonsense	This study	0.949 ± 0.010		
25	11	FAP	c.2727del; p.Thr910LeufsX6	e15	Frameshift	This study	0.964 ± 0.032		
127	27	AFAP	c.3077A>G, p.Asn1026Ser	e15	Missense	36	0.974 ± 0.049		
97	9	FAP	c.3183_3187del; p.Gln1062X	e15	Frameshift	1,2,3	**0.838 ± 0.030**	0.773	1.090
37	13	FAP	c.3183_3187del; p.Gln1062X	e15	Frameshift	1,2,3	1.055 ± 0.093		
100	22	FAP	c.3183_3187del; p.Gln1062X	e15	Frameshift	1,2,3	0.907 ± 0.042		
33	12	FAP	c.3562del; p.Pro1188LeufsX77	e15	Frameshift	This study	0.924 ± 0.073		
124	26	AFAP	c.3631A>G; p.Met1211Val	e15	Missense	This study	0.892 ± 0.050		
UM-4	UM-4	FAP	c.3688C>T; p.Gln1230X	e15	Nonsense	1,2,3	1.015 ± 0.121		
66	16	FAP	c.3838_3839del; p.Leu1280ValfsX4	e15	Frameshift	This study	0.944 ± 0.039		
77	18	FAP	c.3927_3931del; p.Glu1309AspfsX4	e15	Frameshift	1,2,3	0.950 ± 0.023		
76	17	FAP	c.4189G>T; p.Glu1397X	e15	Nonsense	1,3	0.943 ± 0.013		
81	20	FAP	c.4393_4394del; p.Ser1465TrpfsX3	e15	Frameshift	1,2,3	0.932 ± 0.045		
96	21	FAP	c.4393_4394del; p.Ser1465TrpfsX3	e15	Frameshift	1,2,3	0.951 ± 0.039		
23	10	FAP	c.4608_4614del; p.Glu1536AspfsX27	e15	Frameshift	This study	1.028 ± 0.016		
24	10	FAP	c.4608_4614del; p.Glu1536AspfsX27	e15	Frameshift	This study	1.009 ± 0.049		
UM-7	UM-7	AFAP	c.4782_4785del; p.Ala1595ArgfsX54	e15	Frameshift	This study	0.929 ± 0.029		
UM-1	UM-1	AFAP	c.5936_5939del; p.Asn1979ThrfsX64	e15	Frameshift	1	1.054 ± 0.039		
UM-10	UM-10	AFAP	c.5936_5939del; p.Asn1979ThrfsX64	e15	Frameshift	1	1.086 ± 0.099		
UM-8	UM-8	AFAP	Deletion promoter–e4	e1–e4	Gross deletion	This study	**Loss of T**		
54	1	FAP	c.423-2A>G	i3	Splicing	This study	**0.738 ± 0.031**	0.621	0.997
161	2	FAP	c.423-2A>G	i3	Splicing	This study	**0.823 ± 0.069**		
163	2	FAP	c.423-2A>G	i3	Splicing	This study	0.853 ± 0.028		
165	2	FAP	c.423-2A>G	i3	Splicing	This study	**0.717 ± 0.019**	0.497	0.952
UM-2	UM-2	AFAP	c.426_427del; p.Leu143AlafsX4	e4	Frameshift	2	1.155 ± 0.061		
UM-3	UM-3	AFAP	c.426_427del; p.Leu143AlafsX4	e4	Frameshift	2	**1.214 ± 0.020**		
UM-9	UM-9	AFAP	c.464_465del; p.Lys155ArgfsX12	e4	Frameshift	This study	1.133 ± 0.043		
46	3	FAP	c.607C>T; p.Gln203X	e5	Nonsense	1	**1.626 ± 0.008**		
47	3	FAP	c.607C>T; p.Gln203X	e5	Nonsense	1	**1.451 ± 0.049**	2.164	1.083
116	23	AFAP	c.834+1G>A	i7	Splicing	2,3	**0.680 ± 0.027**		
117	23	AFAP	c.834+1G>A	i7	Splicing	2,3	**0.677 ± 0.006**		
118	23	AFAP	c.834+1G>A	i7	Splicing	2,3	**0.728 ± 0.011**		
119	23	AFAP	c.834+1G>A	i7	Splicing	2,3	**0.713 ± 0.006**		
122	23	AFAP	c.834+1G>A	i7	Splicing	2,3	**0.734 ± 0.021**		
10	4	FAP	c.834+1G>A	i7	Splicing	2,3	**0.682 ± 0.029**		
11	4	FAP	c.834+1G>A	i7	Splicing	2,3	**0.722 ± 0.029**		
12	4	FAP	c.834+1G>A	i7	Splicing	2,3	**0.646 ± 0.024**		
13	4	FAP	c.834+1G>A	i7	Splicing	2,3	**0.687 ± 0.042**		
14	4	FAP	c.834+1G>A	i7	Splicing	2,3	**0.725 ± 0.014**		
18	5	FAP	c.1262_1263delinsAA; p.Trp421X	e9	Nonsense	This study	**0.779 ± 0.022**		
21	5	FAP	c.1262_1263delinsAA; p.Trp421X	e9	Nonsense	This study	**0.721 ± 0.024**	0.525	1.069
106	24	AFAP	c.1557_1561del; p.Cys520TyrfsX15	e12	Frameshift	This study	**7.911 ± 0.337**	13.800	6.353
107	24	AFAP	c.1557_1561del; p.Cys520TyrfsX15	e12	Frameshift	This study	**8.706 ± 0.468**		
108	24	AFAP	c.1557_1561del; p.Cys520TyrfsX15	e12	Frameshift	This study	**8.387 ± 0.882**		
105	25	AFAP	c.1699G>T; p.Gly567X	e13	Nonsense	This study	**0.804 ± 0.022**	0.692	1.090
63	6	FAP	c.1660C>T; p.Arg554X	e13	Nonsense	1,2,3	**1.238 ± 0.047**	2.152	1.069
32	7	FAP	c.1660C>T; p.Arg554X	e13	Nonsense	1,2,3	**1.223 ± 0.039**	2.255	1.153

NOTE. Imbalanced ASE values are depicted in bold letters. For further details about the mutations reported we refer to 3 *APC* gene mutation databases: (1) the Universal Mutation Database http://www.umd.be/APC/; (2) http://www.LOVD.nl/APC; and (3) http://www.hgmd.cf.ac.uk/ac/index.php.

SD, standard deviation; e, exon; i, intron.

[a]Reference accession number: NM_000038.[2]

[b]Considered as exon 15 mutation because it leads to a PTC in exon 15.

mal ASE values (0.823 and 0.717) and the third carrier showed a normal borderline value (0.853), which probably reflects a degree of variability between individuals. ASE values were similar among families with the same mutation. For example, in families F-4 (FAP) and F-23 (AFAP), carriers of the same splicing mutation (c.834+1G>A), ASE values were similar, although they did not account for phenotype differences. Allelic imbal-

ance cosegregated with the disease in all cases analyzed (data not shown).

In 9 of the allelic imbalance cases, short-term lymphocyte cultures showed clearer ASE imbalance than observed in cDNA from peripheral blood lymphocytes. Puromycin treatment completely reverted imbalance in 8 of the 9 cases. In the remaining case, only partial reversion was achieved (Table 1). Cultured control lymphocytes showed normal ASE values that remained unchanged after puromycin treatment (data not shown).

ASE Analysis in APC(−)/MUTYH(−) *Polyposis Families*

Twenty-three (5 FAP, 18 AFAP) *APC*(−)/*MUTYH*(−) families (6 and 18 individuals, respectively) were heterozygous for the SNP at exon 11 and subsequently were analyzed. Mean ASE values in controls were 1.001 ± 0.047. Initially, 3 of the 23 families showed abnormal *APC* ASE values (F-29, F-39, and UM-11) (Figure 3 and Table 2). Puromycin treatment of short-term cultured lymphocytes was performed in the 2 families from whom lymphocytes were available (F-29 and F-39) (Table 2). In family F-29, the observed imbalance was more intense in cultured cells than in peripheral blood lymphocyte and was completely reverted by puromycin treatment, as expected. In F-39, with an ASE value of 0.819, the imbalance was no longer observed in cultured lymphocytes.

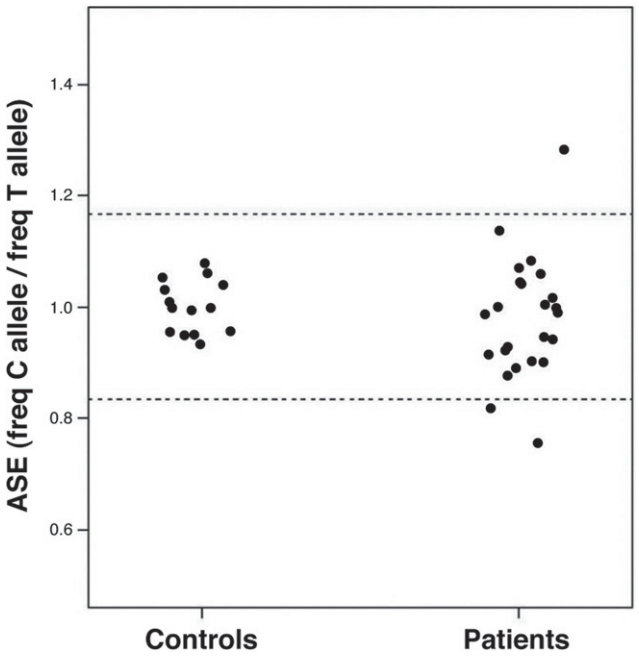


Figure 3. Quantification of ASE in *APC*(−)/*MUTYH*(−) polyposis patients. Distribution of *APC* ASE in controls and polyposis patients with no previously identified mutation. ASE values are the proportion between the C and T allele frequencies of the *APC* coding SNP rs2229992. Each point represents the mean of the triplicate measurements, and each replicate was performed for all samples in each subset. Allelic ratios ranged from 0.756 to 1.284. The normal ASE cut-off range is indicated by dashed horizontal lines (at 1.168 and 0.836).

Thus, 2 of the 23 (9%) *APC*(−)/*MUTYH*(−) families in our series (F-29 and UM-11) harbored ASE imbalance. In F-29, 2 point mutations in the promoter region were identified: g.[112043282C>G(+) 112072889C>T]. These changes were not detected in a small set of 9 controls. However, they were not located in a conserved sequence or a sequence consensus region for transcription factors, and were not associated with clear changes in *APC* expression levels as assessed by quantitative real-time PCR. At the RNA level, the wild type transcript coexisted with a second transcript containing a 60–base pair insertion r.[=, 1408_1409ins1408+1315_1408+1369] in UM-11 RNA. These transcripts also were present in 3 of 5 additional samples consisting of either controls or FAP patients with no ASE imbalance. Interestingly, the UM-11 family did not express a transcript skipping exon 11 that was present in other samples (data not shown). Finally, germline *APC* expression levels were lower in *APC*(−)/*MUTYH*(−) polyposis families (2.963 ± 1.21) than in controls ($P = .04$) and/or *APC*(+) polyposis families with exon 15 mutations ($P = .01$).

Discussion

In this study we adapted the single nucleotide primer extension methodology to perform a detailed allele-specific expression analysis of the *APC* gene in 59 FAP and AFAP families divided into 2 groups: those harboring a pathogenic mutation and those with no identified *APC* or *MUTYH* mutation.

In the presence of a detectable deleterious *APC* gene mutation, ASE imbalance was associated strongly with mutations located outside exon 15. Conversely, balanced ASE was common in cases with exon 15 mutations. Both observations fit, albeit imperfectly, with the nuclear scanning model of NMD in mammals.[10,12] Our observations are in disagreement with the findings of Renkonen et al,[20] who reported no ASE imbalance in a small series of 4 families harboring mutations in exons 6–9 of the *APC* gene, possibly owing to differences in technique or sample processing. To assess the contribution of NMD to the observed imbalance, lymphocytes were short-term cultured and treated with puromycin, a known inhibitor of NMD.[28] Similarly to previous observations, enhanced allelic expression imbalance was observed in vitro before puromycin treatment, pointing to higher degradation of nonsense transcripts associated with culture conditions.[29] Puromycin treatment completely restored normal ASE in most of the cases analyzed. However, only partial reversion was achieved in the case showing the greatest imbalance before treatment (F-24, I-106), and it is unclear whether this can be attributed to the intensity of allelic imbalance, the limitations of the in vitro assay, or other unknown causes.

As has been reported for *MSH2*, *MLH1*, *BRCA1-2*, and *NF1*,[7–11] the nuclear scanning NMD model loosely fits with our observations for the *APC* gene. One of the

Table 2. Characteristics of the *APC*(−)/*MUTYH*(−) Polyposis Families Analyzed for ASE

Individual	Family	Phenotype	ASE ± SD (freq C/freq T)	ASE (−)Puromycin (freqC/freqT)	ASE (+)Puromycin (freqC/freqT)
158	29	FAP	**1.284 ± 0.060**	2.099	0.963
166	30	FAP	1.083 ± 0.047		
170	31	FAP	1.059 ± 0.136		
171	32	FAP	1.045 ± 0.055		
172	32	FAP	1.042 ± 0.011		
175	33	FAP	0.890 ± 0.030		
129	34	AFAP	0.947 ± 0.034		
132	35	AFAP	0.903 ± 0.041		
134	36	AFAP	1.016 ± 0.066		
136	37	AFAP	0.922 ± 0.027		
138	38	AFAP	0.988 ± 0.019		
140	39	AFAP	**0.819 ± 0.024**	0.989	0.998
145	40	AFAP	0.991 ± 0.025		
147	41	AFAP	1.005 ± 0.033		
152	42	AFAP	0.901 ± 0.047		
153	43	AFAP	0.929 ± 0.015		
UM-11	UM-11	AFAP	**0.756 ± 0.122**		
UM-12	UM-12	AFAP	0.943 ± 0.024		
UM-14	UM-14	AFAP	0.914 ± 0.034		
UM-15	UM-15	AFAP	1.002 ± 0.074		
UM-16	UM-16	AFAP	1.138 ± 0.026		
UM-18	UM-18	AFAP	1.070 ± 0.061		
UM-19	UM-19	AFAP	0.877 ± 0.029		
UM-20	UM-20		0.999 ± 0.073		

NOTE. Imbalanced ASE values are depicted in bold letters.
SD, standard deviation.

exceptions is a mutation at intron 14 that creates a skipping of exon 14 and a PTC at the very beginning of exon 15 (data not shown) that associates with imbalanced ASE. This may be owing to the special position of the PTC and its sequence context, both of which are factors known to influence the extent of NMD.

NMD can be considered a modifier of the phenotypic consequences of PTC and has contributed to our understanding of genotype–phenotype correlations in various genetic disorders.[30] In mutations outside exon 15, NMD may promote partial loss of function or haploinsufficiency. Of note, family 24, harboring a frameshift mutation at exon 12, showed the highest ASE imbalance and an attenuated phenotype. It can be speculated that, in this case, the mutant allele is degraded, leading to a bigger proportion of wt APC homodimers and a milder phenotype. In mutations causing PTCs in exon 15, absence of NMD activity may enhance the dominant-negative effect. As such, it is notable that our preliminary observations indicate lower germline *APC* expression levels in polyposis with mutations outside exon 15. An exon 15–mutated allele is not amenable for degradation by NMD, so this mechanism cannot act as an additional regulatory mechanism. However, no clear phenotypic correlation between clinical features and location of the mutation was observed in our series.

Finally, imbalanced ASE was observed in the absence of a detectable mutation in a subset of *APC*(−)/*MUTYH*(−) families assessed by conventional methods. A more detailed analysis of cDNA and promoter regions revealed the presence of alterations in *APC* promoter that may be functionally relevant in one family, whereas in the other family a complex pattern of *APC* transcripts in the exon 11 region was of unknown significance. We cannot rule out the possibility that a genetic change deep within an intron may account for the observed ASE imbalance. In addition, *cis*-acting regulatory SNPs with an allele-specific effect on *APC* also might be responsible for the imbalance.[31] However, the failure to detect imbalances in most of the AFAP cases reflects a putative mechanism for their attenuated nature, which may be related more closely to subtle changes at the RNA level that could lead to delayed phenotypic expression of symptoms. In fact, germline expression of the APC allele is slightly lower in these *APC*(−)/*MUTYH*(−) and shows no apparent correlation with ASE imbalance.

Methodologic issues should be taken into account in the interpretation of results and have both theoretical and practical implications. According to our own and others' results, the SNaPshot approach is a suitable technique with excellent analytic sensitivity.[32–35] Importantly, pyrosequencing (a robust technique) yielded similar normal range values. In any case, the definition of cut-off values was improved by the inclusion of a control set in every experimental run. The putative clinical relevance of borderline values emphasizes the importance of the controls. Sample processing is likely to influence the results. Aged blood has been shown to decrease NMD, thus

precluding the detection of loss of allelic imbalance in the *NF1* gene.[29] The balanced allelic expression of the 2 samples with mutations outside exon 15 may be owing to differences in the isolation process of peripheral blood lymphocytes. Finally, the complex pattern of transcripts at exon 11 should be considered because it might influence ASE results. Notably, transcripts with a skipping of exon 11 have been detected in controls (data not shown).

Our results may have clinical implications. Allelic imbalances detected in a small proportion of *APC*(−)/*MUTYH*(−) families point to the presence of a pathogenic event in the *APC* gene. Although further studies are needed to elucidate the correlation between allelic imbalance and the disease in families with no detectable mutation, it could be used as a diagnostic marker of elusive mutations in the *APC* gene that otherwise might be detected only by the more costly conversion approach[33] or by next-generation sequencing when implemented in the routine clinical setting. Based on our results, we propose a new molecular diagnostic algorithm for polyposis families who undergo *APC* mutation screening beginning with analysis of gross rearrangements (Figure 4). In samples with ASE imbalance, sequencing of exons 1–14 at the cDNA or genomic DNA level would be followed by analysis of the promoter region and, possibly, the 3′ untranslated region. This strategy would have made it unnecessary to sequence the largest exon of the gene (exon 15 with 6574 base pairs) in approximately 21% of the families included in the present study (12 of 58). The pathogenic cause of the disease would remain unknown if sequencing of the whole coding region of *APC* and the promoter and 3′ untranslated regions and cDNA transcript analyses targeting aberrant skipping or cryptic exons revealed no functionally relevant changes. However, a processing protocol for preserving lymphocytes and RNA would be needed to incorporate this RNA-based assay into routine screening. Also, the addition of more polymorphisms as allele-specific markers would improve informativeness and dependability. We do not know whether missense or unclassified variants from introns 1–14 are associated with ASE imbalance. However, transcript-level analysis has shown that a significant proportion of these mutations can be classified as pathogenic and has revealed the presence of PTCs in some cases.[3]

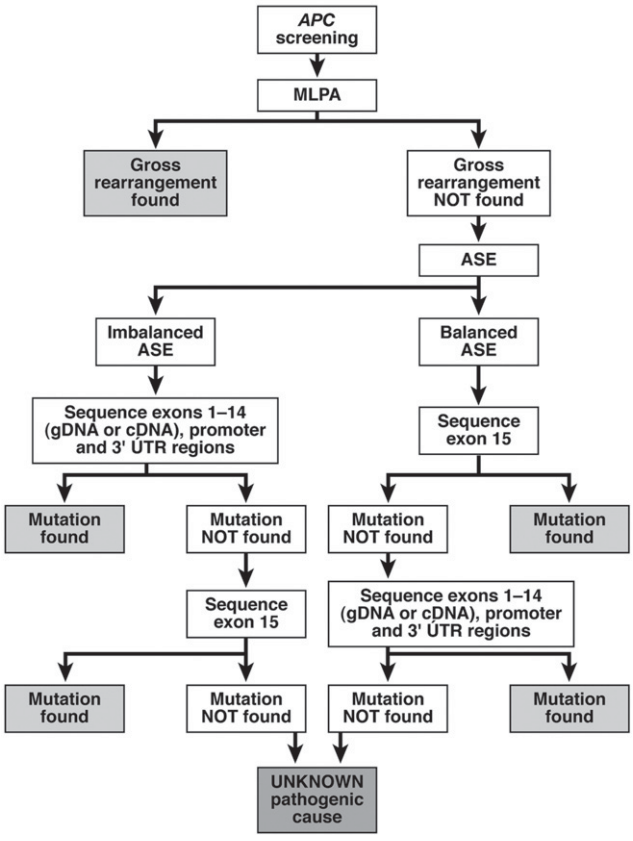


Figure 4. Proposed diagnostic algorithm for *APC* mutation screening including ASE measurement in polyposis families.

Taken together, our results reinforce the importance of RNA-level studies not only for a better understanding of the disease but also for optimizing the molecular diagnostic algorithm for FAP.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at doi: 10.1053/j.gastro.2010.04.047.

References

1. Lipton L, Tomlinson I. The genetics of FAP and FAP-like syndromes. Fam Cancer 2006;5:221–226.
2. Heinimann K, Thompson A, Locher A, et al. Nontruncating APC germ-line mutations and mismatch repair deficiency play a minor role in APC mutation-negative polyposis. Cancer Res 2001;61:7616–7622.
3. Kaufmann A, Vogt S, Uhlhaas S, et al. Analysis of rare APC variants at the mRNA level: six pathogenic mutations and literature review. J Mol Diagn 2009;11:131–139.
4. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985–992.
5. Castellsague E, Gonzalez S, Nadal M, et al. Detection of APC gene deletions using quantitative multiplex PCR of short fluorescent fragments. Clin Chem 2008;54:1132–1140.
6. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C–>T:A mutations in colorectal tumors. Nat Genet 2002;30:227–232.
7. Brinckmann A, Mischung C, Bassmann I, et al. Detection of novel NF1 mutations and rapid mutation prescreening with Pyrosequencing. Electrophoresis 2007;28:4295–4301.
8. Chen X, Weaver J, Bove BA, et al. Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. Hum Mol Genet 2008;17:1336–1348.
9. Renkonen E, Zhang Y, Lohi H, et al. Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. J Clin Oncol 2003;21:3629–3637.
10. Tournier I, Raux G, Di Fiore F, et al. Analysis of the allele-specific expression of the mismatch repair gene MLH1 using a simple DHPLC-Based Method. Hum Mutat 2004;23:379–384.

11. Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, et al. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. Hum Mol Genet 2002;11:2805–2814.
12. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3:285–298.
13. Aretz S, Uhlhaas S, Sun Y, et al. Familial adenomatous polyposis: aberrant splicing due to missense or silent mutations in the APC gene. Hum Mutat 2004;24:370–380.
14. Yan H, Dobbie Z, Gruber SB, et al. Small changes in expression affect predisposition to tumorigenesis. Nat Genet 2002; 30:25–26.
15. Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 1993;329:1982–1987.
16. Venesio T, Balsamo A, Rondo-Spaudo M, et al. APC haploinsufficiency, but not CTNNB1 or CDH1 gene mutations, accounts for a fraction of familial adenomatous polyposis patients without APC truncating mutations. Lab Invest 2003;83:1859–1866.
17. De Rosa M, Scarano MI, Panariello L, et al. Three submicroscopic deletions at the APC locus and their rapid detection by quantitative-PCR analysis. Eur J Hum Genet 1999;7:695–703.
18. Kanter-Smoler G, Fritzell K, Rohlin A, et al. Clinical characterization and the mutation spectrum in Swedish adenomatous polyposis families. BMC Med 2008;6:10.
19. Venesio T, Balsamo A, Sfiligoi C, et al. Constitutional high expression of an APC mRNA isoform in a subset of attenuated familial adenomatous polyposis patients. J Mol Med 2007;85: 305–312.
20. Renkonen ET, Nieminen P, Abdel-Rahman WM, et al. Adenomatous polyposis families that screen APC mutation-negative by conventional methods are genetically heterogeneous. J Clin Oncol 2005;23:5651–5659.
21. Norton N, Williams NM, Williams HJ, et al. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. Hum Genet 2002;110:471–478.
22. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66:589–600.
23. Horii A, Nakatsuru S, Ichii S, et al. Multiple forms of the APC gene transcripts and their tissue-specific expression. Hum Mol Genet 1993;2:283–287.
24. Oshima M, Sugiyama H, Kitagawa K, et al. APC gene messenger RNA: novel isoforms that lack exon 7. Cancer Res 1993;53: 5589–5591.
25. Sulekova Z, Ballhausen WG. A novel coding exon of the human adenomatous polyposis coli gene. Hum Genet 1995;96:469–471.
26. Sulekova Z, Reina-Sanchez J, Ballhausen WG. Multiple APC messenger RNA isoforms encoding exon 15 short open reading frames are expressed in the context of a novel exon 10A-derived sequence. Int J Cancer 1995;63:435–441.
27. Thliveris A, Samowitz W, Matsunami N, et al. Demonstration of promoter activity and alternative splicing in the region 5′ to exon 1 of the APC gene. Cancer Res 1994;54:2991–2995.
28. Andreutti-Zaugg C, Scott RJ, Iggo R. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. Cancer Res 1997;57:3288–3293.
29. Pros E, Larriba S, Lopez E, et al. NF1 mutation rather than individual genetic variability is the main determinant of the NF1-transcriptional profile of mutations affecting splicing. Hum Mutat 2006;27:1104–1114.
30. Khajavi M, Inoue K, Lupski JR. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet 2006;14:1074–1081.
31. Milani L, Gupta M, Andersen M, et al. Allelic imbalance in gene expression as a guide to cis-acting regulatory single nucleotide polymorphisms in cancer cells. Nucleic Acids Res 2007;35:e34.
32. Bray NJ, Buckland PR, Owen MJ, et al. Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet 2003;113:149–153.
33. Valle L, Serena-Acedo T, Liyanarachchi S, et al. Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. Science 2008;321:1361–1365.
34. Yan H, Yuan W, Velculescu VE, et al. Allelic variation in human gene expression. Science 2002;297:1143.
35. Zhang Y, Wang D, Johnson AD, et al. Allelic expression imbalance of human mu opioid receptor (OPRM1) caused by variant A118G. J Biol Chem 2005;280:32618–32624.
36. Menendez M, Gonzalez S, Obrador-Hevia A, et al. Functional characterization of the novel APC N1026S variant associated with attenuated familial adenomatous polyposis. Gastroenterology 2008;134:56–64.

Received July 1, 2009. Accepted April 21, 2010.

Reprint requests

Address requests for reprints to: Gabriel Capellá, MD, Translational Research Laboratory, Institut d'Investigacions Biomédiques de Bellvitge–Institut Català d'Oncologia, Barcelona 08907, Spain. e-mail: gcapella@iconcologia.net; fax: (34) 93-260-7466.

Acknowledgments

The authors thank Laura Valle for critical assessment and helpful comments on the manuscript.

Conflicts of interest

The authors disclose no conflicts.

Funding

This study was supported by the Spanish Ministry of Education and Science (AGL2004-07579-04, SAF 06-6084, and SAF 2009-07319); Fundació "la Caixa" (BM 04-107-0); Fundació Gastroenterologia Dr Francisco Vilardell (F05-01); Spanish Networks Red Temática de Investigación Cooperativa en Cáncer (RD06/0020/1050, 1051), from the Carlos III Health Institute; Acción Transversal del Cáncer; National Institutes of Health/National Cancer Institute (R01 CA81488), and the University of Michigan Comprehensive Cancer Center (grant P30 CA46592); Commission for Universities and Research of the Comissionat per a Universitats i Recerca del Departament d'Innovació, Universitats i Empresa Ministry of Universities, Research and the Information Society of the Catalan government, and a European Social Fund grant (2005FI-00809 to E.C.).

Supplementary Table 1. Primer Pairs Used for cDNA and Promoter Analysis

Covered region	Primer forward	Primer reverse
cDNA sequencing		
exons 1 to 5	5′-AAAACTGGAAACTGAGGCATCT-3′	5′-TCCAATTGCCTTCTGGTCAT-3′
exons 3 to 7	5′-GAGGGTTTGTAAATGGAAGCA-3′	5′-CGGTTTCATGCTTGTTCTGA-3′
exons 5 to 9	5′-GCCAGGATATGGAAAAACGA-3′	5′-GCAAAGTTCGCGACATATCA-3′
exons 7 to 11	5′-GGCAGAATGAAGGTCAAGGA-3′	5′-TTCTGCAATGGCCTGTAGTC-3′
exons 9 to 13	5′-GCATGGACCAGGACAAAAAT-3′	5′-TTTACATCTGCTCGCCAAGA-3′
exons 11 to 15	5′-CGATATGCTGGAATGGCTTT-3′	5′-CCACAAAGTTCCACATGCATTA-3′
APC gene expression levels		
APC gene	5′-GCTATGGCTTCTTCTGGACA-3′	5′-AACGGAGGGACATTTTTGACC-3′
b2-microglobulin	5′-CCAGCAGAGAATGGAAAGTC-3′	5′-GATGCTGCTTACATGTCTCG-3′
Analysis of promoter regions		
promoter 1B_1	5′-CGGAAACCCAAATGATATTGT-3′	5′-GCCATCTTGTGGGAGGTG-3′
promoter 1B_2	5′-AAGCCAGCAACACCTCTCAC-3′	5′-CGGGAGGAGACGTCAATG-3′
promoter 1A_1	5′-GGATTTTGTCCTTCAACCTCA-3′	5′-CAGGGCAGAAGGATTAGCAC-3′
promotor 1A_2	5′-TGCAAAAATCATAGCAATCGAG-3′	5′-AAGACAGTGCGAGGGAAAAC-3′

NOTE. cDNA sequencing: PCR was performed in a final volume of 25 μL using 2 μL of cDNA, 0.2 μmol/L primers, 0.2 mmol/L dNTPs, 2.5 mmol/L MgCl2, and 1.25 U Taq polymerase (Thermoprime Plus DNA Polymerase; ABgene). For fragments e1-5, e3-7, e5-9 and e9-13, 35 cycles of 30 s at 94°C, 30 s at 63°C, and 30 s at 72°C were performed. For fragments e7-11 and e11-15, 10 cycles of 1 min at 94°C, 1 min at 60°C decreasing 0.5°C/cycle, and 1 min at 72°C were followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C. Differential running patterns between samples and controls were assessed visually at agarose gel resolution. *APC* gene expression levels: We performed quantitative PCR with the Roche LightCycler® 480 real-time PCR detection system in all ICO samples analyzed. The same cDNA used for the SNaPshot analysis was used as the template of the three replicates. A pair of primers targeting exon 2 and 3 of the APC gene were used. Reactions were performed in a volume of 10 μL containing 5 μL LightCycler 480 SYBR Green I Master (Roche), 1 μM of each primer and 1 μL cDNA for 45 cycles (95°C for 10 s; 63°C for 10 s; 72°C for 13 s). Then, the samples were subjected to melting curve analysis from 65 to 97°C. Crossing points, i.e. the cycle number when the amplification becomes exponential, were determined using the LightCycler software. β2-microglobulin was used as the housekeeping gene. To evaluate the differences in expression of the APC gene between samples and controls we fitted a linear regression model. Data was conveniently log-scaled (base 2) to improve normality. Analysis of promoter regions: Promoter was analyzed in four reactions. Each PCR were performed in a final volume of 10μL using 50ng de DNA, 0,2 μmol/L primers, 5μL megamix double(Microzone Limited). For fragments 1A, 2A, 1B, 12 cycles of 30 s at 94°C, 30 s at 59°C decreasing 0.5°C/cycle, and 30 s at 72°C and 23 cycles of 30 s at 94°C, 30 s at 53°C, and 30 s at 72°C were performed. For fragment promoter 1B, 12 cycles of 30 s at 94°C, 30 s at 61°C decreasing 0.5°C/cycle, and 30 s at 72°C and 23 cycles of 30 s at 94°C, 30 s at 55°C, and 30 s at 72°C were performed.

CHARACTERIZATION OF NEW FOUNDER ALU-MEDIATED REARRANGEMENTS IN *MSH2* GENE ASSOCIATED WITH A LYNCH SYNDROME PHENOTYPE

Lucia Pérez-Cabornero, **Ester Borràs Flores**, Mar Infante Sanz, Eladio Velasco Sampedro, Alberto Acedo Becares, Enrique Lastra Aras, Jorge Cuevas González, Marta Pineda Riu, Teresa Ramón i Cajal Asensio, Gabriel Capellá Munar, Cristina Miner Pino i Mercedes Durán Domínguez.

Cancer Prevention Research: 4(10):1546-55 (2012).

Resum del treball: S'ha reportat que, en determinades poblacions, les grans delecions genòmiques en els gens *MLH1* i *PMS2* són una causa freqüent de la síndrome de Lynch. En la cohort estudiada, s'han identificat dues noves grans delecions fundadores en el gen *MSH2*. Aquestes mutacions s'han caracteritzat per la determinació de punt de ruptura, l'anàlisi d'haplotips i la correlació genotip-fenotip. A més, s'han identificat mutacions en els gens *MLH1, MSH2* i *MSH6* en 303 pacients de 160 famílies amb sospita de la síndrome de Lynch no relacionats.

Més de 16% (24 de 160) de les famílies tenien mutacions patogèniques (8 en *MLH1*, 15 en *MSH2*, i 1 en *MSH6*). Dotze d'aquestes famílies (50%) són portadores de mutacions que no havien estat prèviament reportades. El 47% de les famílies amb mutació en el gen *MSH2*, presentaven grans reordenaments. A més, la deleció de l'exó 7 i la deleció del exó 4-8 en el gen *MSH2* són mutacions fundadores en la població espanyola. En aquestes famílies, es va observar una segregació d'un haplotip comú, un fenotip similar, i efectes d'anticipació genètica.

Aquestes troballes simplificaran el diagnòstic, l'assessorament i l'atenció clínica de les famílies amb sospita de síndrome de Lynch en la població espanyola.

Research Article
See perspective on p. 1527

Cancer Prevention Research

Characterization of New Founder Alu-Mediated Rearrangements in *MSH2* Gene Associated with a Lynch Syndrome Phenotype

Lucia Pérez-Cabornero[1], Ester Borrás Flores[3], Mar Infante Sanz[1], Eladio Velasco Sampedro[1], Alberto Acedo Becares[1], Enrique Lastra Aras[5], Jorge Cuevas González[2], Marta Pineda Riu[3], Teresa Ramón y Cajal Asensio[4], Gabriel Capellá Munar[3], Cristina Miner Pino[1], and Mercedes Durán Domínguez[1]

Abstract

It has been reported that large genomic deletions in the *MLH1* and *MSH2* genes are a frequent cause of Lynch syndrome in certain populations. Here, a cohort has been screened and two new founder rearrangements have been found in the *MSH2* gene. These mutations have been characterized by break point determination, haplotype analysis, and genotype–phenotype correlation. Mutations have been identified in the *MLH1*, *MSH2*, and *MSH6* genes in 303 subjects from 160 suspected Lynch syndrome unrelated families. All subjects were tested using heteroduplex analysis by capillary array electrophoresis. Multiplex ligation-dependent probe amplification was used to detect rearrangements in mutation-negative index patients and confirmed by reverse transcriptase PCR. The break point of the deletions was further characterized by the array comparative genomic hybridization method. Immunohistochemical staining and microsatellite instability were studied in tumor samples. Hereditary nonpolyposis colorectal cancer–related phenotypes were evaluated. More than 16% (24 of 160) of the families had pathogenic mutations (8 *MLH1*, 15 *MSH2*, and 1 *MSH6*). Twelve of these families (50%) are carriers of a novel mutation. Seven of the 15 positive *MSH2* families (47%) are carriers of a rearrangement. The exon 7 deletion and exon 4 to 8 deletion of *MSH2* are new founder mutations. The segregation of a common haplotype, a similar phenotype, and anticipation effects were observed in these families. These findings will greatly simplify the diagnosis, counseling, and clinical care in suspected Lynch syndrome families and not just in specific geographic areas, so wide distribution may be explained by migration patterns. *Cancer Prev Res; 4(10); 1546–55. ©2011 AACR.*

Introduction

Hereditary nonpolyposis colorectal cancer (HNPCC) syndrome, or Lynch syndrome (MIM 120435), is probably the most common form of inherited colorectal cancer (CRC), accounting for 1% to 5% of cases (1). Affected individuals have a family history of CRC at an early age, characterized by tumor predominance in the proximal colon and an association with extracolonic tumors. Germline mutations in at least 5 mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, *PMS1*, and *PMS2*) have been identified in families fulfilling international criteria for the syndrome, namely, Amsterdam criteria I or II (2–4), or less stringent criteria referred to as the Bethesda guidelines (5), which lead to tumors characterized by widespread microsatellite instability (MSI).

Identification of inherited predisposition is important because it enables targeted clinical surveillance, which significantly reduces cancer morbidity and mortality in Lynch syndrome families (6).

Nearly 90% of the mutations in databases (Leiden Open Variation Database, LOVD) affect either *MLH1* (MIM 120436) or *MSH2* (MIM 609309; refs. 7, 8). The vast majority of mutations are nonsense, missense, splicing, or frameshift mutations, but a recent report (9) indicates that a substantial percentage of HNPCCs are caused by gross genomic rearrangements in *MMR* gene alterations undetectable with traditional methods of mutation analysis. They account for up to 15% of all pathogenic mutations in *MSH2* and *MLH1* (10). The frequency of large rearrangements in *MSH2* as compared with *MLH1* depends on the studied population (11). Different prescreening methods have been proposed.

Authors' Affiliations: [1]Cancer Genetics Laboratory, IBGM-CSIC, University of Valladolid; [2]Anatomical Pathology Department, Hospital Comarcal de Medina del Campo, Crta. Peñaranda de Bracamonte, Medina del Campo, Valladolid; [3]Hereditary Cancer Program, ICO-IDIBELL, Hospitalet de Llobregat; [4]Oncology Department, Hospital de la Santa Creu i Sant Pau, Barcelona; and [5]Oncology Department, Hospital General Yagüe, Burgos, Spain

Corresponding Author: Mercedes Durán Domínguez, Cancer Genetics Laboratory, IBGM-CSIC, University of Valladolid, C/Sanz y Forés 3, 47003 Valladolid, Spain. Phone: 34-983184809; Fax: 34-983184800; E-mail: merche@ibgm.uva.es

doi: 10.1158/1940-6207.CAPR-11-0227

Here, a combined method of heteroduplex analysis by capillary array electrophoresis (HA-CAE) and multiplex ligation-dependent probe amplification (MLPA) has been used to screen our HNPCC population (12). Using this protocol, numerous different mutations have been detected and a high proportion has rearrangements, all in the *MSH2* gene.

Two founder mutations in the *MSH2* gene are presented and characterized here. A similar break point was found in all index cases of the carrier families. These mutations seem to be associated with founder effects, as a common haplotype was associated with each; besides, the novel exon 4 to 8 deletion seems to be associated with anticipation.

Our findings have important implications in the diagnosis and management of such families, and these results will help to simplify genetic testing for Lynch syndrome.

Patients and Methods

Participants

Patients were recruited through the Regional Hereditary Cancer Prevention Program of Castilla y León (Spain). Informed consent was obtained from 303 subjects belonging to 160 unrelated suspected Lynch syndrome families. The control population used was from the National DNA Bank, a collection of representative DNA samples of the Spanish population. Available DNA, RNA, and tumor blocks were obtained for at least one affected person in each family.

Mutation analysis

General screening of point mutation of the *MLH1*, *MSH2*, and *MSH6* genes was carried out using HA-CAE. This method was developed in our laboratory (13), and the validation for *MMR* genes has recently been published (12).

Fragments showing an HA-CAE–altered pattern were sequenced with the BigDye Terminator Sequencing Kit v3.1 (Applied Biosystems) with unlabeled forward and reverse primers on the ABI 3100 DNA Sequencer (4 capillaries; Applied Biosystems).

Genomic rearrangement screening

Negative samples by HA-CAE were tested using MLPA. Genomic rearrangements were evaluated by MLPA (MRC;). MLPA test kits P003 for *MLH1*/*MSH2* and P008 for *MSH6*/*PMS2* were used according to the supplied protocol. Fragment analysis of the PCR products was carried out on the ABI 3130 Genetic Analyzer and gene dosage calculation and analysis were done using GeneMapper software (Applied Biosystems).

Reverse transcriptase PCR

All the alterations detected by MLPA were confirmed by reverse transcriptase PCR (RT-PCR). Three microliters of total RNA was used to synthesize cDNA with a High Capacity cDNA Reverse Transcription kit (Applied Biosystems) using random primers. RT-PCR was carried out with specific primers designed for the coding sequences flanking the putative mutation. To confirm deletions, short amplicons from RT-PCR were sequenced with the following primers:

Exon 4 to 8 deletion: forward in exon 3, 5′-gttggagttgggtatgtggatt-3′; reverse in exon 9, 5′-tgttgactgcatcttcttttcc-3′

Exon 7 deletion: forward in exon 6, 5′-tgctgaataagtgtaaaacccc-3′; reverse in exon 8, 5′-ggagaagtcagaacgaagatcag-3′.

Immunohistochemical and tumor MSI testing

Immunohistochemical (IHC) staining for *MLH1*, *MSH2*, and *MSH6* genes in tissue from at least one individual of the genomic rearrangement families was analyzed by a pathologist in the General Yagüe Hospital, Burgos, Spain.

Paraffin-embedded tumors from these cases were received in our laboratory for MSI study. Paired colon tumor tissues and normal tissues were microdissected and DNA was extracted using the DNAeasy Tissue Kit (Qiagen). A fluorescence multiplex PCR to amplify BAT25, BAT26, D2S123, D5S346, and D17S250 was carried out as described before (14, 15) with slight modifications. Fluorescently labeled PCR products were detected using the ABI 3130 Genetic Analyzer and the GeneScan software. The tumor was categorized as microsatellite stable if none of the 5 markers showed instability, as MSI-low if one of the 5 markers showed instability, and as MSI-high if 2 or more markers showed instability (14).

Break point characterization

An array comparative genomic hybridization (aCGH) strategy was used to confirm the deletions identified by MLPA and to identify the location of deletion break points.

One positive sample, from each different family with a rearrangement, was tested by a human aCGH 44K specially designed by Nimgenetics for coverage of chromosome 2: 47419322-47580004 (NCBI 36).

On the basis of the information obtained from aCGH, primers were designed spanning the putative break points for each case and used in long-range PCR.

Haplotype analysis

Index cases and their relatives were genotyped with 8 microsatellite polymorphic markers surrounding the *MSH2* gene and 2 single-nucleotide polymorphisms in the *MSH2* gene. As controls, 96 unrelated DNA samples from the general population of Castilla y Leon (National DNA Bank) were also genotyped, and allele frequencies were estimated.

The 10 markers used were localized in an approximately 3.6-Mb region encompassing *MSH2* (locus order: cen-D2S123-D2S1248-D2S1247-Clen30-rs3771278-rs2162123-D2S2227-D2S391-Clen27-D2S119-tel) on chromosome 2: 47419322-47580004. Fluorescently labeled primers were used to amplify the microsatellite polymorphic regions. PCR products were analyzed on the Genetic Analyzer 3130 using GenMapper 3.7 software (Applied Biosystems).

The 2 intragenic single base substitutions located within intron 1 and intron 9 of *MSH2* (rs2162123 and rs3771278)

were screened through high-resolution melting technology (LightCycler 480 Instrument), and samples with altered curves were sequenced.

Phenotypic characterization

Genealogic data and phenotype characteristics (gender, age at onset, cancer history, and characteristic feature of tumors) were evaluated in all rearrangement carrier families.

Statistical method

Comparisons between MMR mutation types (point mutations vs. deletions) in the *MSH2* gene were assessed using the Web resource GraphPad Software.

The variables related to the proband family members included type and number of CRC, endometrial cancer, and/or other Lynch syndrome–related cancers and the corresponding ages of diagnosis.

The number and type of cancers were treated as a categorical variable. Categorical data were reported as absolute values (n) and relative frequencies (%), whereas groups were compared by analyzing a 2×2 contingency table using Fisher's exact test. Age was treated as a continuous variable and thereafter dichotomized to less than 50 or 50 years or more. Continuous data were reported in mean values with their corresponding SD, and groups were compared using Student's t test. A 2-sided $P < 0.05$ was considered statistically significant.

Results

Molecular mutation identification

A total of 24 families with a pathogenic germline mutation were detected in *MSH2*, *MLH1*, and *MSH6* by combined HA-CAE–MLPA analysis. Clinicopathologic features, molecular findings of the index patients, and sample numbers are listed in Table 1. Twelve of these families (50%) are carriers of a novel mutation. Seven of them (29.2%) have a rearrangement, all in *MSH2*. Two new rearrangements encompassing exon 7 and exon 4 to 8 deletions were detected in 3 and 4 nonrelated families, respectively.

Identification of the *MSH2* recurrent mutations

One family had previously been detected with exon 7 deletion and 3 families with exon 4 to 8 deletion. (The MLPA assay, RT-PCR products, and sequencing pattern are shown in ref. 12). New cases of this detection are presented here, and the recurrence of these mutations and the founder effect has been investigated. In total, 4 families (VA17, VA20, VA32, and VA134) were studied for exon 4 to 8 deletion and 3 families (VA4, VA169, and VA247) for exon 7 deletion.

Table 1 shows that 58 patients (19% of our tested population) were analyzed for a rearrangement, 60% of whom are carriers.

A high frequency of *MSH2* exon 4 to 8 deletion was observed in *MSH2* mutation carrier families, which accounted for more than 26% (4 of 15).

In total, 33 probands were analyzed, 22 of which are carriers and 11 are not (double the number of carriers).

IHC and MIS were analyzed in 2 families and both presented no staining of the MSH2 protein and MSI-H (see Table 1).

Break point identification

A customized 4×44 Agilent platform was used to map somatic rearrangements (designed by NimGenetics). The aCGH assay provides a prediction of rearrangement break points for the convenient design of primers and sequencing. Eight hundred twenty-two oligonucleotide probes were used to cover *MSH2*. A median distance between nonoverlapping array probes of 500 bp was obtained.

The break points predicted from aCGH are shown in Figure 1. For exon 7, aCGH predicts a 0.01 Mb deletion at the position 47507393-47515906 and presents 71 probes lost (Fig. 1A); for the rearrangement of encoding exons 4 to 8, a 0.04 Mb deletion is predicted between positions 47492237-47527926, with 284 probes lost (Fig. 1B).

PCR primers were designed in the first probe before and after the deletion. To confirm common deletions, PCR was carried out for several index cases of each family with exon 7 deletion and exon 4 to 8 deletion. These resulted in a deletion product of 36.7 kb in exon 4 to 8 deletion (g.13272_49953del36681; NG_007110.1.gb) and 9.4 kb in exon 7 deletion (g.28106_37472del9366; NG_0071-10.1.gb), exclusively observed in carriers of the common deletion and not in deletion-negative controls (see Fig. 1A and B).

The same mutation was found in exon 7 deletion carrier families and in exon 4 to 8 deletion carrier families. The identical break points are positioned within 2 *Alu* elements (Fig. 1A and B). Thus, the deletion is likely to have arisen through an *Alu*-mediated recombination. The presence of identical break point sequences in all cases (ID proband in Table 1) is suggestive of a founder mutation because a frequently recurring recombination event would likely result in at least a few single-nucleotide differences.

Haplotype analysis

Haplotype analysis was carried out to confirm the common genetic origin of the deletion rearrangements (Fig. 2). Most of the positive families for the deletion are from a small area in Castilla y León (in central of Spain), ancestors of families carrying the exon 4 to 8 deletion came from Valle de las Navas and exon 7 deletion ancestors came from Lerma, both in the province of Burgos.

A shared haplotype was observed cosegregating with the mutation (Fig. 2), which was absent in noncarriers of these families. It provides evidence for a common ancestry among these families.

Figures 2A and B represent the pedigrees of the 4 Del4_8 families and the 3 Del_7 families. The phenotype observed in these families is shown, as are the results of the common haplotype segregation.

Table 1. Pathogenic mutations, clinicopathologic features, and molecular findings of the 24 carrier families

Gene	Family code	ID proband	Gender (male, female)	Cancer site and age at diagnosis	Mutation designation	Exon/ intron	Mutation type	Mutation status	Loss of protein expression	MSI status	No. of carrier	No. of noncarrier
MLH1	VA279	483	M	Colon adenoma, 46	c.23del16; STOP11	E-1	Frameshift	New	MLH1/PMS2-	MSI-H	1	2
	VA296	518	M	Colon carcinoma, 36	c.77delA; STOP35	E-1	Frameshift	New			1	
	VA170	300	M	Colon carcinoma, 45	c.119 delT; p.L40fsX	E-2	Frameshift	MLH1_00899			5	3
	VA44	77	F	Rectum, 27	c.306+5G>A	I-3	Splicing	MLH1_00175	MLH1-	MSI-H	5	9
	VA67	102	M	Colon carcinoma, 59	c.306+5G>A	I-3	Splicing	MLH1_00175			3	4
	VA175	309	M	Colon carcinoma, 37	c.306+5G>A	I-3	Splicing	MLH1_00175			3	3
	VA275	429	M	Colon carcinoma, 37–42	c.1865T>A; Leu622His	E-16	Missense pathogenic	MLH1_00643	MLH1-	MSI-H	1	
	VA2	2	M	Transverse colon, 31	c.2221-2224 delCTGC;ins30	E-19	Frameshift	MLH1_01285[a]			4	3
MSH2	VA117	194	F	Uterine-colon, 47–53	c.229_230delAG; p.Ser77CysfsX4	E-2	Frameshift	MSH2_00100			5	3
	VA22	438	M	Colon-rectum, 43–63	c.229_230delAG; p.Ser77CysfsX4	E-2	Frameshift	MSH2_00100	MSH2/MSH6-	MSI-H	7	6
	VA17	86	F	Endometrium, 60	c.646-1019_1386+2420del p.Ile216_Gln462del	E-4-8	Exon 4–8 deletion	New	MSH2-	MSI-H	6	6
	VA20	29	M	Colon carcinoma, 31	c.646-1019_1386+2420del p.Ile216_Gln462del	E-4-8	Exon 4–8 deletion	New			8	2
	VA32	52	F	Endometrium, 51	c.646-1019_1386+2420del p.Ile216_Gln462del	E-4-8	Exon 4–8 deletion	New	MSH2/MSH6-	MSI-H	5	3
	VA134	246	F	Endometrium, 42	c.646-1019_1386+2420del p.Ile216_Gln462del	E-4-8	Exon 4–8 deletion	New			3	
	VA4	4	F	Colon carcinoma, 39	c.1077-3513_1276+5655 p.R359RfsX16	E-7	Exon 7 deletion	New			8	6
	VA169	299	M	Colon carcinoma, 29–37	c.1077-3513_1276+5655 p.R359RfsX16	E-7	Exon 7 deletion	New	MSH2/MSH6-	MSI-H	1	2
	VA247	413	M	Colon/prostate/urinary, 47–57–60	c.1077-3513_1276+5655 p.R359RfsX16	E-7	Exon 7 deletion	New			4	4
	VA188	326	F	Colon carcinoma, 68	c.1226_1227delAG p.Gln409ArgfsX7	E-7		MSH2_00323	MSH2/MSH6-	MSI-H	4	2
	VA251	432	F	Colon/endometrium, 23–35	c.1216C>T; p.Arg406X	E-7	Nonsense	MSH2_00312			1	
	VA6	75	M	Colon-rectum, rectum-sigma, 47	C.1661G>A; p.Ser554Thr	E-10	Splicing	MSH2_00999[a]	MSH2/MSH6-	MSI-H	18	26

(Continued on the following page)

Table 1. Pathogenic mutations, clinicopathologic features, and molecular findings of the 24 carrier families (Cont'd)

Gene	Family code	ID proband	Gender (male, female)	Cancer site and age at diagnosis	Mutation designation	Exon/ intron	Mutation type	Mutation status	Loss of protein expression	MSI status	No. of carrier	No. of noncarrier
	VA174	305	M	Colon carcinoma, 41	c.2240_2241delTA p.Ile747ArgfsX2	E-14	Frameshift	MSH2_01137	MSH2/MSH6-	MSI-H	2	
	VA199	368	M	Urinary tract/colon carcinoma, 50–52	c.2470C>T; p.Gln824X	E-15	Nonsense	MSH2_01163	MSH2/MSH6-	MSI-H	3	2
	VA191	430	F	Endometrium, 40	c.2634G>A skipping exon 15	E-15	Splicing	MSH2_00823			2	
MSH6	VA142	265	F	Endometrium, Breast, 59–61	c.699delT; STOP245	E-4	Frameshift	New	MSH6-		4	1

NOTE: In gray, mutations not described before.
[a]Mutations described by us in LOVD.

Ninety-six unrelated samples were also genotyped as controls and allele frequencies were estimated; these are also shown in Figure 2A and B.

Screening of founder rearrangement deletion in *MSH2*

A PCR test was designed to screen these deletions in first-degree relatives. A routine PCR procedure was optimized. Three primer sequences were used: 1 forward and 2 reverse. The product is a multiplex PCR with the presence of 1 band in the wild type and 2 bands in deletion carrier samples (Fig. 3). This procedure is faster, cheaper, and easier than MLPA.

Genotype–phenotype correlation

Taking into account the fact that 62.5% of mutations we have found are in *MSH2*, largely due to the existence of 2 recurrent mutations in this gene that represent approximately half the mutations in *MSH2*; we have correlated the type of mutation (punctual or rearrangement) with the occurrence of extracolonic tumors and the age at diagnosis (Table 2).

The results of Table 2 show no differences between the 2 groups in the number of tumors developed ($P = 1.0000$) or the age at diagnosis, using the threshold of 50 years ($P = 1.0000$), unlike the tumor type where the prevalence of endometrial and urinary system tumors is higher in founder mutation carriers than in point mutation carriers ($P = 0.47279$ and 0.1247).

On the other hand, it must be emphasized that the median age at onset is different in both groups; 46.7 versus 40.67 years in CRC males and 50 versus 33 years in CRC females. These data, however, were not statistically significant.

Discussion

Lynch syndrome is a heterogeneous disorder with respect to its molecular basis as well as its phenotypic expression.

A variety of point mutations, as well as large genomic rearrangements, have been reported in LOVD (ref. 16).

Here, 160 families were screened using a combined method previously published by our group (12). The overall mutation detection rate in our study is 15% (24 of 160 positive families).

We have found 8 *MLH1* mutation carrier families (Table 1), 3 of which have the *c.306+5G>A* mutation, which has been described as a founder mutation in Spain, and our group has participated in the work (17). Two truncated mutations in exon 1 at *MLH1* are novel, both appearing in a proband with colon cancer before 50 years of age, whereas the *MLH1_01285* mutation has being described by us before (18).

Only one family is an *MSH6* mutation carrier, this is a novel truncated variant. The carrier shows a late-onset endometrial cancer with the absence of any staining of the MSH6 protein and MSS tumor. Similar results have been reported in several studies (19).

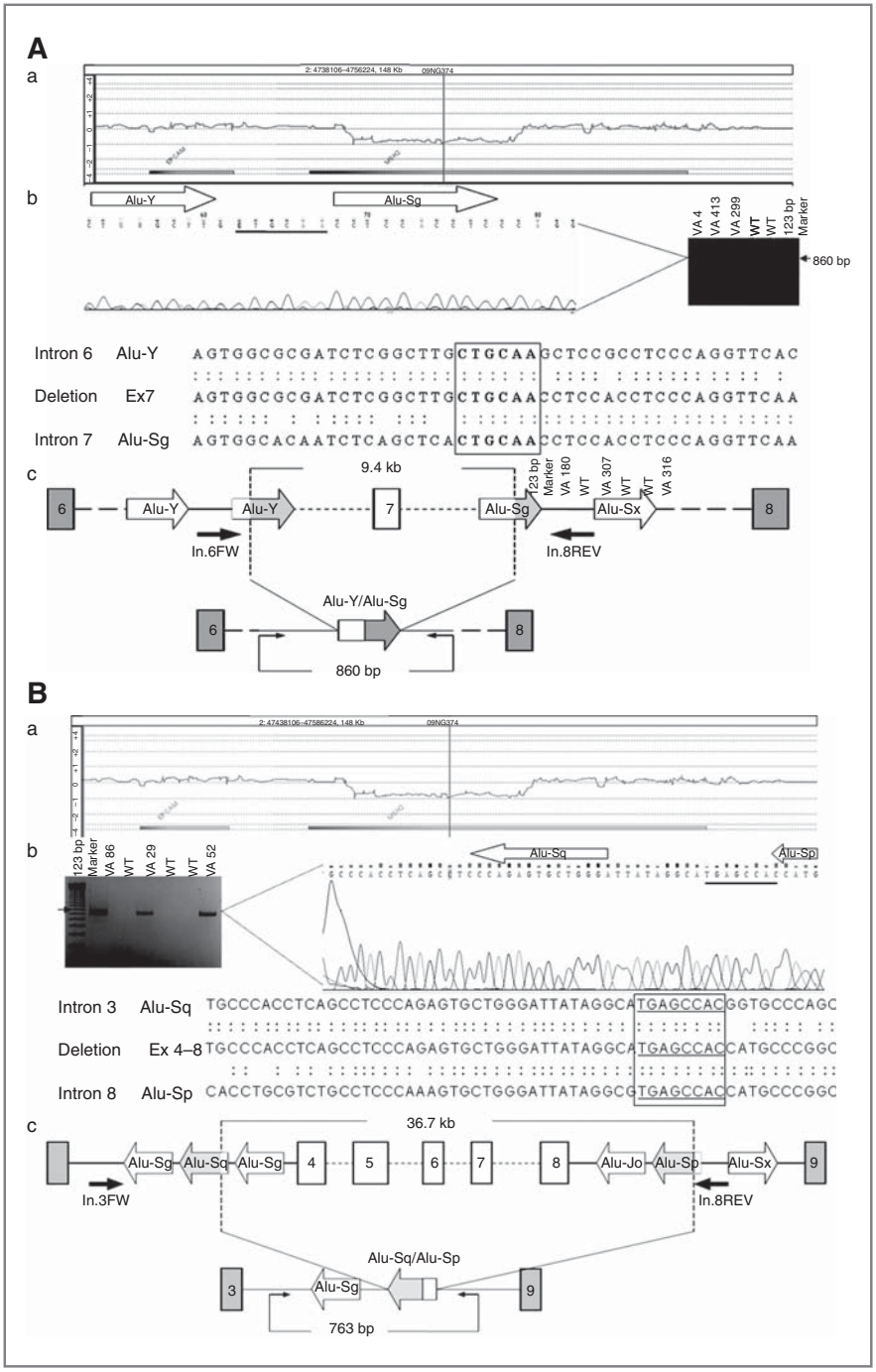


Figure 1. A, exon 7 deletion rearrangement. a, the break points predicted from aCGH. b, rearrangement characterization: PCR amplification, sequencing, and break point sequence. c, break point location and identification of Alu elements. B, exon 4 to 8 deletion rearrangement. a, the break points predicted from aCGH. b, rearrangement characterization: PCR amplification, sequencing, and break point sequence. c, break point location and identification of Alu elements.

The higher rate of mutations in this study appears in the *MSH2* gene (nearly double that in *MLH1*); this is probably because of the high number of recurrent mutations in this gene.

Two mutations that affect the splicing are reported here and which we have described in LOVD. The *c.1661G>A* mutation is present in family VA6, which is a large family; we have studied 44 members, of which 18 are carriers (Table 1). This alteration is associated with the absence of MSH2/MSH6 staining, MSI-H tumor, and a phenotype with colon rectum cancer before the age of 50. The mutation *c.2634G>A* in exon 15 of *MSH2* caused

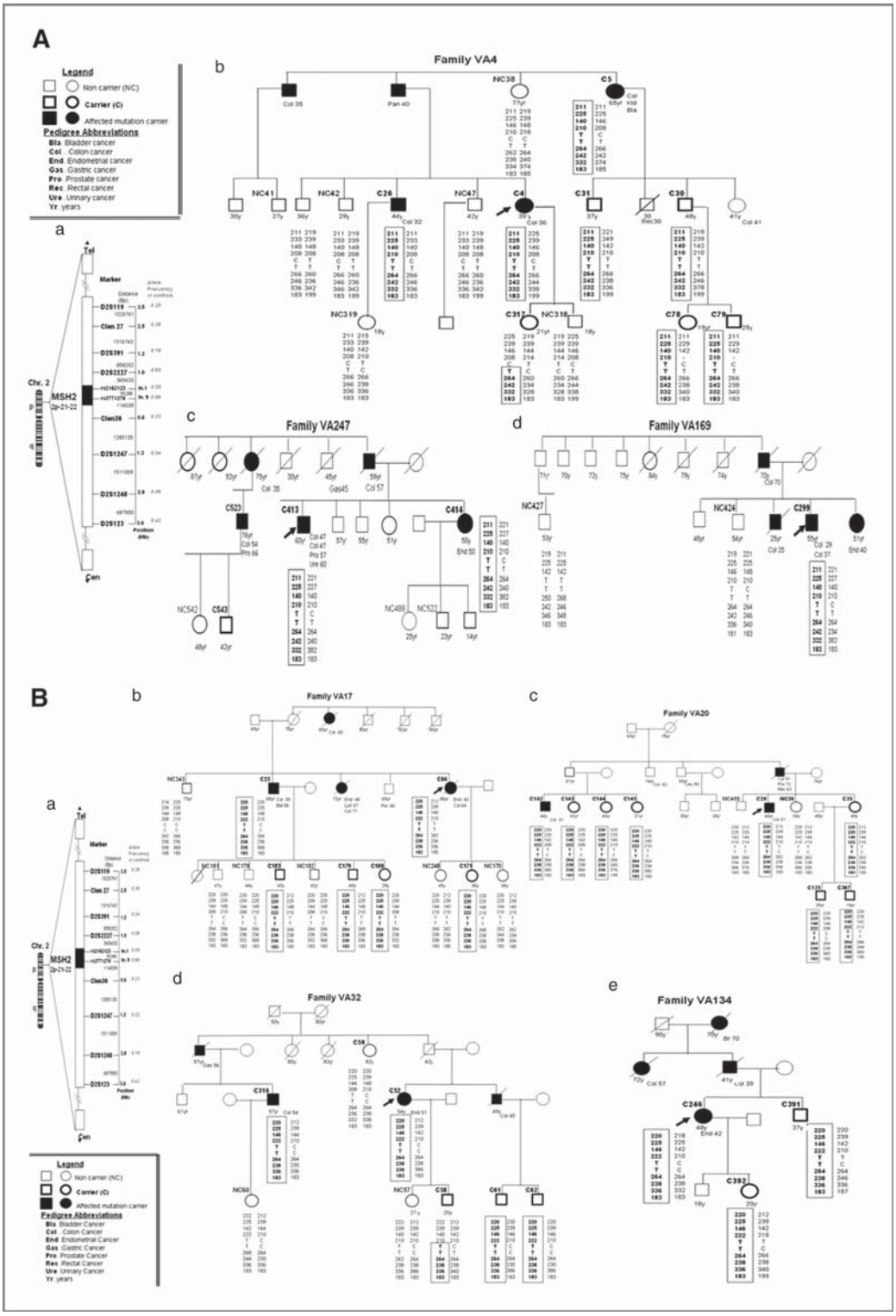


Figure 2. A, a, marker localization and allele frequency in control population of the *MSH2* gene in chromosome 2. b–d, representative pedigrees of exon 7 deletion carrier families. Legend and tumor phenotype are included. B, a, marker localization and allele frequency in control population of the *MSH2* gene in chromosome 2. b–e, representative pedigrees of exon 4 to 8 deletion carrier families. Legend and tumor phenotype are included.

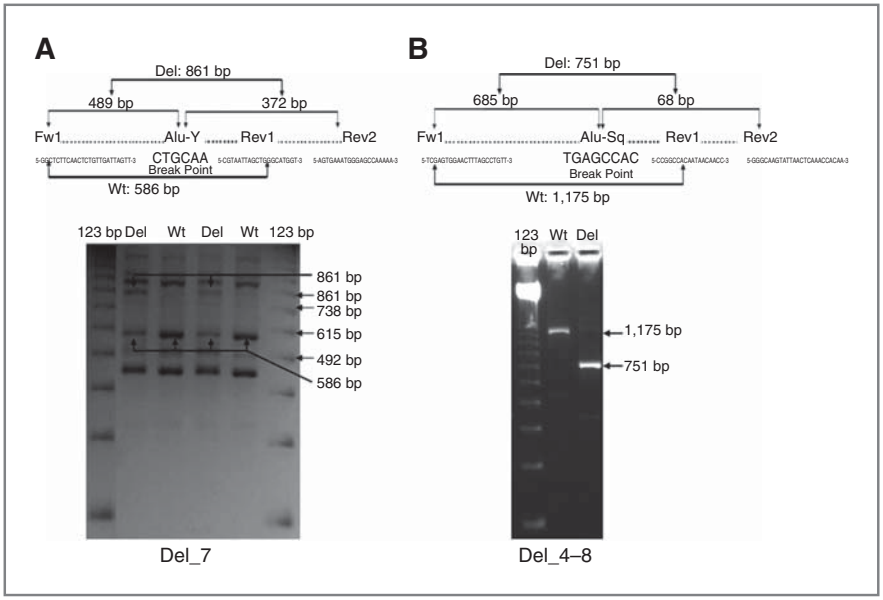


Figure 3. Multiplex PCR of screening rearrangements. A, exon 7 deletion, the wild-type samples show one 586-bp band and deletion samples show 2 bands: 586 bp and an extra band of 861 bp. B, exon 4 to 8 deletion, the wild-type samples show one 1,175-bp band and deletion samples show 2 bands: 1,175 bp and an extra band of 751 bp.

endometrial cancer in a mother and her daughter at 43 and 40 years of age, respectively.

Two recurrent rearrangements have been detected in 7 of 15 *MSH2* carrier mutation families (>46%). Genome rearrangements represent a significant proportion of all pathogenic mutations in the *MMR* genes of patients with colon cancer (20).

An exon 7 deletion was found in 3 nonrelated families, the deletion producing a change in the reading frame and a truncated protein. Other rearrangements in this exon have been described previously (9), but here we exactly characterize this mutation in every index case of the 3 families and all of them have the identical mutation and a shared haplotype (Fig. 2A). The identification of the break points, within the *Y* and *Sg* Alu elements, supports the hypothesis that the recurrent exon 7 deletion is due to an *Alu* repeat–mediated recombination event. These 3 families had a high prevalence of CRC, followed by endometrial cancer and prostate tumor, with the majority of cancers diagnosed before the age of 50 (Fig. 2A). In recent years, recurring mutations for various hereditary cancer syndromes have been identified around the world (21).

The most important evidence in our study is a big deletion that includes 5 exons in the *MSH2* gene, an in-frame mutation that produces a shorter protein and appears in 4 nonrelated Amsterdam criteria fulfilled families.

The study of the break point in index cases of each of the 4 families and the haplotype analyzed confirms the same mutation in all families (Fig. 2B).

We identified that another Alu repeat event that was involved in the exon 4 to 8 deletion (Fig. 2B).

The significant proportion of families in our population with one of these founder rearrangements (near 50% of the families) suggests the need to design a simpler, faster, and cheaper method to detect these mutations. Thus, our sample study protocol could be changed and a prescreening of founder mutations can be started.

We have carried out a multiplex PCR (Fig. 3) that is cheaper than the MLPA method.

Table 2 describes the cancer types in affected carriers and the average age of onset. Most of them developed colon cancer before the age of 45, and the women developed endometrial cancer before the age of 50.

Our data indicate that the anticipation phenomenon is associated with exon 4 to 8 *MSH2* deletion. This is a subjective observation based on evidence from the age of onset and an exhibition of a more aggressive disease severity (number of cancers/tumors or the stage of the tumor) over successive generations (Fig. 2B).

The anticipation phenomenon has been described in other syndromes (22). The presence of anticipation in HNPCC is more controversial, and some reports provide significant evidence that *MHL1* mutations, in particular, may be associated with anticipation (23). Our data suggest that *MSH2* deletions may well be associated with anticipation effects. The median age at colon cancer diagnosis was 44 years in the 4 families (Table 2), but in family VA20, we observed that the difference at onset over 2 generations was 20 years. In family VA17, the number of cancers increased over the generations.

Both *MSH2* rearrangements observed here occurred between intron 4 and 8. It has been described that this is an Alu-rich region because approximately 84% of Alu sequences in *MSH2* are located between the promoter and exon 9 and it is possible that these sequences mediate the recombination observed in this gene (24, 25).

Estimating the age of founder mutations is generally an inexact task (26). It was therefore impossible to estimate

Table 2. Genotype–phenotype correlation in *MSH2* mutation carriers

	Punctual (truncating)	Rearrangements (founder)	Statistical comparison	*P*
Number of families	8	7		
Number of affected	19	15		
Males, *n* (%)	10 (52.6)	6 (40)		
	***n* (%)**	***n* (%)**		
Number of tumors			Several tumors (yes/no)	1.0000
1 tumor	11 (57.9)	8 (53.3)		
>1 tumor	8 (42.1)	7 (46.7)		
Tumor type				
Colon (CRC)	17 (58.6)	13 (50.0)	CRC/EC	0.4727
Endometrial cancer (EC)	4 (13.8)	6 (23.1)		
Urinary system (US)	2 (6.9)	7 (26.9)	CRC/US	
Other cancers	6 (20.7)	–		0.1247
Age at diagnosis, y				
<50	15 (78.9)	11 (73.3%)	Age at diagnosis <50 (yes/no)	1.0000
>50	4 (21.1)	4 (26.7%)		
Age at diagnosis, y	**Average (SD)[a]**	**Average (SD)[a]**		
Total (males and females)	46.37 ± 10.75	40.13 ± 10.46	Average age	0.0987
CRC (males)	46.71 ± 6.24	40.67 ± 9.05	Median age	0.1827
CRC (females)	50.00 ± 23.72	33.00 ± 6.89	Median age	0.1657
Endometrial (females)	43.00 ± 1.41	50.75 ± 7.37	Median age	0.2357

NOTE: [a]Standard deviation.

whether all our founder carrier families share the same haplotype. The fraction of haplotype can be inadequate for recombination events, as we have a small number of carrier families and these show a short genealogy length. The evidence suggests that the origin of these mutations is a recent event.

Given the high proportion of rearrangements, we have evaluated the phenotype in *MSH2* mutation carriers. A more severe phenotype (more different tumors, less onset, etc.) in rearrangement carriers was expected (25), but differences were not found. We only noticed a less primary CRC tumor onset in rearrangement carriers than in other *MSH2* mutations carriers.

We detected rearrangements more often in families with endometrial or urinary cancer. These data are similar to those published by Geray and colleagues (27). A possible explanation could be that both rearrangements occur in the *MSH2* gene and both are founder mutations in our population.

The differences found are not statistically significant, and this does not allow us to orient the diagnosis based on mutation type (punctual or rearrangement) in clinicopathologically suspect patients of *MSH2* mutation.

In conclusion, we have provided genetic evidence that the exon 7 deletion and exon 4 to 8 deletion are both pathogenic founder mutations involved in causing HNPCC in a territory in central Spain. Our data show that large genomic rearrangements occur in these genes with a high frequency and emphasize the need to incorporate techniques to routinely detect them. This should facilitate the genetic diagnosis of Lynch syndrome in our population. The origin of the *MSH2* founder rearrangement can be linked to specific geographic areas, and their current distribution is compatible with the presumed migration pattern in our country. In fact, members of our families have been studied in other cities of Spain (Madrid, Barcelona, and Seville), and it is for this reason that our results are very important.

Our findings will greatly simplify the diagnosis, counseling, and clinical care in suspected families.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the National DNA Bank of Salamanca and the Science Foundation AECC. We also thank Noemy Martínez and Lara Hernández for their excellent technical support and Alan Hynds for his critical reading of the manuscript.

Grant Support

This work has been supported by the University of Valladolid, Valladolid, Spain.

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 May 4, 2011; revised June 16, 2011; accepted July 5, 2011; published OnlineFirst July 21, 2011.

References

1. Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, et al. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology 1993;104:1535–49.
2. de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004;4:769–80.
3. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424–5.
4. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. Gastroenterology 1999;116:1453–6.
5. Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 1997;89:1758–62.
6. Järvinen HJ, Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomäki P, et al. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. Gastroenterology 2000;118:829–34.
7. Peltomaki P, Vasen H: Mutations associated with HNPCC predisposition—update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 2004;20:269–76.
8. Mangold E, Pagenstecher C, Friedl W, Mathiak M, Buettner R, Engel C, 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.
9. van der Klift H, Wijnen J, Wagner A, Verkuilen P, Tops C, Otway R, et al. Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). Genes Chromosomes Cancer 2005;44:123–38.
10. Wang Y, Friedl W, Lamberti C, Jungck M, Mathiak M, Pagenstecher C, et al. Hereditary nonpolyposis colorectal cancer: frequent occurrence of large genomic deletions in MSH2 and MLH1 genes. Int J Cancer 2003;103:636–41.
11. Charbonnier F, Olschwang S, Wang Q, Boisson C, Martin C, Buisine MP, et al. MSH2 in contrast to MLH1 and MSH6 is frequently inactivated by exonic and promoter rearrangements in hereditary nonpolyposis colorectal cancer. Cancer Res 2002;62:848–53.
12. Perez-Cabornero L, Velasco E, Infante M, Sanz D, Lastra E, Hernández L, et al. A new strategy to screen MMR genes in Lynch Syndrome: HA-CAE, MLPA and RT-PCR. Eur J Cancer 2009;45:1485–93.
13. Velasco E, Infante M, Durán M, Pérez-Cabornero L, Sanz DJ, Esteban-Cardeñosa E, et al. Heteroduplex analysis by capillary array electrophoresis for rapid mutation detection in large multiexon genes. Nat Protoc 2007;2:237–46.
14. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248–57.
15. Berg KD, Glaser CL, Thompson RE, Hamilton SR, Griffin CA, Eshleman JR, et al. Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction. J Mol Diagn 2000;2:20–8.
16. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 2011;32:557–63. Available from: http://chromium.liacs.nl/LOVD2/colon_cancer/home.php?action=switch_db.
17. Borràs E, Pineda M, Blanco I, Jewett EM, Wang F, Teulé A, et al. MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. Cancer Res 2010;70:7379–91.
18. Velasco E, Infante M, Durán M, Esteban-Cardeñosa E, Lastra E, García-Girón C, et al. Rapid mutation detection in complex genes by heteroduplex analysis with capillary array electrophoresis. Electrophoresis 2005;26:2539–52.
19. Zhao YS, Hu FL, Wang F, Han B, Li DD, Li XW, et al. Meta-analysis of MSH6 gene mutation frequency in colorectal and endometrial cancers. J Toxicol Environ Health A 2009;72:690–7.
20. Martinez-Bouzas C, Ojembarrena E, Beristain E, Errasti J, Viguera N, Tejada Minguez MI. High proportion of large genomic rearrangements in hMSH2 in hereditary nonpolyposis colorectal cancer (HNPCC) families of the Basque Country. Cancer Lett 2007;255:295–9.
21. Yap HL, Chieng WS, Lim JR, Lim RS, Soo R, Guo J, et al. Recurring MLH1 deleterious mutations in unrelated Chinese Lynch syndrome families in Singapore. Fam Cancer 2009;8:85–94.
22. Hsu L, Zhao LP, Malone KE, Daling JR. Assessing changes in ages at onset over successive generation: an application to breast cancer. Genet Epidemiol 2000;18:17–32.
23. Stella A, Surdo NC, Lastella P, Barana D, Oliani C, Tibiletti MG, et al. Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. Clin Genet 2007;71:130–9.
24. Nakagawa H, Hampel H, de la Chapelle A. Identification and characterization of genomic rearrangements of MSH2 and MLH1 in Lynch syndrome (HNPCC) by novel techniques. Hum Mutat 2003;22:258.
25. Baudhuin LM, Ferber MJ, Winters JL, Steenblock KJ, Swanson RL, French AJ, et al. Characterization of hMLH1 and hMSH2 gene dosage alterations in Lynch syndrome patients. Gastroenterology 2005;129:846–54.
26. Greenwood CM, Sun S, Veenstra J, Hamel N, Niell B, Gruber S, et al. How old is this mutation? A study of three Ashkenazi Jewish founder mutations. BMC Genet 2010;11:39.
27. Geary J, Sasieni P, Houlston R, Izatt L, Eeles R, Payne SJ, et al. Gene-related cancer spectrum in families with hereditary nonpolyposis colorectal cancer (HNPCC). Fam Cancer 2008;7:163–72.

SUPPLEMENTARY MATERIAL

Page 8

Based on the information obtained from the CGH array, primers were designed spanning the putative breakpoints for each case and used in long-range PCR

Mutation	From	To	5´--3´ primer sequence	maximum amplicon size
Exon 7 deletion	MSH2-47507393	MSH2-47515906	Intr6-MSH2-S- GGCTCTTCAACTCTGTTGATTAGTT Intr8-MSH2-A- AGTGAAATGGGAGCCAAAA	860bp
Exon 4_8 deletion	MSH2-47492237	MSH2-47527926	Intr3-MSH2-S- TCGAGTGGAACTTTAGCCTGTT Intr8-MSH2- A- GGGCAAGTATTAACTCAAACCACAA	1979bp

PCR conditions

Buffer Kappa Taq 1X	2,5 µl
Mix dNTPs 0,2 mM/cada uno	1,5 µl
Primers 12 pmol/ each	1-1 µl
Kappa Taq polimerase	0,5 µl
300 ng DNA	2 µl
H2O	
final volume	**25 µl**

°C	Time		Cicles
94	3	min	
94	30	sec	
60	30	sec	X40
72	1	min/kb*	cicles
72	5	min	
15	∞		

Page 9

Fluorescently labeled primers were used to amplify the microsatellite polymorphic regions

multiplex	Microsatellite	oligonucleotide	5`-3`sequence	Amplicon size	Td (ºC)
A	*D2S119*	**MSH2-D2S119-S**	CTTGGGGAACAGAGGTCATT	**214-232bp**	**61**
		MSH2-D2S119-A-FAM	GAGAATCCCTCAATTTCTTTGGA		
B	*Clen27*	**MSH2-Clen27-S-FAM**	AACACTGCCATAGTCACAACTGCCA	**241bp**	**60**
		MSH2-Clen27-A	ACCAATTACATGCACTTAACGTGAT		
	D2S391	**MSH2-D2S391-S-HEX**	GTAATGGAGCCAGTAGGTTACA	**149 bp**	
		MSH2-D2S391-A	AGAGGGTATGATGGAAAAGC		
	D2S2227	**MSH2-D2S2227-S**	CACGCTGTCCATCTCTGAAT	**179-221bp**	
		MSH2-D2S2227-A-HEX	GCAGTTTCTCGGAATAACCA		
C	*Clen30*	**MSH2-Clen30-S**	ATTTGTGACTGTTCAGCTGCTCTCT	**268 bp**	**58**
		MSH2-Clen30-A-HEX	ACCAGCACCATACAGTCAGCTCCTA		
	D2S1248	**MSH2-D2S1248-S-FAM**	GCTCCCATACTCTCACTTG	**380bp**	
		MSH2-D2S1248-A	TTAATACCATCCTCAGTAACC		
D	*D2S1247*	**MSH2-D2S1247-S**	TTTTCCTGCTTCCCACCTG	**250 bp**	**61**
		MSH2-D2S1247-A-FAM	GCAAAAAGGGGCTGTCTTC		
	D2S123	**MSH2-D2S123-S**	TCAACATTGCTGGAAGTTCT	**196bp**	
		MSH2-D2S123-A-HEX	GACTTTCCACCTATGGGACT		

PCR conditions

Buffer Kappa Taq 1X	2,5 **µl**
Mix dNTPs 0.2 mM/each	1 **µl**
oligonucleotides pmoles/each: 1 (F, R) 2 (F, R) 3 (F, R)	12 1-1 **µl /each**
Kappa Taq polimerasa	0,2 **µl**
300 ng DNA	2 **µl**
H2O	
Final volume	**25 µl**

°C	Time		Cicles
94	3	min	
94	30	sec	X40 cicles
X*	30	sec	
72	30	sec	
72	5	min	
15	∞		

The two intragenic single base substitutions located within intron 1 and intron 9 of MSH2 (rs2162123 and rs3771278) were screened through High Resolution Meelting (HRM) thechnology [LightCycler® 480 Instrument (Roche)]

marker	change	position	Primer sequence	T°C	size
rs2162122	T/C	Intron 1	**MSH2-In1-S**- GTCCCTCCCACAATACATGG	**60**	**191pb**
			MSH2-In1-A- ACAGGCTCATATGCGGAAAG		
rs3771278	C/T	Intron 9	**MSH2-In9-S**- GAATGGGTCATTGGAGGTTG	**60**	**256pb**
			MSH2-In9-A- ATCATACAAGGGCCTGTTGG		

PCR conditions

LightCycler®480 High Resolution Meeting X2	5 **µl**
Oligonucleotides 0,2 **µM**/each (S,A)	0,25-0,25 **µl**
Cl2Mg(25mM)	1 **µl**
30 ng ADN	2 **µl**
H2O	
Final volume	**10 µl**

Programs

Program Name	Cycles	Analysis Mode
DESNATURALIZACION	1	None
PCR	40	Quantification
MELTING	1	Melting Curves
HOLD	1	None

DESNATURALIZACION Temperature Targets

Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step Size (°C)	Step Delay (cycles)
95	None	00:10:01	4,4		0	0	0

PCR Temperature Targets

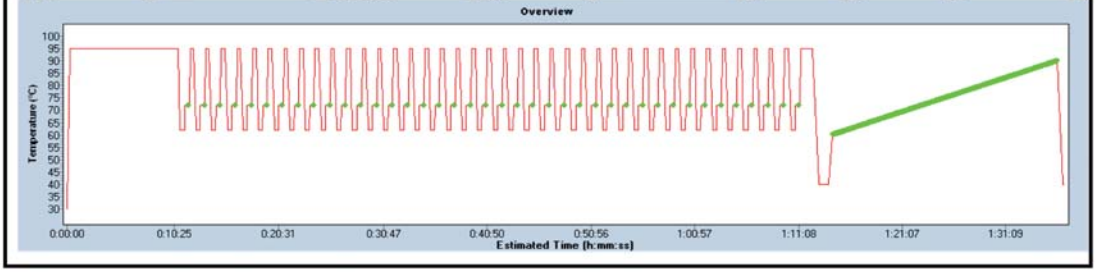
Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step Size (°C)	Step Delay (cycles)
95	None	00:00:10	4,4		0	0	0
62	None	00:00:13	2,2		0	0	0
72	Single	00:00:14	4,4		0	0	0

MELTING Temperature Targets

Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step Size (°C)	Step Delay (cycles)
95	None	00:01:00	4,4				
40	None	00:01:00	1,5				
60	None	00:00:01	1				
90	Continuous		0,02	25			

HOLD Temperature Targets

Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step Size (°C)	Step Delay (cycles)
40	None	00:00:10	1,5		0	0	0



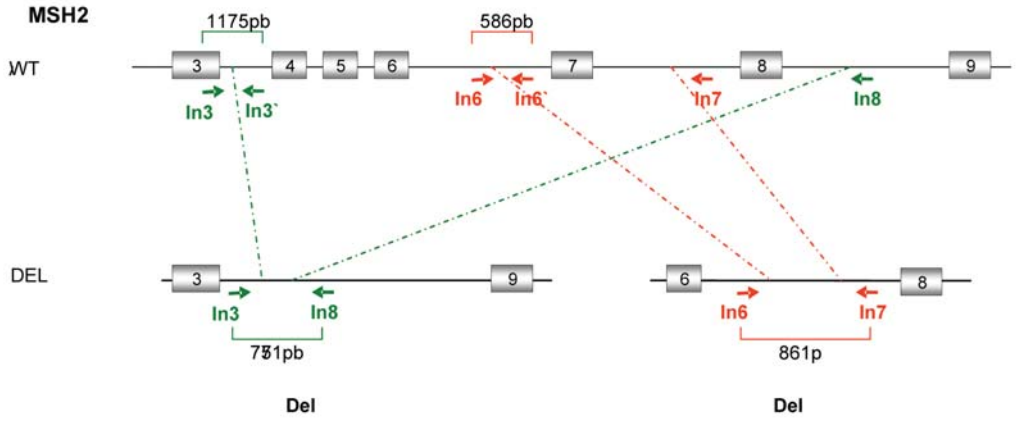
We designed a PCR test to screen these deletions in first degree relatives. A routine PCR procedure was optimized. We used three primer sequences: one forward and two reverse

EXONS 4-8 MSH2 primer sequence	Amplicon size
MSH2-In3S: TCGAGTGGAACTTTAGCCTGTT **MSH2-In8A:** GGGCAAGTATTAACTCAAACCACAA **MSH2-In3´A:** CCGGCCACAATAACAACC	**WT:** 1175pb **Del:** 751pb
EXON 7 MSH2 primer sequence	**Amplicon size**
MSH2-In6S: GGCTCTTCAACTCTGTTGATTAGTT **MSH2-In8A:** CGTAATTAGCTGGGCATGGT **MSH2-In6´A:** AGTGAAATGGGAGCCAAAAA	**WT:** 586pb **Del:** 861pb

PCR conditions

MegaMix-double (microzone)	15ul
MSH2-InXS (8 µM) S:	2ul
MSH2-InXA (8 µM) S´:	2ul
MSH2-InX´A (8 µM) A:	0.5ul
DNA(150ng)	2ul
H20	
Final Volume	**25 µl**

T°C	time		Cicles
94	9	min	
94	1	min	X35
60	1	min	
72	2	min	
72	10	min	
15	∞		



MLH1 METHYLATION SCREENING IS EFFECTIVE IN IDENTIFYING EPIMUTATION CARRIERS

Marta Pineda*, Pilar Mur*, María Dolores Iniesta, **Ester Borràs**, Olga Campos, Gardenia Vargas, Sílvia Iglesias, Anna Fernández, Stephen B Gruber, Conxi Lázaro, Joan Brunet, Matilde Navarro, Ignacio Blanco i Gabriel Capellá. *(*) Aquests autors han contribuït equitativament en aquest treball.*

European Journal of Human Genetics: doi: 10.1038/ejhg.2012.136 (2012).

Resum del treball: Recentment, s'han identificat epimutacions constitucionals en el gen *MLH1* en pacients amb la síndrome de Lynch. L'objectiu d'aquest estudi va ser identificar epimutacions constitucionals en el gen *MLH1*, en un conjunt de pacients amb sospita de la síndrome de Lynch, on *MLH1* es trobava metilat en els tumors i no s'havien detectat mutacions germinals en els gens de reparadors (MMR).

L'anàlisi per MS-MLPA va permetre identificar dues epimutacions constitucionals en el promotor del gen *MLH*1 en dos pacients amb càncer colorectal (5,9%). Aquests resultats van ser confirmats a partir de la seqüenciació amb bisulfit. Els dos individus portadors de les epimutacions havien desenvolupat tumors metacrònics de l'espectre de la síndrome de Lynch a edats joves, i no presentaven història familiar associada.

En un dels casos, la metilació constitucional identificada era monoal·lèlica i mostrava un silenciament transcripcional específic d'al·lel en *MLH1* i *EPM2AIP1*. Aquesta metilació era present en els teixits somàtics normals i era absenten els espermatozoides. L'al·lel metilat va ser transmès per via materna i la metilació es va revertir en una filla que havia heretat el mateix al·lel.

El cribratge de la metilació de *MLH1* en el DNA de limfòcits de pacients amb tumors associats a la síndrome de Lynch, que han estat diagnosticats a edat jove i presenten metilació de *MLH1*, permet la identificació d'individus portadors d'epimutacions en aquest gen. Per altra banda, aquest estudi afegeix evidència a la nova entitat que conformen les epimutacions constitucionals en el gen *MLH1* i la seva herència.

European Journal of Human Genetics (2012), 1–9

www.nature.com/ejhg

ARTICLE

MLH1 methylation screening is effective in identifying epimutation carriers

Marta Pineda[1,6], Pilar Mur[1,6], María Dolores Iniesta[2], Ester Borràs[1], Olga Campos[1], Gardenia Vargas[1], Sílvia Iglesias[1], Anna Fernández[1], Stephen B Gruber[2,3,4], Conxi Lázaro[1], Joan Brunet[5], Matilde Navarro[1], Ignacio Blanco[1] and Gabriel Capellá*[,1]

Recently, constitutional *MLH1* epimutations have been identified in a subset of Lynch syndrome (LS) cases. The aim of this study was the identification of patients harboring constitutional *MLH1* epimutations in a set of 34 patients with a clinical suspicion of LS, *MLH1*-methylated tumors and non-detected germline mutations in mismatch repair (*MMR*) genes. *MLH1* promoter methylation was analyzed in lymphocyte DNA samples by MS-MLPA (Methylation-specific multiplex ligation-dependent probe amplification). Confirmation of *MLH1* constitutional methylation was performed by MS-MCA (Methylation-specific melting curve analysis), bisulfite sequencing and pyrosequencing in different biological samples. Allelic expression was determined using heterozygous polymorphisms. Vertical transmission was evaluated by MS-MLPA and haplotype analyses. MS-MLPA analysis detected constitutional *MLH1* methylation in 2 of the 34 individuals whose colorectal cancers showed *MLH1* methylation (5.9%). These results were confirmed by bisulfite-based methods. Both epimutation carriers had developed metachronous early-onset LS tumors, with no family history of LS-associated cancers in their first-degree relatives. In one of the cases, the identified *MLH1* constitutional methylation was monoallelic and results in *MLH1* and *EPM2AIP1* allele-specific transcriptional silencing. It was present in normal somatic tissues and absent in spermatozoa. The methylated *MLH1* allele was maternally transmitted and methylation was reversed in a daughter who inherited the same allele. *MLH1* methylation screening in lymphocyte DNA from patients with early-onset *MLH1*-methylated LS-associated tumors allows the identification of epimutation carriers. The present study adds further evidence to the emerging entity of soma-wide *MLH1* epimutation and its heritability.
European Journal of Human Genetics advance online publication, 4 July 2012; doi:10.1038/ejhg.2012.136

Keywords: Lynch syndrome; constitutional epimutation; *MLH1*; methylation; MS-MLPA; pyrosequencing

INTRODUCTION

Lynch syndrome (LS) is characterized by an autosomal dominant inheritance of early-onset colorectal cancer (CRC) and increased risk of other cancers.[1,2] It is caused by germline mutations in DNA mismatch repair (*MMR*) genes. *MLH1* or *MSH2* are the most commonly mutated *MMR* genes in LS, whereas mutations in *MSH6* or *PMS2* are significantly less common.[3,4] Occasionally, the presence of constitutional epimutations in *MSH2* and *MLH1* has been reported (reviewed in Hitchins and Ward[5] and Kuiper *et al*[6]).

Constitutional epimutations are those stable changes in gene expression that do not affect DNA sequence and that are present in normal tissues of a given individual.[7] An epimutation that occurs in the germline or early embryo can affect all or most of the soma, and phenocopy genetic disease. *MSH2* epimutations, associated with a strong heritability, have been shown secondary to the presence of deletions in the neighboring *EPCAM* gene.[6] The mutations lead to mosaic methylation of *MSH2* in EPCAM-expressing cells.[8]

Approximately 40 index cases of constitutional *MLH1* methylation have been reported.[9–23] However, the prevalence of *MLH1* constitutional epimutations is still unknown. Most studies addressing this issue have enriched their sampling with patients affected with CRC showing loss of MLH1 protein expression.[13,17,20,22] In other cases, series were enriched for patients with CRC at an age of onset below 50 years.[9,14,17,23]

In a very few cases genetic alterations in *cis* (gross rearrangements and variants in the promoter region) have been identified as responsible for the methylation.[13,16,19] In these cases, an autosomal dominant pattern is readily observed. However, in most cases no genetic cause for the epimutation has been identified (reviewed in Hitchins and Ward[5]). In this context, the inheritance of the epimutation has only been experimentally confirmed in three cases.[10,17,20] The functional impact of these epimutations seems clear. In the few cases analyzed, methylation has been linked to allele-specific silencing of *MLH1* and *EPM2AIP1*.[12,13,15,17] This associates with an allele-specific methylation pattern.[11,17,20,21] In these cases, methylation seems to be widespread affecting all embryonic layers being mosaicism reported.[10,12,20]

The aim of our study was to investigate the prevalence of *MLH1* epimutations in a series of 34 patients with *MLH1*-methylated

[1]Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain; [2]Department of Internal Medicine and Human Genetics, University of Michigan Medical School and School of Public Health, Ann Arbor, MI, USA; [3]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA; [4]USC Norris Conprehensive Cancer Center, Los Angeles, CA, USA; [5]Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IdIBGI, Girona, Spain

*Correspondence: Dr G Capellá, Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL, L'Hospitalet de Llobregat, Gran Via de l'Hospitalet, 199-203, Barcelona 08908, Spain. Tel: +34932607952; Fax: +34932607466; E-mail: gcapella@iconcologia.net

[6]These authors contributed equally to this work.

Received 8 March 2012; revised 22 May 2012; accepted 25 May 2012

CRC and no detected germline *MLH1* mutations. We identified two *bona fide MLH1* epimutation carriers and extensively characterized one of them. The epimutated allele is maternally transmitted, methylation is present in all embryonic layers, erased in spermatozoa and not transmitted to the next generation.

MATERIALS AND METHODS

Patients and samples

Patients were assessed through Cancer Genetic Counselling Units of the Institut Català d'Oncologia (ICO) and the University of Michigan (UM) from 1998 to 2010. A total of 34 individuals (30 ICO, 4 UM) presenting *MLH1*-methylated tumors (methylation levels above 20% in C or D regions) were included in this study (Table 1). The ICO patients were selected from a series of 56 individuals with *MLH1*-deficient CRC and no germline mutations identified in *MLH1*.[24] In all, 29 patients met Bethesda criteria, 1 case met Amsterdam criteria and 4 cases showed other types of CRC familial aggregation. Clinico-pathological information was recorded. Informed consent was obtained from all individuals, and ethics committee approved this study. Sample processing is detailed in Supplementary Methods.

MLH1 promoter methylation analyses

DNA from RKO colorectal tumor cell line (American Type Culture Collection, Manassas, VA, USA) was used as a biallelic *MLH1* methylation control. To generate unmethylated DNA, peripheral blood lymphocyte (PBL) DNA was amplified using the REPLI-g kit (Qiagen, Valencia, CA, USA). A sample of CEPH DNA from the Coriell Institute was used as an unmethylated control in pyrosequencing analyses.[25]

Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA)

SALSA MS-MLPA ME011 kit (MRC Holland, Amsterdam) is based on the use of probes that contain a digestion site specific for the methylation-sensitive *Hha*I enzyme. All reactions were carried out using 100 ng of DNA. The kit includes five probe pairs in *MLH1* promoter (with the respective *Hha*I sites

Table 1 Clinical and molecular features of patients with *MLH1*-methylated CRC

									% somatic *MLH1* methylation		
Case	*Gender*	*Clinical criteria*	*CRC age of onset*	*CRC location*	*TNM*	*Grade*	*Mucinous component*	*Other tumors (age of onset)*	*BRAF*	*C region (−246)*	*D region (−13)*	*MLH1 rs1800734 (c.−93G/A)*
1	M	BC	32	L	T3N0M0	G1	No	CRC (34)	wt	57.6	59.7	GA
2	F	BC	49	R	T2N2M0	G3	Yes		V600E	24.9	36.9	GA
3	M	BC	37	L	T3N0M0	G2	No		wt	29.3	31.7	na
4	F	BC	73	R	T4aN0M0	na	Yes		wt	73.5	70.5	GA
5	M	BC	50	R	T3N1M0	G2	No		wt	28.6	33.6	na
6	F	FA	62	R	T3N0M0	G3	No		wt	61.5	78.5	AA
7	M	BC	42	R	T4N2M0	G2	No	CRC (synch)	wt	24.1	25.2	GG
8	M	BC	29	R	T3N0M0	G2	No		wt	25.1	27.6	na
9	F	BC	47	L	T3N1M0	G2	Yes		wt	38.5	34.9	na
10	F	BC	77	R	T3N0M0	na	Yes		wt	38.2	24.1	GG
11	M	BC	52	R	T3N0M0	G2	No		V600E	35.4	4.4	AA
12	F	BC	62	L	T3N0M0	G2	No		wt	53.7	76.7	GA
13	F	BC	59	R	T3N0M0	G2	No		V600E	39.4	45.8	GG
14	F	BC	77	R	T3N0M0	G2	No		V600E	34.5	28.4	GG
15	F	BC	52	R	T4aN0M0	G2	Yes		V600E	22.9	41.4	GG
16	F	BC	24	R	T3N0M0	G3	No		V600E	57.5	75.1	GG
17	M	FA	78	R	T3N0M0	G2	No		wt	12.5	24.0	GG
18	M	BC	48	R	na	na	No		wt	32.8	34.8	GA
19	M	FA	73	R	T3N0M0	G3	Yes		V600E	19.4	31.2	GA
20	F	BC	50	R	T3N0M0	G2	Yes		V600E	35.8	27.0	GG
21	F	BC	58	R	T3N0M0	G2	No	3 CRC (synch)	V600E	40.6	66.6	GG
22	M	FA	85	R	T4bN0M0	G3	No		V600E	41.4	42.5	GA
23	F	BC	47	L	T3N0M0	G3	Yes		V600E	20.3	39.3	AA
24	F	BC	59	R	T1N0M0	G2	No	CRC (29)	V600E	11.4	20.6	AA
25	M	BC	69	R	T4N0M0	G3	Yes	CRC (synch)	wt	50.3	43.1	GA
26	F	BC	75	R	T2N0M0	G2	No	CRC (64)	V600E	27.1	30.3	GA
27	M	BC	47	L	T3N0M0	G1	No		wt	40.1	21.6	AA
28	M	BC	31	L	T4N0M0	G2	Yes		wt	26.2	32.7	GG
29	F	BC	23	L	T4N1M0	G2	No	GC (26)	wt	79.8	50.4	GA
30	M	BC	86	R	T3N0M0	na	na	BrC (69); RC (78)	wt	na	na	na
31	M	AMS	68	R	T3N0M0	na	na	M (80)	wt	na	na	na
32	F	BC	55	R	T2N0M0	na	na		wt	na	na	na
33	F	BC	52	R	T3N1M0	G3	na		wt	na	na	na
34	F	BC	47	R	T1N0M0	na	No	CRC (29), EC (49)	wt	26.1	37.3	GG

Abbreviations: AMS, Amsterdam criteria; BC, Bethesda criteria; FA, Familial aggregation; M, male; F, female; R, right; L, left; CRC, colorectal cancer; EC, endometrial cancer; GC, gastric cancer; BrC, breast cancer; RC, renal cancer; M, mesothelioma; synch, synchronous; wt, wild-type; na, not available.

located at −659, −383, −246, −13 and +208 relative to the start codon; GenBank accession number U26559) that cover five independent regions: regions A to D of the promoter and intron 1.[26]

Methylation-specific melting curve analysis

Methylation-specific melting curve analysis method consists in a real-time PCR followed by temperature dissociation on DNA previously treated with sodium bisulfite,[27] using the EZ DNA Methylation-Gold Kit (Zymo Research, Orange, CA, USA). Experimental conditions and primers are detailed in Supplementary Methods, Figure S1 and Table S1.

Bisulfite sequencing

A total of 1 μl of bisulfite-converted DNA was used in a PCR reaction for the amplification and subsequently sequencing of *MLH1* promoter regions C and D.[26] Experimental conditions and primers are detailed in Supplementary Methods, Figure S1 and Table S1.

Clonal bisulfite sequencing

A total of 1 μl of bisulfite-modified DNA was amplified, cloned and sequenced. Experimental conditions and primers are detailed in Supplementary Methods, Figure S1 and Table S1.

Pyrosequencing

In all, 2 μl of bisulfite-converted DNA were used in a PCR reaction for the amplification of regions C and D of the *MLH1* promoter[26] using HotStarTaq master mix (Qiagen) and biotinylated primers (Supplementary Table S1 and Figure S1). Primers were designed using the Pyromark Assay Design Software 2.0 (Qiagen). Experimental conditions are detailed in Supplementary Methods.

MLH1 allelic expression analyses

For allelic expression analyses at the c.655A>G SNP (rs1799977) within *MLH1* exon 8, the relative levels of the A/G alleles were determined in genomic DNA and cDNA by single-nucleotide primer extension (SNuPE) and pyrosequencing, as described in Supplementary Methods.

EPM2AIP1 allelic expression analysis

Amplification and sequencing of rs9311149 flanking region, within *EPM2AIP1* gene, was performed as previously described.[12] For allelic expression analysis at rs9311149, the relative levels of G/T alleles were determined in genomic DNA and cDNA by SNuPE as described in Supplementary Methods, using primers listed in Supplementary Table S1.

Direct sequencing of *MLH1* promoter

Screening for mutations within the *MLH1* promoter was performed by PCR amplification and sequencing as described.[28] One reverse amplification primer has been modified (Supplementary Table S1).

Haplotype analysis

Haplotype analysis was performed using four intragenic *MLH1* single-nucleotide polymorphisms (rs1800734, rs9876116, rs1799977 and rs4234259) and seven microsatellite markers (D3S1609, D3S1612, D3S2369, D3S1611, D3S3623, D3S1298, D3S3564) covering 12 Mb around *MLH1*, as previously described.[29] To deduce the methylation-associated haplotype, intrafamilial segregation analysis was performed under the assumption that the number of crossovers between adjacent markers was minimal.

Second hit analysis

Loss-of-heterozygosity (LOH) analysis was performed on DNA extracted from paraffin-embedded tumor tissue and compared with PBL DNA at informative microsatellites (see haplotype analysis) and SNP rs1799977, either by genotyping or SNuPE (see Supplementary Methods), respectively. *MLH1* somatic mutation status was assessed in tumor DNA by direct sequencing and multiplex ligation-dependent probe amplification (SALSA MLPA P003-B1; MRC Holland).

BRAF V600E screening

A 196-bp region of human *BRAF* gene spanning the hotspot mutation c.1799T>A (V600E) was amplified by PCR (Supplementary Table S1) as described.[24] The PCR products were purified using Illustra GFX DNA and Gel Band Purification kit (GE Healthcare, Buckinghamshire, UK). *BRAF* V600E mutation detection was performed by SNuPE using the ABI PRISM SNaPshot Multiplex Kit (Applied Biosystems, Foster City, CA, USA) and a specific primer.

RESULTS

Clinical and molecular features of patients with *MLH1*-methylated CRC

In all, 34 patients (15 males; 19 females) were analyzed (Table 1). Mean age at diagnosis was 55 (range 23–86 years). Twenty-six tumors (76%) were located in the right colon and ten (33%) were classified as mucinous. Only six patients (18%) had lymph node involvement and none of them had distal metastasis. *BRAF* mutations were detected in 13 tumors (38%). A common SNP rs1800734 (c.−93G>A) within the *MLH1* promoter was found to be heterozygote in 10 cases (38%) and homozygote A in 5 (19%). In eight individuals (24%), additional LS-associated tumors were diagnosed, three synchronous and five metachronous (Table 1). Molecular characterization of these additional tumors (Table 2) allowed demonstrating the existence of two *MLH1*-methylated tumors in four individuals (cases 1, 7, 29 and 34).

Identification of new LS cases harboring a constitutional *MLH1* epimutation

The methylation status of *MLH1* promoter was analyzed by MS-MLPA in DNA extracted from PBLs. Constitutional methylation was only detected in 2 individuals (cases 1 and 34) of the 34 patients included (5.9%). It represented 2 out of 100 LS cases in ICO series (2%). In both cases, methylation in *MLH1* promoter was detected in the five regions analyzed, including C and D promoter regions, which was correlated with transcriptional silencing[26] (Table 3).

Sequencing analysis of the whole *MLH1* promoter (from c.−1469 to intron 1) in PBL DNA from cases 1 and 34 did not detect any variant affecting the binding of MLPA probes nor *Hha*I restriction sites. Likewise, it revealed that case 1 was heterozygous for SNP rs1800734 (c.−93G>A) and case 34 was heterozygous for SNP rs34566456 (c.−607G>C). No other variants – including c.−27C>A and c.85G>T[16] – were identified within the promoter region.

Case 1 is a 47-year-old male who underwent urgent sigmoidectomy due to intestinal occlusion secondary to a sigmoid adenocarcinoma (pT3N0M0, stage II) at the age of 32. After 2 years, the patient was diagnosed with an adenocarcinoma of the hepatic flexure (pT3N0M0, stage II) and a subtotal colectomy was carried out. Microsatellite analysis showed MSI, loss of MLH1 and PMS2 expression, absence of *BRAF* V600E mutation and somatic *MLH1* methylation in both tumors (Table 2). The patient had no family history of cancer in his first-degree relatives as it is shown in his pedigree (Figure 1a).

Case 34 is a 55-year-old female who was diagnosed of a sigmoid adenocarcinoma (pT3N1M0, stage III) at the age of 29 years and underwent a sigmoidectomy. After 15 years, the patient was diagnosed with an adenocarcinoma of the hepatic flexure (pT1N0M0, stage I). At the age of 49 years, she was diagnosed of an endometrial adenocarcinoma (pT1N0M0). Microsatellite analysis showed instability of the five analyzed markers in the second CRC, and instability of bat26 and MONO-27 in the endometrial cancer. Both second colorectal and endometrial tumors showed loss of MLH1 and

Table 2 Molecular features of tumors from patients affected by multiple LS-associated tumors

				IHC					% somatic MLH1 methylation	
Case	Tumor type	Age of onset	MSI analysis	MLH1	MSH2	MSH6	PMS2	BRAF	C region (−246)	D region (−13)
1	CRC[a]	32	+	−	+	+	−	wt	57.6	59.7
	CRC	34	+	−	+	+	−	wt	60.5	62.8
7	CRC[a]	42	+	−	+	+	ND	wt	24.1	25.2
	CRC	42	+	−	+	+	−	wt	28.9	24.2
21	CRC[a]	58	+	−	+	+	ND	V600E	40.6	66.6
	CRC	58	−	+	+	+	+	ND	ND	ND
	CRC	58	−	+	+	+	+	ND	ND	ND
	CRC	58	−	ND	ND	ND	ND	ND	ND	ND
24	CRC	29	NA	NA	NA	NA	NA	NA	NA	NA
	CRC[a]	59	+	−	+	+	ND	V600E	11.4	20.6
25	CRC[a]	69	+	−	+	+	ND	wt	50.3	43.1
	CRC	69	NA	NA	NA	NA	NA	NA	NA	NA
26	CRC	64	NA	NA	NA	NA	NA	NA	NA	NA
	CRC[a]	75	+	−	+	+	−	V600E	27.1	30.1
29	CRC[a]	23	+	−	+	+	ND	ND	79.8	50.4
	GC	26	+	−	+	+	ND	ND	63.0	73.0
34	CRC	29	NA	NA	NA	NA	NA	NA	NA	NA
	CRC[a]	47	ND	−	ND	ND	−	wt	55.5	48.8
	EC	49	+	−	+	+	ND	wt	26.1	37.3

Abbreviations: NA, not available; ND, not done.
[a]Tumors included in the initial series listed in Table 1.

Table 3 Analysis of *MLH1* methylation using MS-MLPA in samples from the proband and relatives

			% MLH1 methylation				
Family	Individual	Sample	A region (−659)	B region (−383)	C region (−246)	D region (−13)	Intron 1 (+208)
A	I.1	PBL	0	0	0	0	0
	II.1 (case 1)	CRC 1	**61.2**	**83.7**	**57.6**	**59.7**	**60.9**
		CRC 2	**62.3**	**86.9**	**60.5**	**62.8**	**63.5**
		PBL	**60.5**	**76.7**	**56.0**	**56.2**	**60.2**
		fibroblasts	**55.8**	**53.2**	**64.0**	**52.4**	**63.0**
		colonic mucosa	**52.3**	**78.9**	**58.3**	**48.5**	**62.6**
		sperm	0	0	0	0	0
	II.2	PBL	0	0	0	0	0
	III.1	PBL	0	0	0	0	0
	III.2	PBL	0	0	0	0	0
B	case 34	EC	**33.6**	**59.4**	**26.1**	**37.3**	**28.5**
		CRC	**58.0**	**56.3**	**55.5**	**48.8**	**56.4**
		PBL	**35.9**	**45.3**	**25.1**	**27.6**	**27.7**
		RKO	**110.1**	**113.2**	**103.0**	**88.2**	**103.4**

Peripheral blood lymphocytes (PBL), skin fibroblasts, colorectal tumors (CRC 1 and 2), normal adjacent mucosa and sperm from case 1 (II.1), PBL from his relatives, and PBL, CRC and endometrial cancer (EC) from case 34, were analyzed. DNA from RKO cell line (methylated in *MLH1*) is used as a positive control. Representative data from two independent experiments is shown. Methylation levels above 20% are shown in bold.

PMS2 expression, absence of *BRAF* V600E mutation and somatic *MLH1* methylation (Table 2). Patient's mother was affected by a breast cancer at the age of 77 years (Figure 1b).

Methylation-specific melting curve analysis confirmed the presence of a methylated peak in C and D promoter regions in both cases (Figure 2a and Supplementary Figure S2). Likewise, bisulfite sequencing showed the presence of both methylated C as well as non-methylated T (bisulfite-converted non-methylated C) alleles at each CpG site in the samples of interest (Figure 2a and Supplementary Figure S2). Average methylation levels in PBL of the case 1 were 34% and 39% in C and D regions, respectively, as assesed by pyrosequencing (Figure 2a; Table 4). Clonal bisulfite sequence analysis confirmed hemiallelic methylation in PBL DNA confined to allele A of the rs1800734 (Figure 2b). In case 34, average methylation levels in PBL were 20% and 19% in C and D regions, respectively (Supplementary Figure S2; Table 4).

Functional impact of the *MLH1* epimutations

The *MLH1* promoter is bi-directional for transcription of *MLH1* and *EPM2AIP1* genes. In case 1, the neutral heterozygous polymorphism c.655G>A (rs1799977) within *MLH1* exon 8 was used to determine the effect of the epimutation on *MLH1* transcriptional activity. Monoallelic expression of *MLH1* transcript, associated to G allele, was demonstrated by pyrosequencing and SNuPE (Figure 3). ASE (allele-specific expression) values obtained in patient and control sample were 0.05 and 1.17 when analyzed by pyrosequencing, and 0.02 and 0.98 by SNuPE, respectively. In case 34, the absence of coding heterozygous polymorphisms in *MLH1* prevented its transcriptional analysis.

SNuPE analysis at rs9311149 of *EPM2AIP1* evidenced complete silencing of *EPM2AIP1* G allele in case 1 (Figure 3b, right panel) and partial silencing of the same allele in case 34 (Supplementary Figure S2b), further reinforcing the functional impact of the constitutional methylation. The obtained ASE values were 0.02 in case 1, 0.48 in case 34 and 1.00 in control sample.

Characterization of the *MLH1* epimutation

MLH1 methylation pattern. Follow-up of case 34 and her family has proved difficult. Thus, for the purpose of detailed characterization, we

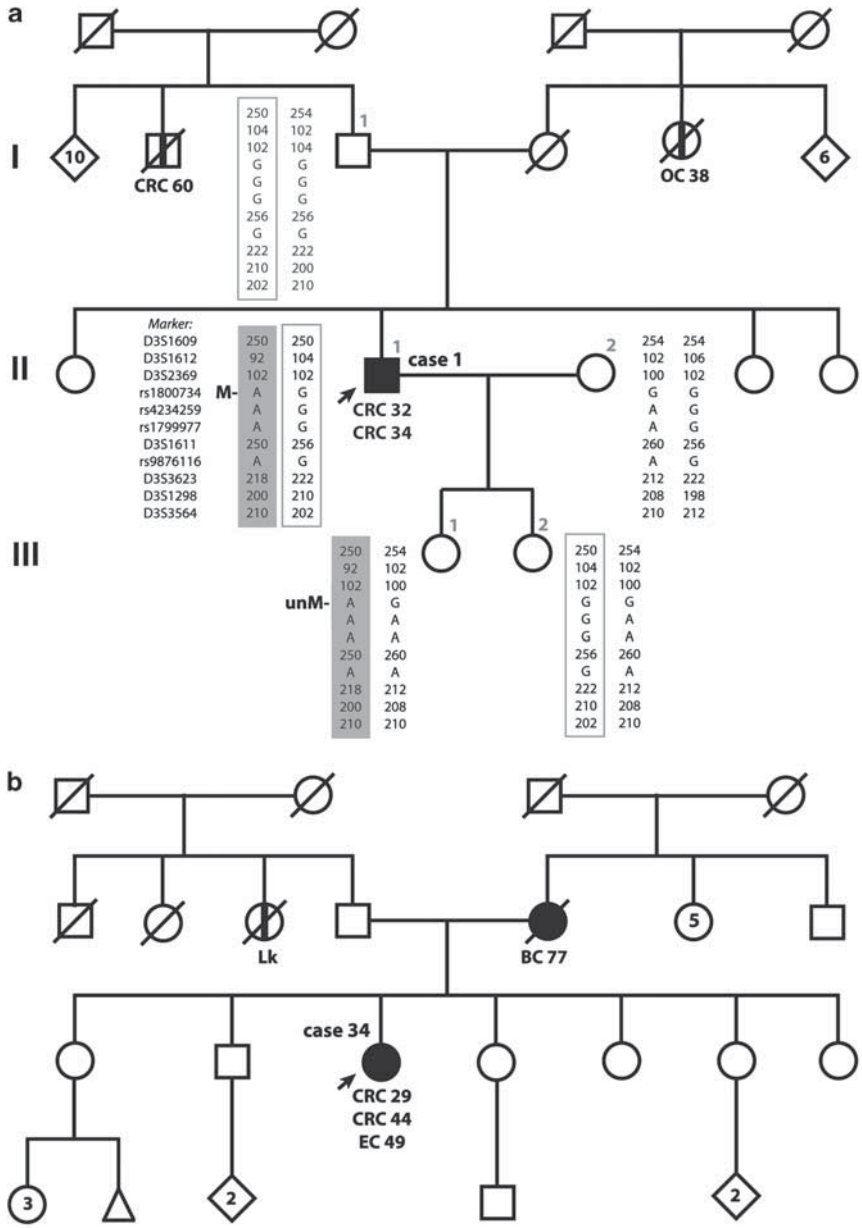


Figure 1 Family pedigree of the epimutation carriers. Circles, females; squares, males; filled, cancer affected; vertical line at center, non-confirmed cancer affected. Cancer localization (CRC, colorectal cancer; OC, ovarian cancer; EC, endometrial cancer; BC, breast cancer; Lk, leukemia) and age at diagnosis are indicated. (**a**) Pedigree and haplotypes of case 1. The epimutation carrier (II.1) is indicated by an arrow. Generations are indicated on the left margin in Roman numerals and analyzed relatives are identified by numbers. Haplotypes, generated by analyzing SNP and microsatellite markers flanking or within *MLH1*, are detailed according to the key indicated in individual II.1. The paternally inherited allele in II.1 is in a square and the maternally derived allele is highlighted in dark gray. The presence of methylation (M) or its absence (unM) is indicated. (**b**) Pedigree of case 34. The epimutation carrier is indicated by an arrow.

have focused in the characterization of case 1. First, we wanted to explore whether methylation was present in all embryonic layers and in the germline of case 1. MS-MLPA analysis in skin fibroblasts (ectoderm) and colorectal mucosa (endoderm) revealed similar levels of *MLH1* methylation than in PBL (Table 3), indicating hemiallelic methylation in all embryonic layers. In contrast, no methylation was detected in patient sperm as evidenced by MS-MLPA and pyrosequencing analyses (Tables 3 and 4). Direct sequencing of the PBL and sperm for *MLH1* promoter C region evidenced the presence of both alleles at rs1800734 in both samples (data not shown). These results indicate the reversion of the epimutation in patient spermatozoa.

Inheritance pattern of the epimutant allele. To further investigate the inheritance pattern of the allele harboring the epimutation, we analyzed the *MLH1* promoter methylation status as well as a haplotype of 12 Mb around *MLH1* in available PBL DNA from patient's first-degree relatives. MS-MLPA analysis showed no evidence of *MLH1* methylation in relatives (Table 3). Haplotype analysis revealed that the

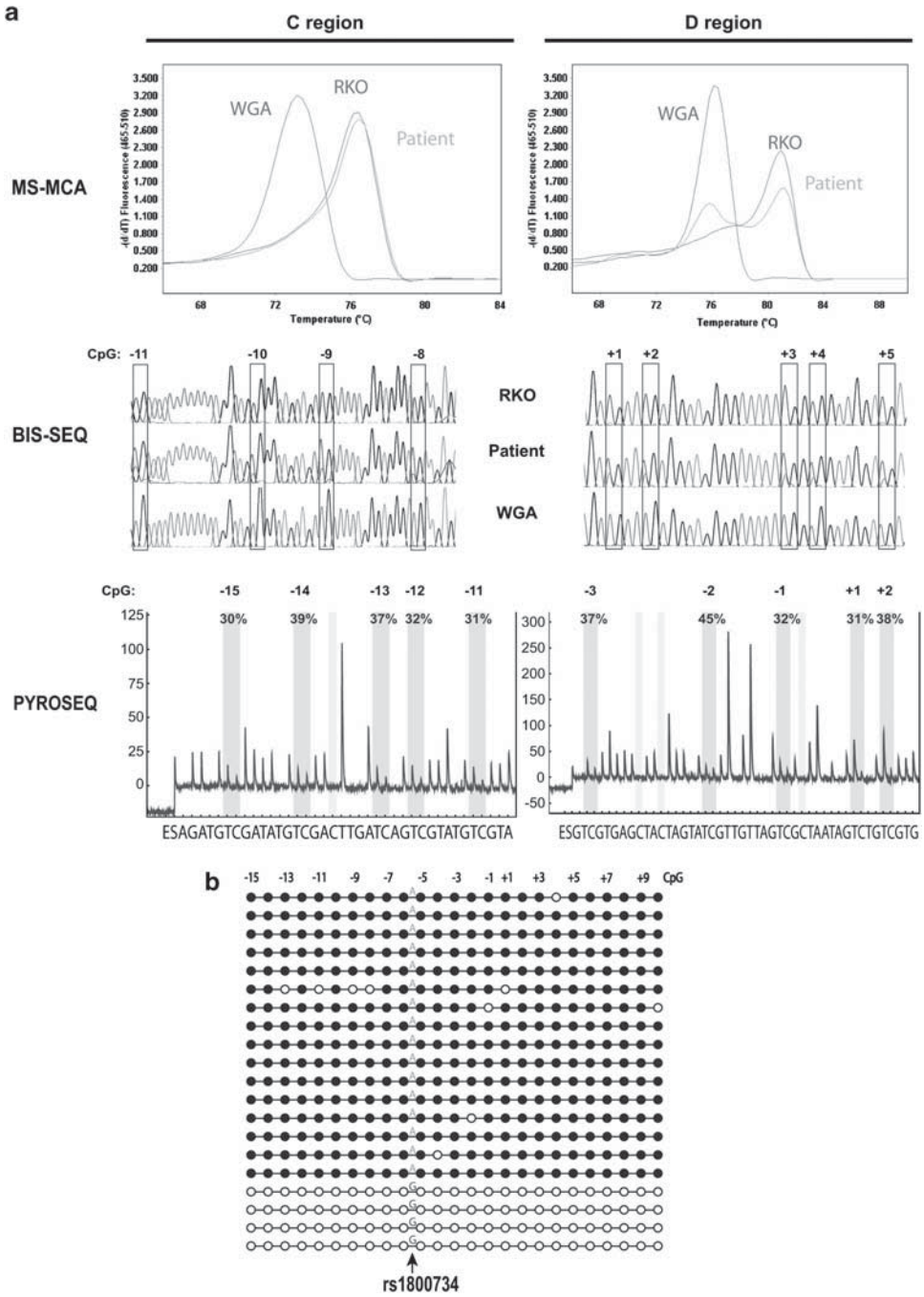


Figure 2 Confirmation of the constitutional *MLH1* epimutation of case 1. (**a**) Analysis of the *MLH1* promoter C and D regions by methylation-specific melting curve analysis (MS-MCA), bisulfite sequencing (BIS-SEQ) and pyrosequencing (PYROSEQ). Top panel: MS-MCA of *MLH1* promoter. In the analysis of C region, WGA DNA (unmethylated control) and RKO DNA (methylated control) show single melting peaks at 73 and 77°C, respectively. In D region, WGA and RKO melting peaks temperature are 76 and 82°C, respectively. Analysis by MS-MCA in PBL DNA from the patient 1 (green line) shows the presence of the methylated peak in both regions. Middle panel: sequence analysis of bisulfite-converted DNA. WGA DNA shows T at each CpG analyzed, consistent with complete modification of the DNA. RKO DNA shows C at each CpG. Patient DNA shows a mixture of T and C at CpG sites, attributable to partial methylation. Bottom panel: representative pyrograms obtained in the analysis of C and D *MLH1* promoter regions in PBL DNA from the patient. The peaks within the shaded area of the pyrogram correspond to the CpG interrogated. Percentage methylation at each site is calculated as the C:T ratio of peak heights (representing methylated:unmethylated cytosine). *x* axis represents the nucleotide dispensation order. *y* axis units are arbitrary representing light intensity. (**b**) Clonal bisulfite sequencing of the *MLH1* promoter in PBL DNA from the epimutation carrier 1. Each horizontal line represents a single allele. CpG dinucleotides are depicted by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. The allele at rs1800734 (c.−93G>A) is indicated as A or G. Methylation is confined to the A allele. Each CpG analyzed is numbered according to its position relative to the translation initiation codon.

epimutated allele is only shared by the patient and one of his daughters (Figure 1a). The lack of availability of biological material from the mother has precluded us from analyzing the presence of the epimutation in her. These results confirmed that the epimutated allele is maternally inherited in the patient, and that methylation is erased in the patient's daughter who inherited the same allele.

Table 4 Quantification of *MLH1* promoter methylation by pyrosequencing

MLH1 *promoter C region*

			CpG position								
Family	*Individual*	*Sample*	*−15*	*−14*	*−13*	*−12*	*−11*	*Mean*	*SD*	*Min*	*Max*
A	II.1 (case 1)	PBL	32.0	38.1	36.1	31.7	33.6	34.3	2.8	31.7	38.1
		sperm	2.1	0.0	3.8	2.1	1.4	1.9	1.4	0.0	3.8
B	case 34	PBL	22.1	21.6	20.1	17.1	17.7	19.7	2.3	17.1	22.1
		RKO	95.5	96.5	94.2	92.6	95.9	94.9	1.6	92.6	96.5
		CEPH	2.2	2.15	3.6	2.55	2.3	2.6	0.6	2.2	3.6

MLH1 *promoter D region*

			CpG position										
Family	*Individual*	*Sample*	*−6*	*−5*	*−4*	*−3*	*−2*	*−1*	*1*	*2*	*Mean*	*SD*	*Min*	*Max*
A	II.1 (case 1)	PBL	39.0	50.0	38.9	36.4	43.8	33.3	32.1	39.4	39.1	5.8	32.1	50.0
		sperm	0.0	5.3	0.0	1.6	6.5	2.9	0.0	1.7	2.3	2.5	0.0	6.5
B	case 34	PBL	19.7	20.5	19.1	19.2	17.2	18.3	19.8	19.2	19.1	1.0	17.2	20.6
		RKO	95.5	92.6	84.0	90.2	76.1	72.7	81.0	93.7	85.7	8.6	72.7	95.5
		CEPH	3.4	5.7	0.0	0.0	9.4	3.8	2.7	2.9	3.5	3.0	0.0	9.4

Each sample was run in triplicates. Methylation at each specific CpG was calculated as the mean of the triplicates. Values for each specific CpG within the region are given in percentage. Average percentage of methylation of the whole region was calculated as the mean for the five CpGs analyzed in C region and the eight CpGs in the D region. Both peripheral blood lymphocytes (PBL) and sperm from the proband (II.1) were analyzed. DNA from the colorectal cancer cell line RKO was used as positive control. CEPH DNA was used as negative control. Each CpG analyzed is numbered according to its position relative to the translation initiation codon.

Inactivation of the non-methylated allele in tumor tissue. We explored the nature of the putative second hit in the patient's sigmoid colon cancer. Full exonic sequencing of the *MLH1*-coding region did not identify any additional mutation. LOH was evidenced at *MLH1* rs1799977 and D3S1611 (data not shown). Retention of heterozygosity was observed at the distal marker D3S3564, whereas LOH was not evaluable at markers D3S1612, D3S3623 and D3S1298 due to their instability. These results point to the loss of the wild-type *MLH1* allele in tumor DNA. MLPA analysis in tumor DNA was not conclusive, probably owing to the poor quality of tumor FFPE-DNA.

DISCUSSION

We identified two *bona fide MLH1* epimutations and one of them has been extensively characterized. In previous reports, *MLH1* epimutations were detected in 8–13% of patients with tumors showing MLH1 loss of expression.[13,17,20,22] We have detected this alteration in 2 out of 30 patients with *MLH1*-methylated CRC meeting Bethesda or Amsterdam criteria (6.7%) and in 2 of 14 patients with an age of onset below 50 years (14.2%), in whom no germline *MLH1* mutation was identified. This is in line with the prevalence reported by van Roon *et al*[23] in patients with *MLH1*-methylated tumors enriched for cases with an early age of onset. If we take into consideration only the ICO series, *MLH1* epimutations represent so far 2% of all LS cases.

In accordance with previous reports (reviewed in Hitchins and Ward[5]), the cases identified in this study had developed multiple LS tumors at an early age. This may not only reflect the phenotype associated with the epimutation but also the selection criteria used so far in most studies. Of note, methylation was not only detected in metachronous colon tumors but also in endometrial carcinomas as well. *BRAF* mutation was absent in four analyzed tumors from the identified epimutation carriers. However, the presence of somatic *BRAF* V600E mutation has been previously reported in tumors from three epimutation carriers,[10,12,23] representing 15.8% (3/22) of the reported cases. In our set of cases, the degree of *MLH1* methylation is highly variable among tumors from both epimutation carriers and the remaining patients. Epimutations have been detected in two of four cases where multiple tumors showed somatic *MLH1* hypermethylation.

PBL methylation levels correlated with the observed transcriptional silencing, suggesting the presence of mosaicism in case 34. Dosage of the methylated allele is important. In line with previous observations, approximately 50% of the alleles were methylated in case 1.[10,11,14,17,20,21] As reported, the functional impact of the epimutation seems clear, as it associates with monoallelic expression of *MLH1* and *EPM2AIP1* transcripts[12,13,15,17] and an allele-specific methylation pattern.[14,17,20,21] LOH in an intragenic *MLH1* microsatellite marker was detected, consistent with somatic loss of the unmethylated allele. In fact, LOH has been found to be the most frequent mechanism of inactivation of wild-type allele in tumors from epimutation carriers.[12]

So far, in all cases identified but one, the methylated allele was of maternal origin.[10,12,15,17,20] The epimutation was found in the maternally inherited allele. Although we were unable to definitively demonstrate whether the epimutation was inherited or *de novo*, this may further support the notion that this type of aberration is more likely to accumulate during the oogenesis. We were able to perform a more detailed study of the index case and descendants. While *MLH1* methylation was present in every embryonic layer of the index case, a complete erasure was observed in the spermatozoa, as reported by Hitchins *et al*.[17,30] The lack of methylation in spermatozoa does not necessarily mean that inheritance cannot occur. In fact, this was clearly demonstrated in one descendant who inherited the epimutation out of three harboring the same allele.[17] In our case, the epimutated allele was transmitted unmethylated to one of his daughters.

In spite of an extensive search, we have not been able to identify a genetic alteration underlying the epimutated allele. Genetic aberrations in *cis* (gross rearrangements in two cases (one deletion of *MLH1* exons 1 and 2, and one duplication involving the whole gene) and in a third one the variant c. −27C>A within the promoter region) have been identified as responsible for *MLH1* methylation.[13,16,19] Dominant transmission pattern is observed in these cases.

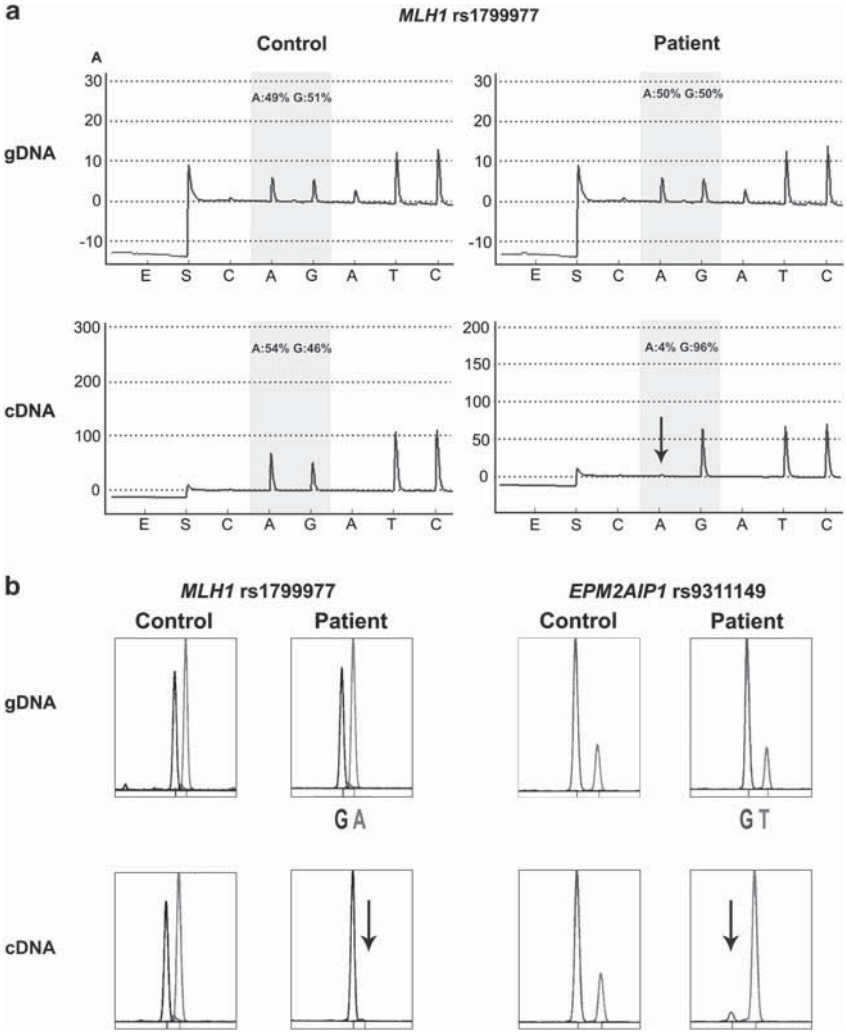


Figure 3 Transcriptional inactivation of *MLH1* and *EPM2AIP1* alleles. (**a**) Illustrative example of the pyrogram across the expressible *MLH1* rs1799977 (c.655A>G) in genomic DNA (gDNA) (top panels) and cDNA (bottom panels) derived from a heterozygous healthy control (left panels) and the epimutation carrier (right panels). The peaks within the shaded area of the pyrogram are the nucleotides at the SNP site, quantified with respect to neighboring nucleotides. Their relative values are given as percentage values above the pyrogram trace. There was a transcriptional inactivation of the 'A' allele (indicated with a downward arrow) in the cDNA of the patient with the *MLH1* epimutation. *x* axis represents the nucleotide dispensation order. *y* axis units are arbitrary representing light intensity. (**b**) Representative results of the SNuPE analysis at *MLH1* rs1799977 (c.655A>G) (left panel) and *EPM2AIP1* rs9311149 (right panel) in gDNA and cDNA derived from a heterozygous control and the epimutation carrier. Transcriptional silencing of the A allele at *MLH1* rs1799977 and T allele at *EPM2AIP1* rs9311149 in the cDNA of the patient was observed.

Dominant inheritance has been also observed in cases where no genetic alterations are detected.[10,12,15,17,20] In these cases, methylation was mosaic and associated to a shared haplotype.

Although we cannot completely rule out that aberrations have been missed, the lack of family history and the lack of vertical transmission are compatible with a *de novo* methylation occurred in the early embryo, where there is no apparent predisposing genetic mechanism that would allow for the restoration of methylation after the gametogenesis. However, this is an unsettled issue. The epimutation carrier identified in this study showed methylation confined to the A allele at rs1800734, although allele-specific methylation is not restricted to either A or G allele in other reported cases.[14,17,20,21] It is intriguing that the A allele at rs1800734 associates with somatic *MLH1* promoter methylation and increased risk of MSI CRC.[23,31–35] In addition, it has been shown that this polymorphism modifies the efficiency of *MLH1/EPM2AIP1* transcription.[36]

It is difficult to translate these findings into specific recommendations for these patients and their relatives. At this time caution is mandatory. In the presence of a detected constitutional epimutation, genetic screening of descendants is important. However, in the presence of an inherited non-methylated allele in lymphocyte DNA, two options are available. On the one hand, descendants can be counseled as relatives of a LS case where direct genetic testing has been non-informative. In this setting, it is assumed that lack of methylation in the inherited allele does not rule out that a mosaic status is present in the patient or that a non-detected genetic alteration predisposing to a late acquisition of methylation is present in this family. Alternatively, recommendation can be made based on the degree of personal and familial history of cancer. Further knowledge is needed to translate these research findings into useful information for management of patients and families.

The increasing detection of epimutations has lead to the suggestion that the diagnostic algorithm of LS might be improved. So far, the detection of somatic *MLH1* hypermethylation is often used to exclude patients from further MMR mutation analysis, based on cost effectiveness considerations.[24,37] The patients with somatic *MLH1* hypermethylation could now be considered as candidates to screen for constitutional *MLH1* epimutations. Based on the clinical presentation of the reported cases[5] and our experience, this screening could be restricted to those diagnosed earlier than 50 years or with multiple tumors the first one before the age of 60. If this was the case, MS-MLPA could be a good methodological approach. The robustness and informativeness already shown for paraffin-embedded tissues[24] has been confirmed when being used in the germline. In any case, confirmation with at least another technique (ie, pyrosequencing) would be mandatory.

In summary, *MLH1* methylation screening in PBL from patients with early-onset *MLH1*-methylated CRC allows the identification of epimutation carriers. Using this strategy we have identified two *bona fide MLH1* epimutations. In one of them, the methylated allele is from maternal origin, is present in all embryonic layers and is absent in spermatozoa. The characterization of these cases provides further evidence of the emerging entity of soma-wide MLH1 epimutation and its heritability.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the patients who participated in this study, Gemma Aiza for technical assistance, Javier Carmona for his assistance with pyrosequencing and Dr Juana Fernández for her assistance in skin fibroblast isolation and culture. This work was supported by grants from Ministerio de Ciencia e Innovación (SAF 06-06084; 09-07319), Fundació Gastroenterologia Dr Francisco Vilardell (F05-01), Ministerio de Educación y Ciencia Spanish Networks RTICCC (RD06/0020/1050, 1051), Acción en Cáncer (Instituto de Salud Carlos III), Fundación Científica AECC and NCI U19 CA 148107-02.

1 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.
2 Vasen HF, Moslein G, Alonso A *et al*: Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J Med Genet* 2007; **44**: 353–362.
3 Marra G, Boland CR: Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J Natl Cancer Inst* 1995; **87**: 1114–1125.
4 Peltomaki P, Vasen H: Mutations associated with HNPCC predisposition – update of ICG-HNPCC/INSiGHT mutation database. *Dis Markers* 2004; **20**: 269–276.
5 Hitchins MP, Ward RL: Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J Med Genet* 2009; **46**: 793–802.
6 Kuiper RP, Vissers LE, Venkatachalam R *et al*: Recurrence and variability of germline EPCAM deletions in Lynch syndrome. *Hum Mutat* 2011; **32**: 407–414.
7 Hesson LB, Hitchins MP, Ward RL: Epimutations and cancer predisposition: importance and mechanisms. *Curr Opin Genet Dev* 2010; **20**: 290–298.
8 Ligtenberg MJ, Kuiper RP, Chan TL *et al*: Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. *Nat Genet* 2009; **41**: 112–117.
9 Auclair J, Vaissiere T, Desseigne F *et al*: Intensity-dependent constitutional MLH1 promoter methylation leads to early onset of colorectal cancer by affecting both alleles. *Genes Chromosomes Cancer* 2011; **50**: 178–185.
10 Crepin M, Dieu MC, Lejeune S *et al*: Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. *Hum Mutat* 2012; **33**: 180–188.
11 Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD: A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res* 2002; **62**: 3925–3928.
12 Goel A, Nguyen TP, Leung HC *et al*: De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* 2011; **128**: 869–878.
13 Gylling A, Ridanpaa M, Vierimaa O *et al*: Large genomic rearrangements and germline epimutations in Lynch syndrome. *Int J Cancer* 2009; **124**: 2333–2340.
14 Hitchins M, Owens S, Kwok CT, Godsmark G, Algar U, Ramesar R: Identification of new cases of early-onset colorectal cancer with an MLH1 epimutation in an ethnically diverse South African cohort. *Clin Genet* 2011; **80**: 428–434.
15 Hitchins M, Williams R, Cheong K *et al*: MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2005; **129**: 1392–1399.
16 Hitchins MP, Rapkins RW, Kwok CT *et al*: Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5′UTR. *Cancer Cell* 2011; **20**: 200–213.
17 Hitchins MP, Wong JJ, Suthers G *et al*: Inheritance of a cancer-associated MLH1 germline epimutation. *N Engl J Med* 2007; **356**: 697–705.
18 Miyakura Y, Sugano K, Akasu T *et al*: Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. *Clin Gastroenterol Hepatol* 2004; **2**: 147–156.
19 Morak M, Koehler U, Schackert HK *et al*: Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing Lynch syndrome. *J Med Genet* 2011; **48**: 513–519.
20 Morak M, Schackert HK, Rahner N *et al*: Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter methylation in blood cells clinically displaying HNPCC. *Eur J Hum Genet* 2008; **16**: 804–811.
21 Suter CM, Martin DI, Ward RL: Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004; **36**: 497–501.
22 Valle L, Carbonell P, Fernandez V *et al*: MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. *Clin Genet* 2007; **71**: 232–237.
23 van Roon EH, van Puijenbroek M, Middeldorp A *et al*: Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? *BMC Cancer* 2010; **10**: 180.
24 Gausachs M, Mur P, Corral J *et al*: MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. *Eur J Hum Genet* 2012 (in press).
25 Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R: Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 1990; **6**: 575–577.
26 Deng G, Peng E, Gum J, Terdiman J, Sleisenger M, Kim YS: Methylation of hMLH1 promoter correlates with the gene silencing with a region-specific manner in colorectal cancer. *Br J Cancer* 2002; **86**: 574–579.
27 Azuara D, Rodriguez-Moranta F, de Oca J *et al*: Novel methylation panel for the early detection of colorectal tumors in stool DNA. *Clin Colorectal Cancer* 2010; **9**: 168–176.
28 Mueller J, Gazzoli I, Bandipalliam P, Garber JE, Syngal S, Kolodner RD: Comprehensive molecular analysis of mismatch repair gene defects in suspected Lynch syndrome (hereditary nonpolyposis colorectal cancer) cases. *Cancer Res* 2009; **69**: 7053–7061.
29 Borras E, Pineda M, Blanco I *et al*: MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. *Cancer Res* 2010; **70**: 7379–7391.
30 Hitchins MP, Ward RL: Erasure of MLH1 methylation in spermatozoa-implications for epigenetic inheritance. *Nat Genet* 2007; **39**: 1289.
31 Whiffin N, Broderick P, Lubbe SJ *et al*: MLH1-93G>A is a risk factor for MSI colorectal cancer. *Carcinogenesis* 2011; **32**: 1157–1161.
32 Allan JM, Shorto J, Adlard J *et al*: MLH1 -93G>A promoter polymorphism and risk of mismatch repair deficient colorectal cancer. *Int J Cancer* 2008; **123**: 2456–2459.
33 Chen H, Taylor NP, Sotamaa KM *et al*: Evidence for heritable predisposition to epigenetic silencing of MLH1. *Int J Cancer* 2007; **120**: 1684–1688.
34 Mrkonjic M, Roslin NM, Greenwood CM *et al*: Specific variants in the MLH1 gene region may drive DNA methylation, loss of protein expression, and MSI-H colorectal cancer. *PLoS One* 2010; **5**: e13314.
35 Campbell PT, Curtin K, Ulrich CM *et al*: Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. *Gut* 2009; **58**: 661–667.
36 Perera S, Mrkonjic M, Rawson JB, Bapat B: Functional effects of the MLH1-93G>A polymorphism on MLH1/EPM2AIP1 promoter activity. *Oncol Rep* 2011; **25**: 809–815.
37 Perez-Carbonell L, Alenda C, Paya A *et al*: Methylation analysis of MLH1 improves the selection of patients for genetic testing in Lynch syndrome. *J Mol Diagn* 2010; **12**: 498–504.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS:

Samples

DNA extraction of colorectal mucosa and tumour tissue from paraffin-embedded material was done after enrichment for normal and tumour cells using the QIAmp DNA Mini Kit (Qiagen). Microsatellite instability testing was performed in tumor DNA using the MSI Analysis System (Promega). Genomic DNA was extracted from peripheral blood lymphocytes (PBL) using the FlexiGene DNA kit (Qiagen, Hilden, Germany). Different samples were acquired from the epimutation carrier: skin fibroblasts, peripheral blood lymphocytes, colorectal tumor and normal adjacent mucosa, and sperm. For fibroblast isolation, a skin biopsy was cut into small pieces and digested with 160 U/ml collagenase type 1 (Sigma, St. Louis, MO) and 0.8 U/ml dispase grade 1 (Roche Diagnostics, Penzberg, Germany).[1] Fibroblasts were grown with Dulbecco's modified Eagle's medium (Gibco, Invitrogen), 10% fetal bovine serum (Gibco, Invitrogen), and penicillin/streptomycin (Gibco, Invitrogen) at 37ºC and 5% CO2. DNA from cultured fibroblasts was extracted using the Gentra Puregene Cell Kit. Sperm was washed twice in 1x SSC/ 1% SDS, then washed in 1x SSC and incubated in 0.2x SSC/ 1% SDS/ 1M 2-mercaptoethanol for 1 hr at room temperature. DNA was extracted from spermatozoa using a standard phenol-chloroform method and ethanol precipitation. Total RNA was extracted from PBL using Trizol reagent (Invitrogen, Carlsbad, CA), according to the manufacturer's instructions. cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) and random primers (Invitrogen).

MLH1 promoter methylation analyses

MS-MCA. Bisulfite converted DNA was used in a nested PCR reaction for the amplification of regions C and D of the *MLH1* promoter.[2] Each promoter region was preamplified using external primers (Table S1). Eighty ng of bisulfite modified DNA was added to MegaMix double solution (Microzone Ltd., UK) containing 2μM of each primer. PCR conditions were: 95 ºC at 10 min followed by 15 cycles of 30s at 95 ºC, 30s at 50°C, 30s at 72ºC and a final elongation step at 72 °C for 10min. The nested PCR was carried out in a LightCycler 480 II (Roche) using 1μl of amplified *MLH1* promoter fragments in 9μl of Light Cycler 480 SYBR Green I Master Kit (Roche) containing 0.5μM of internal primers. The amplification protocol was: 95ºC for 10min, followed by 40 cycles of 95ºC for 10s, 50°C for 20s, and 72ºC for 25s. Melting curve analysis was performed by heating the PCR products from 60ºC to 98ºC with an increase of 0.2°C/s whereas fluorescence was monitored continuously.

Bisulfite-Sequencing. One µl of bisulfite converted DNA was used in a 10ul-PCR reaction for the amplification of *MLH1* promoter regions C and D[2] using Double Megamix (Microzone Ltd., UK) and 0,2µM of primers (Table S1). The cycling program included 10 min at 94ºC, 35 cycles of 30s at 95°C, 30s at 50°C and 30s at 72ºC and final extension at 72ºC for 10 min. PCR products were

purified using ExoSAP-it (Affymetrix, Inc.) and sequenced using the amplification primers and BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems, Carlsbad, CA).

Clonal Bisulfite-Sequencing. One µl of bisulfite modified DNA was amplified in a PCR reaction using EcoStar DNA polymerase (Ecogen, Spain) and 0,3µM of primers (Table S1). PCR products were purified by ExoSAP-it (Affymetrix, Inc.) and cloned into pGEM-T vector (PromegaCorp, Madison, WI). In order to confirm that transformed cells contained the fragment of interest we performed a colony-PCR using M13 primers. Amplification conditions were: 10 min at 94ºC; and 35 cycles of 1min at 94°C, 1min at 55°C, 1min at 72ºC and final extension at 72ºC for 10 min. The PCR products were analyzed by agarose gel electrophoresis. Twenty individual clones were sequenced using M13 primers and BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems).

Pyrosequencing. In the case of region D, we designed 2 different sequencing primers, the first primer called as "D1" analyzed the first 3 CpGs and the second primer "D2" analyzed the last 5 CpGs in the region of interest. The program used for amplification was: 95°C for 15 min, 35 cycles of 94°C for 1 min, 1 min at the annealing temperature (Table S1), 72°C for 1 min and a final extension at 72°C for 10 min. Five µl of PCR product were evaluated for % methylation using the PyroMark Q96 MD pyrosequencer (Qiagen,Valencia, CA). If the sample failed at more than one site, it was repeated using 10 µl of PCR product. Purification and subsequent processing of the biotinylated single-stranded DNA was performed according to the manufacter's recommendations at the PyroMark Q96 Vacuum Prep Workstation (Qiagen). The pyrosequencing primers were used in a final concentration of 0.3µM. The pyrosequencing reaction was performed using each specific sequencing primer on a PyroMark Q96 MD pyrosequencer system with the Pyromark Gold Q96 reagents kit. The sequences interrogated were GAGYGGATAGYGATTTTTAAYGYGTAAGYGTATATTTTTTTAGGTAG for promoter C region, GATGGYGTAAGTTATAGTTGAAGGAAGAAYGTGAGTAYGAGGTATTGAGGTGATTGGTTGAA GG for promoter "D1" region, and YGTTGAGTATTTAGAYGTTTTTTTGGTTTTTTTGGYGTTAAAATGTYGTTYGTGGTAGGGGTTA TT for promoter "D2" region. The relative levels of the C (representing methylated) and T (representing unmethylated) nucleotides at Y positions of target CpGs sites were determined using the Pyro Q-CpG Software (Qiagen). Each sample was run in triplicates. Methylation at each specific CpG was calculated as the mean of all triplicates. Average % methylation of the whole region was calculated as the mean for the 5 CpGs in C region and the 8 CpGs at the D region.

MLH1 **allelic expression analyses**

SNuPE: PCR flanking the rs1799977 of genomic DNA and cDNA was performed using Double Megamix (Microzone Ltd., UK) and 0.1 µM of primers (Table S1). cDNA amplification conditions were: 5 min at 94ºC, 35 cycles of 1 min at 94°C, 1 min at 59°C and 1 min at 72ºC and final extension at 72ºC for 7 min. For genomic DNA the cycling program included 10 min at 94ºC, 35 cycles of 30s at 95°C, 30s at 55°C and 30s at 72ºC and final extension at 72ºC for 10 min. PCR

products were purified using ExoSAP-it (Affymetrix, Inc.) and sequenced if necessary using amplification primers and BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems). The amplified band was analyzed using the ABI PRISM SNaPshot kit (Applied Biosystems) and a specific primer (Table S1). SNaPshot reactions were carried out in a 10µl volume containing SNaPshot Multiplex Ready Reaction Mix, specific primer (0.2µM) and the purified PCR product. The cycling program included 25 cycles of 96°C for 10s, 50°C for 5s and 60ºC for 30s. Extension products were purified with 1U of shrimp alkaline phosphatase (Amersham, UK) for 15 min at 37ºC and 15min at 75ºC. The purified products were run in an ABI Prism 3130 DNA sequencer and analyzed by GeneMapper v4.0 (Applied Biosystems).

Pyrosequencing: Quantitative pyrosequencing assays were designed as previously described.[3] PCR and sequencing primers are shown in Table S1. After PCR amplification of genomic DNA and cDNA, products were sequenced on a PyroMark Q24 pyrosequencing instrument (Qiagen, Valencia, CA). A control in which the template was omitted was used to detect background signal. A nucleotide dispensation order of CAGATCTGA was used to interrogate the sequence of interest A/GTCTTTGGAAA. The proportion of A allele versus G alleles of rs1799977 were obtained using PyroMark Q24 AQ software calculations. The mean of triplicates for both DNA and cDNA were calculated for each sample.

To obtain ASE (allele-specific expression) values, both for SNuPE and pyrosequencing, we used the previously described method: cDNA (peak height major allele/ peak height minor allele) / gDNA (peak height major allele/ peak height minor allele). The final ASE value was calculated as the mean of the ASE values obtained for the triplicates studied in each sample.[4] ASE values of 1.0 indicate equal levels of expression form both alleles. ASE values <<1.0 indicate reduced expression.

REFERENCES

1 Serra E, Ars E, Ravella A *et al*: Somatic NF1 mutational spectrum in benign neurofibromas: mRNA splice defects are common among point mutations. *Hum Genet* 2001; 108: 416-429.

2 Deng G, Peng E, Gum J, Terdiman J, Sleisenger M, Kim YS: Methylation of hMLH1 promoter correlates with the gene silencing with a region-specific manner in colorectal cancer. *Br J Cancer* 2002; 86: 574-579.

3 Kwok CT, Ward RL, Hawkins NJ, Hitchins MP: Detection of allelic imbalance in MLH1 expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Fam Cancer* 2010; 9: 345-356.

4 Valle L, Serena-Acedo T, Liyanarachchi S *et al*: Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. *Science* 2008; 321: 1361-1365.

Table S1. Primers used in the current study. (*) Biotinylated labeled primer. The location of primers used in the study of the *MLH1* promoter methylation is shown in Figure S1.

Gene	Analysis	Primer name	Forward primer (5'-3')	Reverse primer (5'-3')	Ta (ºC)
MLH1					
	MS-MCA	MLH1C_PCR_ext	TATTTTTGTTTTTATTGGTTGG	TAAATACCAATCAAATTTCTCAA	50
		MLH1C_PCR_int	TGTTTTTATTGGTTGGATATTT	CCAATCAAATTTCTCAACTCTATA	50
		MLH1D_PCR_ext	AGGTATTGAGGTGATTGGTTG	CAATTCTCAATCATCTCTTTAATAACA	50
		MLH1D_PCR_int	GGTGATTGGTTGAAGGTATTTT	ATCATCTCTTTAATAACATTAACTAACC	50
	Promoter bisulfite sequencing	MLH1C_BS	TTTTAAAAAYGAATTAATAGGA	AAATACCAATCAAATTTCTCAA	50
		MLH1D_BS	AAATTTGATTGGTATTTAAGTT	CATTCTCAATCTCTTTAA	50
	Clonal promoter bisulfite sequencing	MLH1C-D_BS	TTTTAAAAAYGAATTAATAGGA	CATTCTCAATCTCTTTAA	50
	Promoter bisulfite pyrosequencing	MLH1C_PCR	GGTATTTTTGTTTTTATTGGTTGGATAT	ACTCTATAAATTACTAAATCTCTT*	57
		MLH1C_Seq	TAAAAAGAATTAATAGGAA		57
		MLH1D_PCR	TTGAGAAATTTGATTGGTATTTAAGTTGT	ACATTAACTAACCCCTAAATAACTTCCCC*	58
		MLH1D1_Seq	TGAAGGGTGGGGTTG		58
		MLH1D2_Seq	GATTGGTTGAAGGTATTTT		58
	Promoter sequencing	MLH1promoter_PCR	AACCCTTTCACCATGCTCTG	CCTCGTGCTCACGTTCTTC	59
		MLH1promoter_Seq1	TACATGCTCGGGCAGTACCT		54
		MLH1promoter_Seq2	TGAAGAGAGAGCTGCTCGTG		54
	ASE (SNuPE)	rs179997_PCR_cDNA	CACAATGCAGGCATTAGTTTCTC	AGGTACAGGAATGGGTGTGTG	59
		rs179997_PCR_gDNA	GTTTCAGTCTCAGCCATGAG	ACACATGATTCACGCCACAG	55
		rs179997_snupe		TTCTCGACTAACAGCATTTCCAAAGA	50
	ASE (pyrosequencing)	rs179997_Pyr_cDNA	GCCTCAACCGTGGACAATATTC	GCTACGGTTTTATCCTCACATCCA*	64
		rs179997_Pyr_gDNA	GCCTCAACCGTGGACAATATTC	CGACATACCGACTAACAGCATTTC*	56
		rs179997_Pyr_Seq	GGACAATATTCGCTCC		50
EPM2AIP1					
	ASE (SNuPE)	rs9311149_PCR	GTCCTGTTGTAGCAGTGAATAT	GCAGCATTGGAGAATTGGTA	59
		rs9311149_Seq1	TAGGTCCTTACCAGTTACTG		54
		rs9311149_Seq2		CATCATTAGGGAAGATCTAG	54
		rs9311149_snupe	TCCTTGAAACACTTGAACACTTGAT		50
BRAF					
	BRAF V600E screening (SNuPE)	BRAF_PCR	CCTAAACTCTTCATAATGCTT	ATAGCCTCAATTCTTACCAT	55
		BRAF_snupe	TAAAAATAGGTGATTTTGGTCTAGCTACA		50

Figure S1. Location of primers used in the study of *MLH1* promoter methylation. Map of the CpG islands encompassing the *MLH1* and *EPM2AIP1* promoters (adapted from MethPrimer program). Two CpG islands are identified (in blue), comprising the C and D promoter regions. Each small vertical red line represents a CpG site. Primer position is indicated by squares. Amplified bands covering regions C, D or both are shown as blue, brown and black lines, respectively. The translation start sites of *MLH1* and *EPM2AIP1* are indicated by +1.

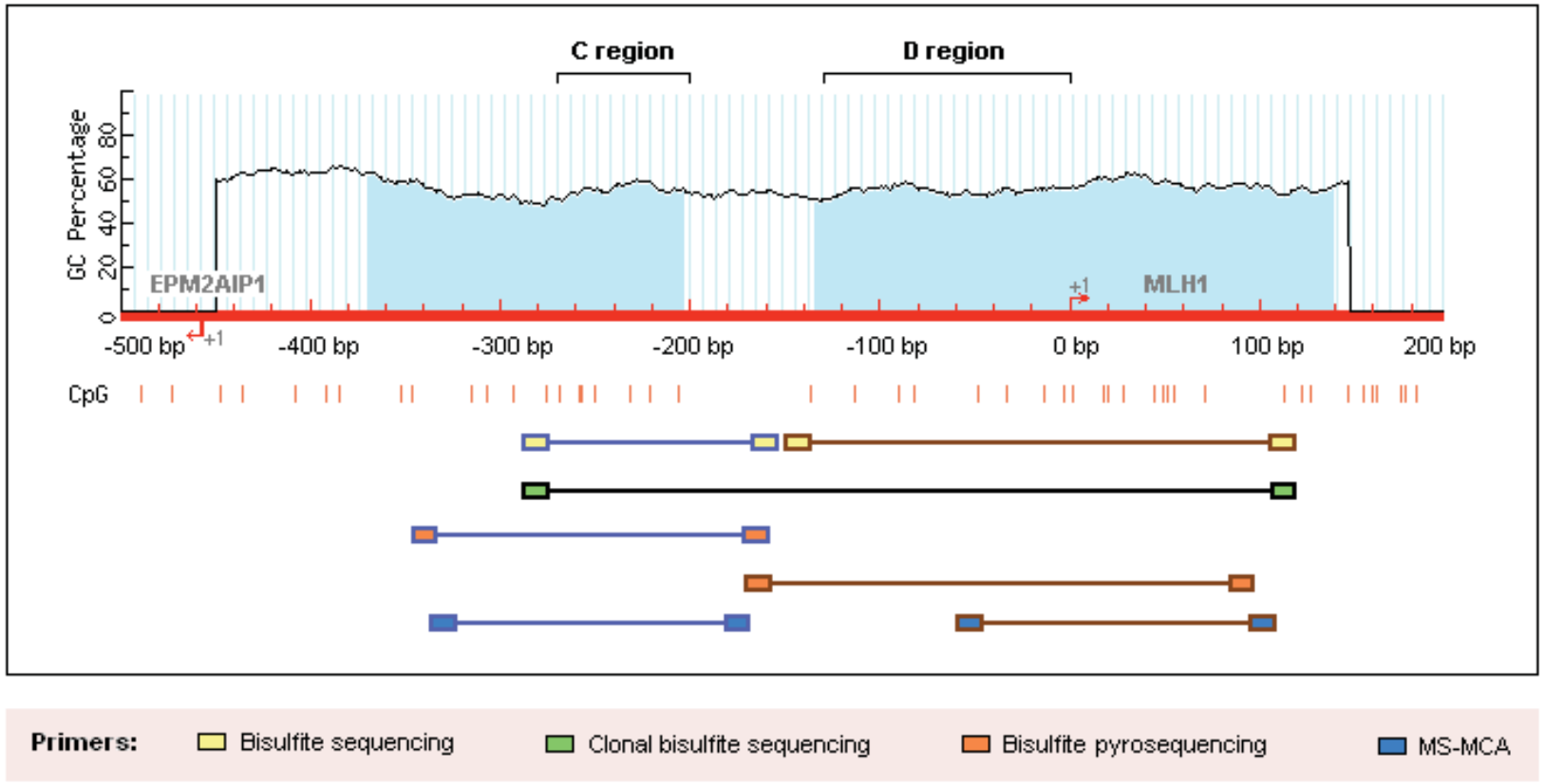
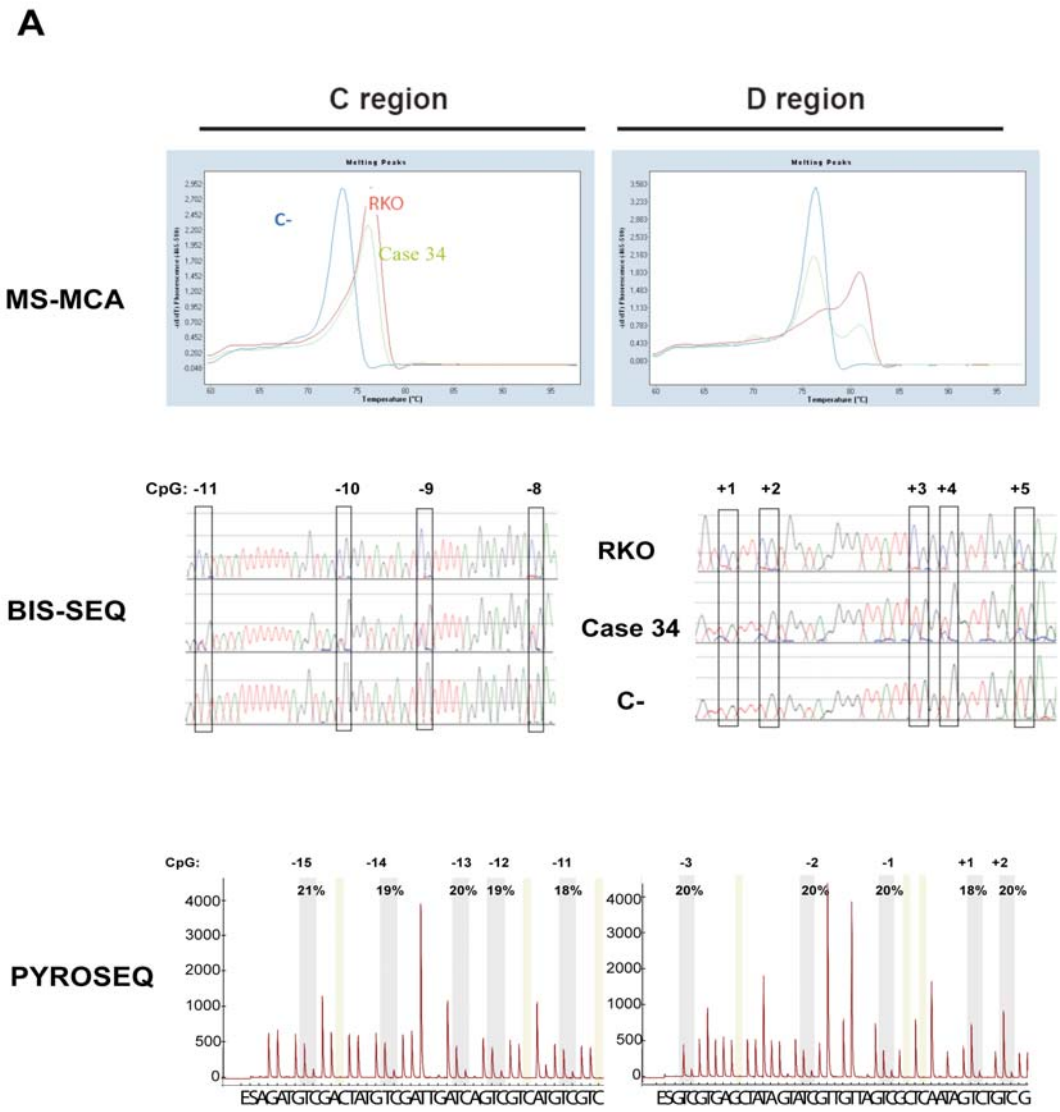
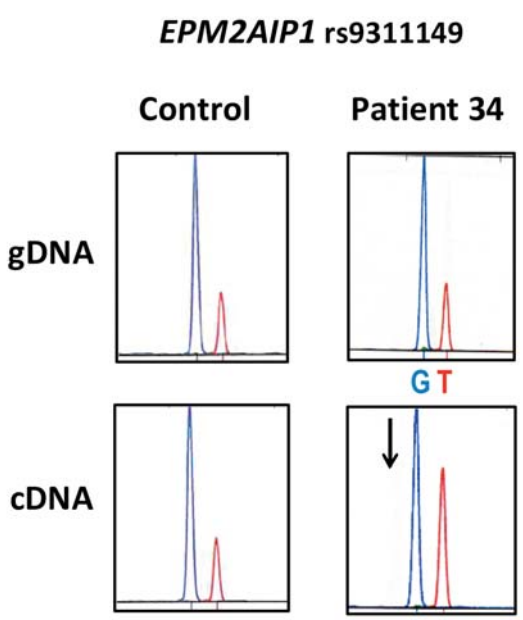


Figure S2. Confirmation of the constitutional *MLH1* epimutation of case 34. A. Analysis of the *MLH1* promoter C and D regions by MS-MCA, bisulfite sequencing (BIS-SEQ) and pyrosequencing (PYROSEQ). Top panel: MS-MCA of *MLH1* promoter. In the analysis of C region, unmethylated control (C-) and RKO DNA (methylated control) show single melting peaks at 73ºC and 77ºC, respectively. In D region, unmethylated control and RKO melting peaks temperature are 76ºC and 82ºC, respectively. Analysis by MS-MCA in PBL DNA from the patient 34 (green line) shows the presence of the methylated peak in both regions. Middle panel: Sequence analysis of bisulfite converted DNA. Unmethylated control shows T at each CpG analyzed, consistent with complete modification of the DNA. RKO DNA shows C at each CpG. Patient DNA shows a mixture of T and C at CpG sites, attributable to partial methylation. Bottom panel: Representative pyrograms obtained in the analysis of C and D *MLH1* promoter regions in PBL DNA from the patient. The peaks within the shaded area of the pyrogram correspond to the CpG interrogated. Percentage methylation at each site is calculated as the C:T ratio of peak heights (representing methylated:unmethylated cytosine). *X*-axis represents the nucleotide dispensation order. *Y*-axis units are arbitrary representing light intensity. **B. Transcriptional inactivation of *EPM2AIP1* allele.** Representative results of the SNuPE analysis at *EPM2AIP1* rs9311149 in genomic DNA and cDNA derived from a heterozygous control and the epimutation carrier. Partial transcriptional silencing of the T allele at *EPM2AIP1* rs9311149 in the cDNA of the patient was observed.



B



IDENTIFICATION OF A FOUNDER *EPCAM* DELETION IN SPANISH LYNCH SYNDROME FAMILIES

Pilar Mur*, Marta Pineda*, Atocha Romero, Jesús del Valle, **Ester Borràs**, Carla Guarinos, Alicia Canales, Matilde Navarro, Joan Brunet, Daniel Rueda, Teresa Ramón y Cajal, Víctor Moreno, Conxi Lázaro, José Luis Soto, Trinidad Caldés, Ignacio Blanco, Gabriel Capellá *(*) Aquests autors han contribuït equitativament en aquest treball.*

Manuscrit sotmès a publicació

Resum del treball: Les delecions en línia germinal en l'extrem 3' del gen *EPCAM* s'han involucrant amb la síndrome de Lynch. L'objectiu d'aquest treball va ser caracteritzar a nivell molecular les famílies espanyoles portadores de delecions en *EPCAM*. Caracterització de la deleció en *EPCAM* en 3 famílies no relacionades que procedien del sud-est d'Espanya va demostrar que compartien una deleció de 8,7 Kb, que abasta els exons 8 i 9 (g.77526_86198del8673). La mateixa deleció s'havia identificat anteriorment en una altra família espanyola. Els tumors diagnosticats en aquestes famílies eren gairebé exclusivament tumors colorectal. El patró de metilació en *MSH2* va mostrar mosaisme: els nivells mitjans de metilació van ser majors en el colon normal i en els tumors colorectals (mitjana de 27,6 i 22,3%, respectivament), al comparar-los amb els valors de metilació dels limfòcits i de les mostres de mucosa bucal (mitjana de 1,3 i 0,7%, respectivament).A més, es va detector un haplotip comú associat a la mutació que comprèn 9.9Mb,en quatre famílies portadores de la deleció g.77526_86198del8673 del gen *EPCAM*. Aquest estudi proporciona coneixements sobre les característiques clíniques i moleculars de les epimutacions en mosaic al gen *MSH2* associades a delecions en el gen *EPCAM*. La identificació de la primera mutació fundadora a Espanya identificada en el gen *EPCAM* té implicacions útils per al diagnòstic molecular de síndrome de Lynch en la població espanyola.

Identification of a Founder *EPCAM* Deletion in Spanish Lynch Syndrome Families

Pilar Mur[1]*, Marta Pineda[1]*, Atocha Romero[2], Jesús del Valle[1], Ester Borràs[1], Carla Guarinos[3], Alicia Canales[4], Matilde Navarro[1], Joan Brunet[5], Daniel Rueda[4], Teresa Ramón y Cajal[6], Víctor Moreno[7], Conxi Lázaro[1], José Luis Soto[3], Trinidad Caldés[2], Ignacio Blanco[1], Gabriel Capellá[1].
(*) These authors contributed equally to this work.

[1]Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL, L'Hospitalet de Llobregat, Spain; [2]Molecular Oncology Laboratory, Hospital Clínico San Carlos, IdISSC; Madrid; [3]Hereditary Cancer Program, Hospital Universitario, Elche, Comunidad Valenciana, Spain; [4]Laboratory of Molecular Biology, Hematology Division, Hospital Universitario 12 de Octubre, Madrid, Spain; [5]Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IdIBGI, Girona, Spain; [6]Department of Medical Oncology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; [7]Biomarkers and Susceptibility Unit, Cancer Prevention and Control Program, IDIBELL-Catalan Institute of Oncology, CIBERESP and University of Barcelona. Spain.

ABSTRACT

BACKGROUND & AIM: Germline deletions at the 3' end of *EPCAM* have been involved in the etiology of Lynch syndrome. The aim of this study was to characterize at the molecular level Spanish families harboring *EPCAM* deletions. **METHODS:** Non-commercial MLPA probes and long-range PCR amplification were used to characterize each deletion. A multiplex PCR-based test was designed to specifically amplify the mutated or the wildtype allele. Haplotyping was performed by analyzing 8 extragenic microsatellite markers and 5 intragenic *MSH2* SNPs. Methylation of *MSH2* promoter was analyzed by Methylation Specific-MLPA. **RESULTS:** Tumors diagnosed in 7 Spanish families harboring *EPCAM* deletions were almost exclusively colorectal. Mosaicism in *MSH2* methylation in *EPCAM* deletion carrier samples was observed being average methylation levels higher in normal colon and colorectal tumors (27.6 and 31.1%, respectively), than in lymphocytes and oral mucosa (1.1 and 0.7%, respectively). Characterization of the *EPCAM* deletion in 3 unrelated families coming from Southeastern Spain demonstrated that they shared an 8.7 Kb deletion encompassing exons 8 and 9 (g.77526_86198del8673). The same rearrangement was previously reported in another Spanish family (Guarinos et al., J Mol Diagn 2010). Carriers of the deletion shared a common mutation-associated haplotype comprising 9.9Mb. Additionally, in two independent families, a novel deletion of 11.6 kb in the *EPCAM* gene, g.77446_89071del11626, was identified. **CONCLUSION:** The present study provides knowledge on the clinical and molecular characteristics of mosaic *MSH2* epimutations associated with deletions in the *EPCAM* gene. The identification of an *EPCAM* founder mutation has useful implications for the molecular diagnosis of Lynch syndrome in Spain.

INTRODUCTION

Lynch syndrome (LS), caused by germline mutations in DNA mismatch repair (MMR) genes, is characterized by an autosomal dominant inheritance of early-onset colorectal cancer (CRC) and increased risk of other cancers (1, 2). Tumors from LS patients are characterized by microsatellite instability (MSI) and loss of MMR protein expression. *MLH1* and *MSH2* are the most commonly mutated MMR genes in LS, whereas mutations in *MSH6* or *PMS2* are significantly less common (3, 4). Occasionally, the presence of constitutional epimutations in *MSH2* and *MLH1* has been reported (5, 6).

Constitutional epimutations are those stable changes in gene expression that do not affect DNA sequence and that are present in normal tissues of a given individual. Approximately 45 families with *MSH2* epimutations have been reported so far (6). They associate with an autosomal dominant inheritance and have been shown secondary to the presence of deletions in the neighbouring *EPCAM* gene (formerly known as *TACSTD1*), located upstream of *MSH2*, that codifies for the EPCAM protein (6, 7). Germline deletions at the 3' end of *EPCAM* eliminate the transcription termination signal inducing the expression of a *EPCAM-MSH2* fusion transcript resulting in silencing of *MSH2* by the hypermethylation of its promoter (7, 8). The level of *MSH2* methylation correlates with EPCAM expression levels, usually high in those epithelial target tissues of Lynch syndrome (7). Tumors developed in *EPCAM* deletion carriers are almost exclusively colorectal (19-21). Twenty different deletions of variable lengths involving at least exons 8 and 9 of *EPCAM* gene have been identified (6, 9). All breakpoints involved *Alu* sequences and most of them appeared to occur by homologous recombination [reviewed in (6)]. Recently, the extent of the *EPCAM* deletions

has been related to the phenotype of deletion carriers (10).

Founder mutations are a characteristic of isolated populations and have presumably descended from a common ancestor. A number of founder mutations have been identified in MMR genes [reviewed in (11); (7, 12-21)]. Only one of them affected the *EPCAM* gene (7). In certain populations, founder mutations in MMR genes explain a substantial fraction of Lynch syndrome, providing important targets for molecular diagnostics in these particular populations (22-24).

The aim of our study was to characterize at the molecular level Spanish families harboring a deletion in the *EPCAM* gene. By doing so we identified the first founder *EPCAM* mutation reported in Spanish LS families.

METHODS

Patients and samples

Mutational screening of MMR genes were performed in suspected Lynch syndrome patients from the Catalan Institute of Oncology, either fulfilling Amsterdam or Bethesda criterias and associated with MSI phenotype or loss of MMR protein expression in tumors. In all, 139 pathogenic mutations in any of the MMR genes were identified (unpublished results). Patients were assessed at Genetic Cancer Counseling Units of ICO (Catalonia, north-east Spain). In addition, probands from 5 unrelated *EPCAM* carrier families identified at other Spanish centers (the Hereditary Cancer Biobank of Valencia and Genetics Cancer Units of Barcelona and Madrid) were included. Analyzed cases were selected on the basis of a loss of MSH2 protein and/or inconclusive immunostaining in Lynch suspected cases with no *MSH2* gene mutations detected.

Detailed family histories from at least three generations and geographic origins were obtained. Genealogical investigations based on interviews failed to identify relationships between individuals from different families. An informed consent was obtained from all subjects evaluated, and an ethics committee approved this study.

Clinical data collection included the location of tumors, the age at diagnosis, and the result of MSI testing and immunohistochemistry analysis. DNA extraction of colorectal mucosa and tumour tissue from paraffin-embedded material was done after enrichment for normal and tumour cells using the QIAmp DNA Mini Kit (Qiagen, Hilden, Germany). Genomic DNA from peripheral blood lymphocytes (PBL) was extracted using the FlexiGene DNA kit (Qiagen). DNA extraction of oral mucosa was performed by Gentra Puregene Buccal Cell Kit (Qiagen).

Forty-eight DNA samples from healthy control individuals randomly selected from a hospital based CRC case-control study (25) were used in the haplotype analysis (see below).

MSH2 promoter methylation analyses

Hypermethylation of *MSH2* promoter was analyzed by means of SALSA MS-MLPA probemix (ME011 kit; MRC Holland, Amsterdam) according to the manufacturer's recommended protocol. MS-MLPA is based on the use of probes that contain a digestion site specific to the methylation-sensitive *Hha*I enzyme. All reactions were carried out using 100 ng of DNA. The kit includes 3 probe pairs in the *MSH2* (with the respective *Hha*I sites located at -269, -193 and +124 relative to the start codon; GenBank NM_ 000251.1). MS-MLPA experiments were performed in duplicate. To generate unmethylated DNA, peripheral blood lymphocyte (PBL) DNA was amplified using the REPLI-g kit (Qiagen).

Analysis of germline rearrangements in *EPCAM* gene

Analysis of germline rearrangements in *EPCAM* gene was performed by MLPA (Multiplex Ligation-dependent Probe Amplification) technique, using SALSA MLPA kit P003-B1 (MRC-Holland, Amsterdam, The Netherlands) and 100 ng of PBL DNA. Fragment analysis was performed on an ABI 310 Genetic Analyzer and the results analyzed using the GeneScan and Genotyper software (Applied Biosystems, Foster City, CA). Calculations using peak height were performed as described (26). An average dosage quotient close to 1 is expected for individuals with two copies, whereas values close to 0.5 indicate loss of one copy. Confirmation of *EPCAM* deletions in exon 9 was achieved using SALSA MLPA kit P008 (MRC-Holland).

Characterization of *EPCAM* rearrangements

To better characterize the *EPCAM* rearrangements, we designed 3 home-made MLPA probes as described (27). MLPA reaction was performed as reported (28). The signal of each probe was adjusted after visual examination of preliminary results by modifying the concentration of the probe-mix.

In order to determine the deletion breakpoints, we performed long-range PCR of gDNA using the Expand Long Template PCR System (Roche Diagnostics GmbH, Mannheim, Germany) or

TaKaRa LA Taq (Takara Bio Inc., Otsu, Shiga Japan) using different combinations of primers pairs (**Supplementary Table S1**). PCR products containing the expected rearrangement were purified using GFX PCR DNA and the Gel Band Purification Kit (Illustra, GE Healthcare UK limited, Buckinghamshire, United Kingdom). Isolated PCR fragments were sequenced by walking on both strands using the Big Dye V3.1 Terminator Kit (Applied Biosystems, Forster City, CA, USA) and an automated sequencer (**Supplementary Table S1**). The interspersed repeats in regions around the breakpoints were examined using the Repeat Masker Program. A multiplex PCR-based test was designed using a set of primers that specifically amplified the mutated or the wild-type allele. Primer sequences are detailed in **Supplementary Table S1** (conditions available upon request).

Haplotype analysis

Haplotype analysis was performed using 8 extragenic microsatellite markers (D2S2331, D2S2328, D2S119, D2S288, D2S1248, D2S123, D2S378, D2S2957) and 5 intragenic *MSH2* single nucleotide polymorphisms (rs2303426, rs3732182, rs12998837, rs3732183, rs2303428) covering 21 Mb around *EPCAM-MSH2* (**Supplementary Table S1**; conditions available on request). DNA samples from 17 *EPCAM* deletion carriers and 5 non-carriers of the four Lynch syndrome families and 50 control individuals were analyzed, including individuals that come from the same area of origin of the *EPCAM* families. 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 haplotype in the control population was estimated using PHASE2 [10 000 iterations after 50 000 burn-in and thinning factor 5; (29)].

RESULTS

MLPA analysis of *MMR* genes in Lynch syndrome suspected patients lead to the identification of 7 *EPCAM* deletion probands (**Figure 1**). Further DNA testing in these families identified 17 additional *EPCAM* mutation carriers, 4 obligated carriers and 6 non-carriers (**Table 1**). Tumors diagnosed in carriers were of the digestive tract and primarily colorectal (**Table 1**). The mean age at first diagnosis was 43.7 years (SD 8.3, range 30-56). At the time of the study 7 deletion carriers were asymptomatic; 2 of them older than 60 years (**Table 1**).

The hypermethylation status of *MSH2* gene promoter was analyzed by MS-MLPA in different samples (2 oral mucosa, 6 PBL, 6 normal colon or 6 CRC) from some of the *EPCAM* deletion carriers (3 carriers of family B, 2 carriers from families A and C and 1 carrier from families D, F and G) (**Table 2**). *MSH2* methylation levels were higher than 10% in at least one of the 3 MLPA probes in the 6 normal colonic tissues and in 5 of the 6 tumors analyzed. The average methylation of all probes in normal colon samples and colorectal tumors was 27.6% (mean range 11.8-55.5%) and 31.1% (mean range 2.4-72.5%), respectively. In contrast, lower levels of methylation were found in 6 PBL (mean range 0-3.6%) and 2 oral mucosa samples (0-1.3%).

Home-made MLPA analysis was used to initially characterize the *EPCAM* deletion in 3 unrelated families coming from Southeastern Spain, confirming their extension to the 3'UTR of *EPCAM* (**Table 1**, A to C, and **Figure 2**). Subsequently, long-range PCR analysis in genomic DNA from the 3 probands identified the same rearrangement. It was an 8.7Kb deletion encompassing *EPCAM* exons 8 and 9 and the 3'UTR (g.77526_86198del8673 from AC079775.6) (**Figure 2**). The breakpoint junction involved *Alu* elements (*AluSp* and *AluSx*) and revealed a 32bp stretch of microhomology, suggesting homologous recombination as a likely mechanism involved in this rearrangement (**Supplementary Figure S1**).

A multiplex PCR assay was designed for the diagnostic testing of the identified g.77526_86198del8673 mutation (**Supplementary Figure S2**), ruling out its presence in the other *EPCAM* positive families (**Table 1**, D to G). In two of these cases (D and E) further characterization identified a novel *EPCAM* deletion of 11.6kb (g.7746_89071del11626) (**Figure 2**). Both families met Amsterdam I criteria (**Table 1**). The breakpoint was located within two *AluSp* elements and revealed a 19bp stretch of microhomology (**Supplementary Figure S1**).

At the time of characterization, the *EPCAM* g.77526_86198del8673 deletion was reported in another Spanish family (9). The identification of the same *EPCAM* rearrangement in 4 Spanish families pointed to a common ancestor. Of note, ancestors of these families came from Southeastern Spain (**Table1**). To confirm this hypothesis, available DNA samples from mutation carriers and non-carriers from the 4 families were used in the haplotype analysis. A mutation-associated haplotype comprising 9,9Mb on chromosome 2 was shared by all carriers (**Table 3**). The haplotype stretched from microsatellite D2S2328 upstream of *EPCAM* to microsatellite D2S1248 downstream of *MSH2*. In families A, B and C the shared haplotype

extended approximately 11Mb (from marker D2S2328 to D2S123). In the family reported by Guarinos and collaborators (9), the common haplotype was reduced due to an intrafamilial recombination event at marker D2S123. The haplotype identified among carriers was not observed in the control population studied (data not shown). The obtained results confirmed the founder origin of the *EPCAM* mutation g.77526_86198del8673 in Spanish Lynch syndrome families. Unfortunately, the age of the mutation could not be estimated due to the small number of families.

DISCUSSION

Molecular characterization of *EPCAM* deletions let to the identification of g.77526_86198del8673 and g.77446_89071del11626 in 3 and 2 unrelated Spanish LS families, respectively. Here we also describe the *EPCAM* g.77526_86198del8673 deletion as the first reported *EPCAM* founder mutation in Spanish LS families. It is the only *EPCAM* mutation identified so far in the series at the Catalan Institute of Oncology (north-eastern Spain), accounting for 2.2% of all LS cases. This is in line with the prevalence of *EPCAM* deletions reported in south-eastern Spain, Germany and The Netherlands (1.9%, 2.8% and 1.1%, respectively) (6, 9).

The results obtained in the methylation analysis of *EPCAM* deletion carrier samples confirmed the previously reported mosaicism in the *MSH2* methylation pattern (7, 30). Minimal or undetectable *MSH2* promoter methylation levels were detected in PBL and oral mucosa samples, whereas methylation was evident in all colonic tissue samples and in 5 of the 6 tumors analyzed. Interestingly, in one CRC *MSH2* methylation level was low (average mean 2.4%). This fact might be caused by loss of the methylated allele due to a third hit. Unfortunately, the small amount of available sample prevented further analyses. It has been proposed that the analysis of *MSH2* methylation status in tumors for the confirmation of *EPCAM* mutations as a LS cause (9) as well as for the selection of candidate patients to *EPCAM* mutational analysis (31, 32). Based on our results, we recommend the use of normal colonic tissues instead of tumors for the aforementioned studies.

Some studies have demonstrated that the loss of EPCAM protein expression detected by immunohistochemistry (IHC) in tumors showing loss of MSH2 protein expression is highly specific in the identification of *EPCAM* deletion carriers (33-35), suggesting EPCAM IHC might be included in the diagnostic algorithm of LS. In our study, patients diagnosed with tumors showing loss of MSH2 expression were initially screened for germline rearrangements using an MLPA kit that includes probes for the analysis of both *MSH2* and *EPCAM* rearrangements. In this context, EPCAM IHC should be relevant in cases with a non-evaluable MSH2 IHC result.

The spectrum of tumors developed in carriers from Spanish families harboring the *EPCAM* deletions was mostly colorectal, with no endometrial cancer reported. Our results add further evidence to previous studies concluding that *EPCAM* deletion carriers have high risk of colorectal tumors and low risk of endometrial cancer compared to carriers of *MSH2* mutations (10, 14, 36-39). This observation might support the adaptation of screening recommendations in *EPCAM* female carriers with a focus on CRC (37). Indeed, endometrial cancer has been reported in only 3 from a series of 94 female *EPCAM* carriers, all of them harboring large *EPCAM* deletions extending close to the *MSH2* promoter. This suggests that the extent of the deletion could be related to the risk of endometrial cancer (10). If this is eventually confirmed, molecular characterization of *EPCAM* deletions should be mandatory to ensure proper risk assessment of female carrier patients.

The novel *EPCAM* deletion g.77446_89071del11626 and the founder g.77526_86198del8673 shared the same *AluSp* sequence at the 5´ breakpoint. Moreover, the founder g.77526_86198del8673 *EPCAM* mutation involves the same *Alu* repeat pair than the g.77436_86109del8674 deletion, previously reported in 3 unrelated families (6, 8). The microhomology observed at the breakpoints has been associated with deletions of various lengths, probably originating from recombination events at different positions within the same *Alu* repeat pair (40).

Previous to this study, only one founder *EPCAM* mutation (c.859-1462_*1999del) has been reported (7). It was a 4.9kb deletion found in 16 Dutch and 1 American families (6, 7, 37). In that case, the mutation-associated haplotype comprised 6,2Mb (7). The mutation-associated haplotype reported here comprises a bigger 9,9Mb region. In Spain, six MMR founder mutations have been previously described: 4 in *MSH2* and 2 in *MLH1* (12, 16, 41, 42). Interestingly, the founder *EPCAM* g.77526_86198del8673 represents the third Alu-mediated founder deletion within *EPCAM-MSH2* region identified in the Spanish Lynch syndrome population (42, 43). The increasing number of MMR founder mutations identified and the consequent design of quick and cheaper tests to detect them, will contribute to the development of population-specific cost-effective screening strategies for LS.

In conclusion, we have provided genetic evidence that the *EPCAM* g.77526_86198del8673 deletion is a Spanish founder mutation causing LS. Our findings will have useful implications for molecular diagnosis of LS in our setting.

ACKNOWLEDGMENTS

We thank the patients who participated in this study and Gemma Aiza for technical assistance.

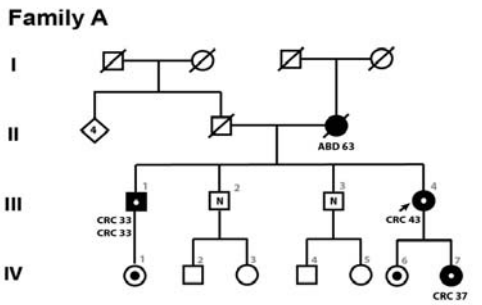
REFERENCES

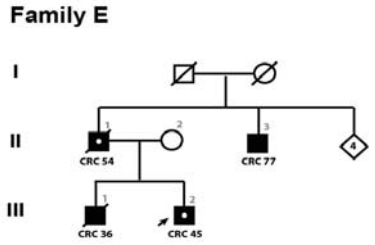
1. Lynch HT, Lynch JF, Lynch PM, Attard T. Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. Fam Cancer 2008;7:27-39.
2. Vasen HF, Moslein G, Alonso A, Bernstein I, Bertario L, Blanco I*, et al.* Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). J Med Genet 2007;44:353-62.
3. Marra G, Boland CR. Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. J Natl Cancer Inst 1995;87:1114-25.
4. Peltomaki P, Vasen H. Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 2004;20:269-76.
5. Hitchins MP, Ward RL. Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. J Med Genet 2009;46:793-802.
6. Kuiper RP, Vissers LE, Venkatachalam R, Bodmer D, Hoenselaar E, Goossens M*, et al.* Recurrence and variability of germline EPCAM deletions in Lynch syndrome. Hum Mutat 2011;32:407-14.
7. Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M*, et al.* Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 2009;41:112-7.
8. Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. Hum Mutat 2009;30:197-203.
9. Guarinos C, Castillejo A, Barbera VM, Perez-Carbonell L, Sanchez-Heras AB, Segura A*, et al.* EPCAM germ line deletions as causes of Lynch syndrome in Spanish patients. J Mol Diagn 2010;12:765-70.
10. Kempers MJ, Kuiper RP, Ockeloen CW, Chappuis PO, Hutter P, Rahner N*, et al.* Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. Lancet Oncol 2011;12:49-55.
11. Pineda M, Gonzalez S, Lazaro C, Blanco I, Capella G. Detection of genetic alterations in hereditary colorectal cancer screening. Mutat Res 2010;693:19-31.
12. Borras E, Pineda M, Blanco I, Jewett EM, Wang F, Teule A*, et al.* MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. Cancer Res 2010;70:7379-91.
13. Pinheiro M, Pinto C, Peixoto A, Veiga I, Mesquita B, Henrique R*, et al.* A novel exonic rearrangement affecting MLH1 and the contiguous LRRFIP2 is a founder mutation in Portuguese Lynch syndrome families. Genet Med 2011;13:895-902.
14. Perez-Cabornero L, Infante Sanz M, Velasco Sampedro E, Lastra Aras E, Acedo Becares A, Miner Pino C*, et al.* Frequency of rearrangements in Lynch syndrome cases associated with MSH2: characterization of a new deletion involving both EPCAM and the 5' part of MSH2. Cancer Prev Res (Phila) 2011;4:1556-62.
15. Borelli I, Barberis MA, Spina F, Cavalchini GC, Vivanet C, Balestrino L*, et al.* A unique MSH2 exon 8 deletion accounts for a major portion of all mismatch repair gene mutations in Lynch syndrome families of Sardinian origin. Eur J Hum Genet 2012; in press.
16. Menéndez M, Castellví-Bel S, Pineda M, de Cid R, Muñoz J, González S*, et al.* Founder effect of a pathogenic MSH2 mutation identified in Lynch syndrome Spanish families. Clin Genet 2010;78:186-90
17. Goldberg Y, Porat RM, Kedar I, Shochat C, Galinsky D, Hamburger T*, et al.* An Ashkenazi founder mutation in the MSH6 gene leading to HNPCC. Fam Cancer 2010;9:141-50.
18. Raskin L, Schwenter F, Freytsis M, Tischkowitz M, Wong N, Chong G*, et al.* Characterization of two Ashkenazi Jewish founder mutations in MSH6 gene causing Lynch syndrome. Clin Genet 2011;79:512-22.
19. Tang R, Hsiung C, Wang JY, Lai CH, Chien HT, Chiu LL*, 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.
20. Tomsic J, Liyanarachchi S, Hampel H, Morak M, Thomas BC, Raymond VM*, et al.* An American founder mutation in MLH1. Int J Cancer 2012;130:2088-95.
21. Tomsic J, Senter L, Liyanarachchi S, Clendenning M, Vaughn CP, Jenkins MA*, et al.* Recurrent and founder mutations in the PMS2 gene. Clin Genet 2012; in press.
22. Froggatt NJ, Green J, Brassett C, Evans DG, Bishop DT, Kolodner R*, 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.
23. Foulkes WD, Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C*, 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.

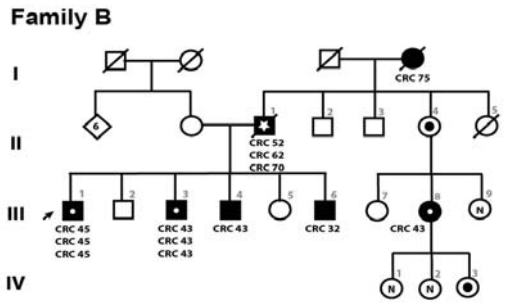
24. Nystrom-Lahti M, Kristo P, Nicolaides NC, Chang SY, Aaltonen LA, Moisio AL*, et al.* Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat Med 1995;1:1203-6.
25. Moreno V, Gemignani F, Landi S, Gioia-Patricola L, Chabrier A, Blanco I*, et al.* Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. Clin Cancer Res 2006;12:2101-8.
26. Taylor CF, Charlton RS, Burn J, Sheridan E, Taylor GR. Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. Hum Mutat 2003;22:428-33.
27. del Valle J, Feliubadalo L, Nadal M, Teule A, Miro R, Cuesta R*, et al.* Identification and comprehensive characterization of large genomic rearrangements in the BRCA1 and BRCA2 genes. Breast Cancer Res Treat 2009;122:733-43.
28. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002;30:e57.
29. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003;73:1162-9.
30. Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF*, et al.* Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. Nat Genet 2006;38:1178-83.
31. Niessen RC, Hofstra RM, Westers H, Ligtenberg MJ, Kooi K, Jager PO*, et al.* Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer 2009;48:737-44.
32. Rumilla K, Schowalter KV, Lindor NM, Thomas BC, Mensink KA, Gallinger S*, et al.* Frequency of deletions of EPCAM (TACSTD1) in MSH2-associated Lynch syndrome cases. J Mol Diagn 2011;13:93-9.
33. Huth C, Kloor M, Voigt AY, Bozukova G, Evers C, Gaspar H*, et al.* The molecular basis of EPCAM expression loss in Lynch syndrome-associated tumors. Mod Pathol 2012;25:911-6.
34. Kloor M, Voigt AY, Schackert HK, Schirmacher P, von Knebel Doeberitz M, Blaker H. Analysis of EPCAM protein expression in diagnostics of Lynch syndrome. J Clin Oncol 2011;29:223-7.
35. Musulen E, Blanco I, Carrato C, Fernandez-Figueras MT, Pineda M, Capella G*, et al.* Usefulness of epithelial cell adhesion molecule expression in the algorithmic approach to Lynch syndrome identification. Hum Pathol 2012; in press.
36. Grindedal EM, Blanco I, Stormorken A, Maehle L, Clark N, Gonzalez S*, et al.* High risk of endometrial cancer in colorectal cancer kindred is pathognomonic for MMR-mutation carriers. Fam Cancer 2009;8:145-51.
37. Lynch HT, Riegert-Johnson DL, Snyder C, Lynch JF, Hagenkord J, Boland CR*, et al.* Lynch syndrome-associated extracolonic tumors are rare in two extended families with the same EPCAM deletion. Am J Gastroenterol 2011;106:1829-36.
38. Goecke T, Schulmann K, Engel C, Holinski-Feder E, Pagenstecher C, Schackert HK*, 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.
39. Lagerstedt Robinson K, Liu T, Vandrovcova J, Halvarsson B, Clendenning M, Frebourg T*, et al.* Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. J Natl Cancer Inst 2007;99:291-9.
40. Vissers LE, Bhatt SS, Janssen IM, Xia Z, Lalani SR, Pfundt R*, et al.* Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. Hum Mol Genet 2009;18:3579-93.
41. Medina-Arana V, Barrios Y, Fernandez-Peralta A, Herrera M, Chinea N, Lorenzo N*, 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.
42. Perez-Cabornero L, Borras Flores E, Infante Sanz M, Velasco Sampedro E, Acedo Becares A, Lastra Aras E*, et al.* Characterization of new founder Alu-mediated rearrangements in MSH2 gene associated with a Lynch syndrome phenotype. Cancer Prev Res (Phila) 2011;4:1546-55.
43. Hitchins MP, Burn J. Alu in Lynch syndrome: a danger SINE? Cancer Prev Res (Phila) 2011;4:1527-30.

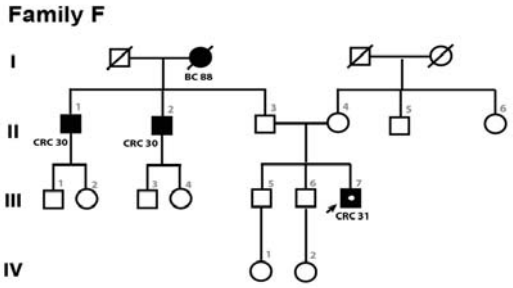
FIGURES

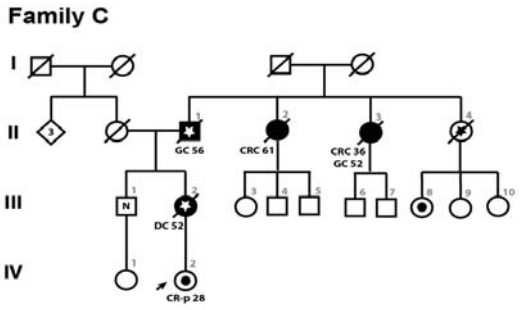
Figure 1. Pedigrees of *EPCAM* deletion carrier families. Probands are indicated by an arrow. Cancer localization (CRC, colorectal cancer; GC, gastric cancer; DC, duodenal cancer; ABD, abdominal cancer; LC, lung cancer; CR-p, colorectal polyp) and age at diagnosis are indicated in affected members. Carriers are depicted by dots, obligate carriers by stars and noncarriers by the symbol N.

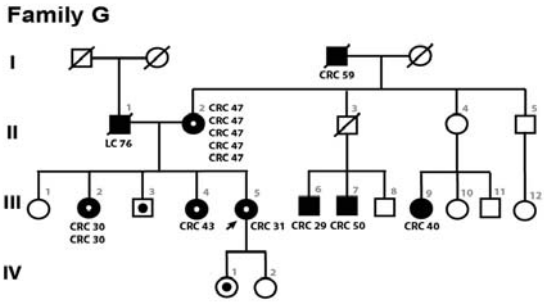












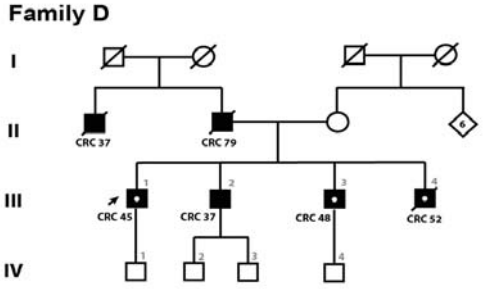
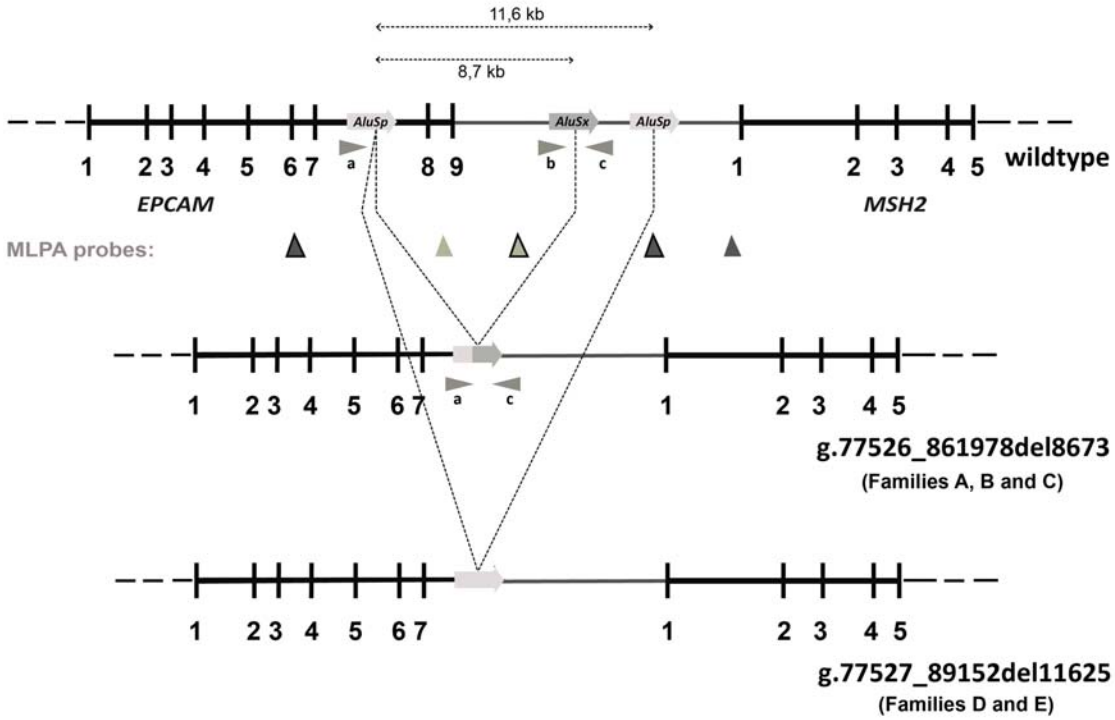


Figure 2. Characterization of the *EPCAM* g.77526_86198del8673 and g.77446_89071del11626 rearrangements. Schematic representation of the genomic structures of the wild-type and deleted *EPCAM* alleles. Positions of the used MLPA probes are indicated by arrowheads (home-made probes are outlined in black), and arrowheads located within the deleted region are shown in light gray. Primers used in the diagnostic predictive test of g.77526_86198del8673 are shown as dark gray triangles.



TABLES

Table 1. Clinical and molecular characteristics of carriers of *EPCAM* deletions. *Index case; Abbreviations: ICO, Institut Català d'Oncologia; HCSC, Hospital Clínico San Carlos; HSP, Hospital Sant Pau; HU12O, Hospital Universitario 12 de Octubre; AC, Amsterdam criteria; BC, Bethesda criteria; M, male; F, female; CRC, colorectal cancer; DC, duodenal cancer; GC, gastric cancer; h-g ad, high-grade colorectal adenoma; MSI, microsatellite instability: +, instable; IHC, immunohistochemical analysis of MMR proteins in tumor tissue: + expression; - loss of expression; NE, not evaluable result; ND, not determined.

Family	Referral center	Region of origin	*EPCAM* deletion	Clinical criteria	Individual	Gender	Current age/ Age at death	Mutational status	Type of tumor	Age at diagnosis	MSI analysis	IHC		
												MLH1	MSH2	MSH6
A	ICO, Girona	Almería, Andalucía	g.77526_86198del8673	AC	III-1	M	37	Carrier	CRC	33	+	+	NE	NE
									CRC	33				
					III-4*	F	43	Carrier	CRC	45	+	+	NE	NE
					IV-1	F	22	Carrier	-					
					IV-6	F	31	Carrier	-					
					IV-7	F	42	Carrier	CRC	37				
B	ICO, Badalona	Almería, Andalucía	g.77526_86198del8673	AC	II-1	M	71	Obligated c.	CRC	52				
									CRC	62	+	+	-	-
									CRC	70				
					II-4	F	73	Carrier	-					
					III-1*	M	56	Carrier	CRC	45	+	+	-	-
									CRC	45			-	
									CRC	45			-	
					III-3	M	44	Carrier	CRC	43	+	+	-	-
									CRC	43			-	
									CRC	43			-	
					III-8	F	45	Carrier	CRC	43				
					IV-3	F	20	Carrier	-					
C	ICO, Barcelona	Jaén, Andalucía	g.77526_86198del8673	AC	II-1	M	56	Obligated c.	GC	56				
					II-4	F	25	Obligated c.	-					
					III-2	F	52	Obligated c.	DC	52	+	+	-	-
					III-8	F	61	Carrier	-					
					IV-2*	F	28	Carrier	h-g ad	28	+	+	-	-
D	HCSC and HU12O, Madrid	Cuenca, Castilla la Mancha	g.77527-89152_del11625	AC	III-1*	M	52	Carrier	CRC	45	+	+	-	-
					III-3	M	61	Carrier	CRC	48	+	+	-	-
									CRC	54				
					III-4	M	52	Carrier	CRC	52	+	+	-	-
E	HCSC, Madrid	Madrid	g.77527-89152_del11625	AC	II-1	M	54	Carrier	CRC	54	+	+	-	-
					III-2*	M	57	Carrier	CRC	45	+	+	-	-
F	HSP, Barcelona	Tarragona, Catalunya	ND	BC	III-7*	M	35	Carrier	CRC	31		+	-	-
G	HU12O, Madrid	ND	ND	AC	II-2	F		Carrier	CRC	47	+	+	-	-
									CRC	47				
									CRC	47				
									CRC	47				
									CRC	47				
					III-2	F	56	Carrier	CRC	30				
									CRC	30				
					III-3	M	54	Carrier	-					
					III-4	F	53	Carrier	CRC	43				
					III-5*	F	51	Carrier	CRC	31	+	+	-	-
					IV-1	F	21	Carrier	-					

Table 2. MS-MLPA methylation analysis of the *MSH2* promoter in lymphocyte, oral mucosa, normal colon and CRC DNA from *EPCAM* deletion carriers. Methylation levels above 25% and in the range 10-25% are highlighted in dark and light grey, respectively. (*) Nucleotide position relative to the translation start codon.

Family	Individual	Sample	***MSH2* methylation (%)** MLPA probe position*:			***MSH2* methylation mean (%)**
			-270	**-193**	**+126**	
A	III-1	*Lymphocyte*	0.0	0.0	0.0	0.0
		Colorectal tumor	18.9	85.5	34.4	46.3
	III-4	*Lymphocyte*	0.0	0.0	0.0	0.0
		Oral mucosa	0.0	0.0	0.0	0.0
		Normal colon	12.8	24.3	18.2	18.4
		Colorectal tumor	8.9	18.2	27.6	18.2
B	II-3	*Normal colon*	9.7	27.2	12.2	16.4
	III-1	*Normal colon*	17.9	58.6	38.5	38.4
	III-3	*Lymphocyte*	7.5	3.2	0.0	3.6
		Normal colon	21.3	34.8	18.5	24.9
		Colorectal tumor	7.1	0.0	0.0	2.4
C	III-2	*Normal colon*	5.1	21.7	8.6	11.8
	IV-2	*Lymphocyte*	5.2	0.0	0.0	1.7
		Oral mucosa	0.0	0.0	4.0	1.3
		Normal colon	56.2	57.0	53.3	55.5
D	III-2	*Lymphocyte*	0.9	0.0	2.5	1.1
		Colorectal tumor	25.7	10.1	26.9	20.9
F	III-7	*Colorectal tumor*	26.2	27.4	24.8	26.1
G	III-5	*Lymphocyte*	0.0	0.0	0.0	0.0
		Colorectal tumor	72.3	11.9	133.3	72.5

Table 3. Haplotypes associated to *EPCAM* g.77526_86198del8673 mutation in the identified families. Position on chromosome 2 is according to NC_000002.11. Intragenic *MSH2* markers are highlighted in light gray. Dark gray shading indicates shared haplotype. The symbol "/" indicates that the phase of the mutation haplotype cannot be established. The symbol "//" indicates that recombination has been occurred within the family. The frequencies of disease-associated alleles among a panel of control DNAs are shown at the bottom. (*) Previously described by Guarinos and collaborators (9).

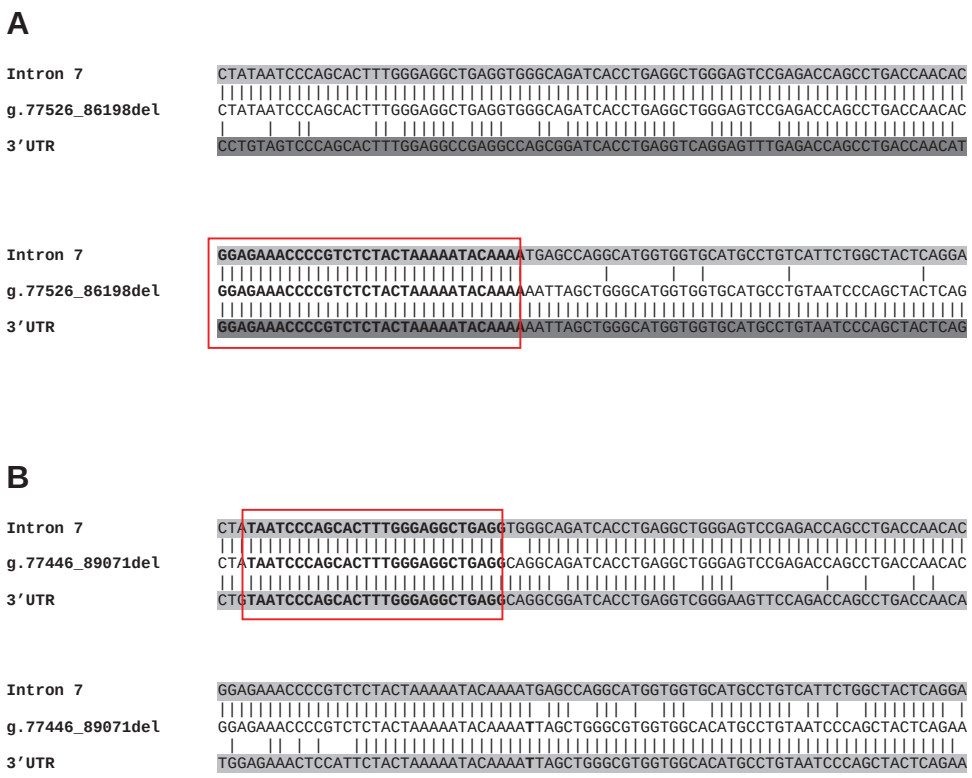
Marker (Mb on chr 2)	**D2S2331** (38.94)	**D2S2328** (40.67)	**D2S119** (44,07)	**D2S288** (46.54)	**g.77526_86198del8673** (47.61-47.62)	**rs2303426** (47.63)	**rs3732182** (47.69)	**rs12998837** (47.69)	**rs3732183** (47.69)	**rs2303428** (47.70)	**D2S1248** (50.60)	**D2S123** (51.28)	**D2S378** (57.30)	**D2S2957** (60.47)
Family														
A	**131**	128/**132**	212/**220**	101/**103**	**Yes**	C/G	**G**	**A**	**G**	**T**	**344**/352	**178**	117	226
B	129//**131**/133	**132**	**220**	**103**	**Yes**	**G**	**G**	**A**	**G**	**T**	**344**	**178**	**123**	**234**
C	129//**131**	**132**	**220**	**103**	**Yes**	**G**	T/G	T/A	A/**G**	**T**	**344**	**178**	109/**123**	222/**234**
Guarinos, 2010 (*)	121//125/**131**	**132**	**220**	**103**	**Yes**	**G**	**G**	**A**	**G**	**T**	**344**	**178**//194	**123**//117	**234**//226
Shared haplotype		**9.9 Mb**												
Shared alleles		132	220	103	Yes	G	G	A	G	T	344			
Control frequencies (%)		**9.4**	**31.3**	**44.8**	**0**	**41.7**	**NA**	**NA**	**NA**	**88.5**	**15.6**			

SUPPLEMENTARY MATERIAL

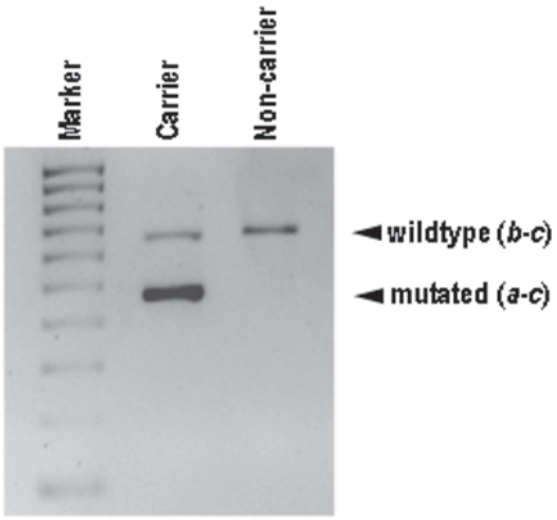
Supplementary Table S1. Primers used in the current study.

Analysis	Application	Primer name	Sense	Sequence (5'-3')	Ta (ºC)
BREAKPOINT	LR-PCR	EPCAM_LR1_F	forward	CCCGTACTTCTACTTTACCTCTAGTCCTAGTCT	60
		EPCAM_LR1_R	reverse	ACCAAATGGTTCGCTGCTGCCGTTAAT	
		EPCAM_LR3_F	forward	TTGTGCTGGTGAGTACAGAACAAG	60
		EPCAM_LR3_R	reverse	CAACCCCACTTGGCCTAAATA	
		EPCAM_LR6_F	forward	CAGAATTCAGAAAGAGGTTTGAGG	60
		EPCAM_LR6_R	reverse	CAGGCCCAGCCTTTTAACAT	
		EPCAM_LRHCSC_F	forward	TTGCTTGTAAAGGTGAATCTGG	55
		EPCAM_LRHCSC_R	reverse	TTATTGCAAATCGAGAAGTACG	
	Sequencing	EPCAM_SQ_2F	forward	GGTGGTTACTTAGTTTGTGCAT	58
		EPCAM_SQ_2R	reverse	TTCCACTCCCACATGGTAA	59
		EPCAM_SQ_4F	forward	GCCTGTAATCGCAGCACTTT	59
		EPCAM_SQ_4R	reverse	CTTGTTGCCCAAGCTGGAGT	62
		EPCAM_SQ_5F	forward	GTGGCTCATGCCTGTAATCC	59
		EPCAM_SQ_5R	reverse	CCCTTGACAACCGCTAATCT	58
		EPCAM_SQ_6F	forward	CAGAATTCAGAAAGAGGTTTGAGG	59
		EPCAM_SQ6_F2	forward	GGCTCAAGCAGTTCTTTTGG	55
		EPCAM_SQ6_R	reverse	CAGGCCCAGCCTTTTAACAT	60
		EPCAM_SQ6_R2	reverse	CATGTTGGCCAGGCTGGTCT	65
		EPCAM_SQ7_F	forward	TGACCAACACGGAGAAACCC	62
		EPCAM_SQ7_R	reverse	ATGAGCCAAGATCGCGCCACT	67
		EPCAM_SQ8_F	forward	AATGGAGTCTCGCTGGCCG	65
		EPCAM_SQ8_R	reverse	GAGATACAGTTTCGCTCTTGTTGCCC	65
		EPCAM_SQ9_F	forward	GTAGAGATGGGGTTTTGCC	56
		EPCAM_SQ9_R	reverse	CCTCCCAGGTTCAAGCAATTC	62
		EPCAM_SQ10_F	forward	GCATCCAGGAAGAATCAGGAC	60
	Diagnostic PCR	EPCAM_MP9_R2	reverse	AGAGGGCTAAGATCAGTGGGAAT	55
		EPCAM_MP6_F3	forward	GCATTGCTTGTAAAGGTGAATCTG	
		EPCAM_MP6_F4	forward	TGGGTTCTATCCTGCCTTTG	
HAPLOTYPE	Microsatellite analysis	D2S2331_F	forward	PET-ATTAGCACTTACCTGGCACA	54
		D2S2331_R	reverse	AGTTTATGCTGTGATTAATACCTGG	
		D2S2328_F	forward	NED-CCGAGCAATTTCACTCTGG	53
		D2S2328_R	reverse	CTTTGGCAAACGAGCG	
		D2S119_F	forward	HEX-CTTGGGGAACAGAGGTCATT	54
		D2S119_R	reverse	GAGAATCCCTCAATTTCTTTGGA	
		D2S288_F	forward	FAM-GTCTCAGGGAATAGGAAGGA	52
		D2S288_R	reverse	ACTGGGCCAGTGATTGTTC	
		D2S1248_F	forward	PET-GCTCCCATACTCTCACTTG	52
		D2S1248_R	reverse	PETTTAATACCATCCTCAGTAACC	
		D2S123_F	forward	FAM-TCAACATTGCTGGAAGTTCT	51
		D2S123_R	reverse	FAM-GACTTTCCACCTATGGGACT	
		D2S378_F	forward	HEX-TTTCCTTTGGAGGAAGAGTG	51
		D2S378_R	reverse	HEX-TTATTACACTGCTTACCAGCG	
		D2S2957_F	forward	NED-TGGAATGGGAGAAGGTAACA	53
		D2S2957_R	reverse	NED-CTGTCTCTACACATCCTGTTGG	
	SNP analysis	MSH2_Exon1_F	forward	TCGCGCATTTTCTTCAACCA	67
		MSH2_Exon1_R	reverse	GTCCCTCCCCAGCACG	
		MSH2_Exon10_F	forward	GTTGCCCAGTGCTGGTGCTC	60
		MSH2_Exon10_R	reverse	TGTTGCGACAGCTGACTGCTC	
		MSH2_Exon13_F	forward	CGCGATTAATCATCAGTGT	57
		MSH2_Exon13_R	reverse	CACAGGACAGAGACATACATT	

Supplementary Figure S1. Alignment of the deleted *EPCAM* alleles with *EPCAM* intron 7 and 3'UTR sequences. The red square indicates the microhomology at the breakpoint region. AluSp and AluSx sequences are highlighted in light and dark grey, respectively. A. Alignment of the *EPCAM* g.77526_86198del8673 allele. B. Alignment of the *EPCAM* g. 77446_89071del 11626 allele.



Supplementary Figure S2. Result of the multiplex PCR-based diagnostic test of the *EPCAM* g.77526_86198del8673 in a carrier and a non-carrier individual. Primers *b* and *c* amplify a 686-bp fragment specific to the wildtype allele, and primers *a* and *c* amplify a 488-bp fragment specific to the deleted allele (location of primers is shown in Figure 2). Marker lane represents the 100-bp DNA ladder.



TELOMERE LENGTH IN LYNCH SYNDROME: DYNAMICS AND ASSOCIATION WITH GENETIC ANTICIPATION

Nuria Seguí, Marta Pineda, Elisabet Guinó, **Ester Borràs**, Matilde Navarro, Fernando Bellido, Victor Moreno, Conxi Lázaro, Ignacio Blanco, Gabriel Capellá, Laura Valle.

Manuscrit sotmès a publicació

Resum del treball: La variació en la longitud dels telòmers s'ha associat amb un major risc de malignitat, encara que els estudis en el càncer colorectal són controvertits. A més, l'escurçament dels telòmers s'ha associat amb l'anticipació genètica en un nombre de malalties genètiques, incloent varis síndromes de càncer hereditari. No obstant això, no hi ha estudis concloents realitzats la síndrome de Lynch. Aquesta síndrome està causada per mutacions germinals en els gens de reparació del DNA. El nostre objectiu és estudiar la longitud dels telòmers en la síndrome de Lynch i avaluar la seva associació amb l'anticipació genètica. La longitud dels telòmers de leucòcits es va mesurar en 244 portadors de la mutació en els gens reparadors en 96 famílies amb síndrome de Lynch i en 234 controls utilitzant un mètode de PCR múltiplex quantitativa. Els individus portadors de mutació afectes de càncer mostraven telòmers significativament més curts que els portadors de la mutació sense càncer. A més, els individus portadors afectats de càncer, i en particular aquells amb una mutació *MSH2*, van mostrar la disminució més pronunciada en la longitud dels telòmers en relació amb l'edat. L'estudi de l'anticipació en parelles de pares i fills, es va dur a terme en 59 famílies, i va mostrar els fills desenvolupaven el càncer una mitjana de 10-12 anys abans que els seus pares. No obstant això, l'anticipació observada no es va associar amb escurçament dels telòmers. En conclusió, els nostres resultats demostren que els individus portadors de mutacions en els gens reparadors i afectes de càncer mostren diferents patrons i dinàmiques en la longitud dels telòmers. A més, l'anticipació en l'edat del diagnòstic observada en la síndrome de Lynch no es pot explicar per la longitud telomèrica

Telomere Length in Lynch Syndrome: Role as Cancer Risk Factor and Association with Genetic Anticipation

N. Seguí[1], M. Pineda[1], E. Guinó[2], E. Borràs[1], M. Navarro[1], F. Bellido[1], V. Moreno[2,3], C. Lázaro[1], I. Blanco[1], G. Capellá[1] and L. Valle[1]

[1]Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Barcelona, Spain; [2]Unit of Biomarkers and Susceptibility, Catalan Institute of Oncology, IDIBELL and CIBERESP, Barcelona, Spain; [3]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Spain.

Short title: Telomere Length in Lynch Syndrome

ABSTRACT

Telomere length variation has been associated with increased risk of several types of tumors, and telomere shortening, with genetic anticipation in a number of genetic diseases including hereditary cancer syndromes. No conclusive studies have been performed for Lynch syndrome (LS), a hereditary colorectal cancer syndrome caused by germline mutations in the DNA mismatch repair (*MMR*) genes. Here we evaluate telomere length in LS, both as a cancer risk factor and as the mechanism involved in the anticipation in the age of cancer onset observed in successive generations of LS families. Leukocyte telomere length was measured in 244 *MMR* gene mutation carriers from 96 LS families and in 234 controls using a monochrome multiplex quantitative PCR method. Cancer-affected mutation carriers showed significantly shorter telomeres than cancer-free mutation carriers. Also, cancer-affected carriers showed the most pronounced decline of telomere length with age, compared with unaffected carriers. The study of age anticipation in parent-child pairs from 59 families showed that children developed cancer at an average of 12 years earlier than their parents. However, the observed anticipation was not associated with telomere shortening. Nevertheless, our findings suggest that telomere attrition might explain the dependence of cancer risk on the parent-of-origin of *MMR* gene mutations previously reported. In conclusion, cancer-affected *MMR* gene mutation carriers have distinct telomere-length pattern and dynamics, suggesting a role as cancer risk factor for LS. However, the anticipation in the age of onset observed in LS is not explained by telomere shortening.

AUTHOR SUMMARY

Hereditary colorectal cancer (CRC) constitutes about 5% of all diagnosed CRC cases. Among CRC hereditary syndromes, Lynch syndrome (LS) is the most common one, predisposing families to colorectal, endometrial and other extracolonic tumors. The cause of the increased cancer risk in these families is the presence of a mutation in one gene involved in the repair of DNA errors: *MLH1*, *MSH2*, *MSH6* or *PMS2*. Nevertheless, there is large variation in disease expression, e.g. in the age of cancer onset, among and within families, where the same gene mutation is present, suggesting the presence of risk modifying factors. This variability is also represented by the anticipation in the age of cancer onset that occurs in successive generations. Telomeres are repetitive DNA sequences at the end of chromosomes that protect them, guaranteeing chromosome stability and integrity. Telomere length alterations have been proposed as cancer risk
factors and telomere shortening explains genetic anticipation in several diseases including some hereditary cancer disorders. Here we find that shortened telomeres may act as a cancer risk factor in LS, and we discard the association of telomere shortening with the anticipation in the age of cancer onset across generations observed in LS families.

INTRODUCTION

Lynch syndrome (LS) (MIM 120435) is the most common hereditary syndrome that predisposes to colorectal cancer (CRC) and other extracolonic tumors, accounting for 2%–5% of the total burden of CRC. It is caused by germline mutations in the DNA mismatch repair (*MMR*) genes *MLH1* (MIM 120436), *MSH2* (MIM 609309), *MSH6* (MIM 600678) and *PMS2* (MIM 600259), as well as in *EpCAM* (MIM 185535). There is large phenotypic variation in the age of onset among and within families [1, 2]. Moreover, progressively earlier age of cancer onset in successive generations has been reported [3-10]. However, the underlying molecular mechanisms of both the overall phenotypic variation and the anticipation in the age of onset are still unknown.

Telomere shortening has gained considerable interest as a potential biomarker of cancer risk and also as a mechanism associated with

genetic anticipation. Moreover, we have recently shown that a common genetic variant located in the telomerase gene (*hTERT*; MIM 187270) increases cancer risk in LS patients at early ages, and that this SNP is associated with shortened telomere length in cancer-affected *MMR* gene mutation carriers [11]. Telomeres consist of multiple short repeats (TTTAGG) at the end of chromosomes and protect them against large-scale genomic rearrangements. In most somatic cells, telomeres shorten with each replication cycle, eventually leading to cellular senescence or apoptosis. Interestingly, telomere length anomaly appears to be one of the earliest and most prevalent genetic alterations in the process of malignant transformation [12, 13]. Given the proposed role of telomere shortening in early processes of carcinogenesis, attention has been directed to its potential role as a cancer risk biomarker. With this purpose, a number of retrospective and prospective observational studies have been conducted, although results are inconsistent among studies and tumor types [14]. Regarding CRC, despite the conflicting results found in retrospective studies, the prospective ones have found no convincing association between leukocyte telomere length and CRC risk [14-16].

Also, telomere length attrition has been proposed as a mechanism of anticipation in different inherited diseases, being associated with early onset and/or severity of disease in genetic disorders such as dyskeratosis congenita [17-19], Li-Fraumeni [20, 21] and familial breast and ovarian cancer [22, 23].

Here we aim to elucidate the role of telomere length in LS-associated cancer risk and to study its implication in the anticipation in the age of onset observed in successive generations of LS families.

RESULTS

Telomere length and shortening dynamics

Relative telomere length (RTL) was assessed in 244 *MMR* gene mutation carriers, both cancer-affected (n=144) and unaffected (n=100), and in 234 controls, including non-carriers of the same LS families (n=144) and unrelated cancer-free controls (n=90) (Figure 1). We found that among mutation carriers, cancer cases had significantly shorter age-adjusted telomeres than unaffected individuals (p=0.032). Although statistically non-significant, worthy of note is the fact that unaffected mutation carriers tend to have longer telomeres than unaffected controls (p=0.069, including non-carriers and cancer-free controls within the control group).

The negative slope of the linear regression line that best fit the relative telomere length data for the controls, as expected, indicated that telomere length shortened with age (Pearson's correlation coefficient r=-0.187; p=0.004), as did for *MMR* gene mutation carriers (r=-0.297; $p=2.37 \times 10^{-6}$). The differences observed between both groups were not statistically significant (p=0.204) (Figure 2A). Among *MMR* gene mutation carriers, cancer-affected individuals showed faster telomere attrition with age (r=-0.344; $p=2.4 \times 10^{-5}$), than cancer-free carriers (r=-0.094; p=0.351) (for the difference, p=0.045) (Figure 2B). No statistically significant differences were observed between *MSH2* (r=-0.446; p=0.003) and *MLH1* (r=-0.318; p=0.003) mutation carriers (for the difference, p=0.435).

Telomere length and anticipation in the age of cancer onset

The occurrence of age anticipation in cancer-affected mutation carriers was analyzed in 59 LS families, classified as *MLH1* (32 families), *MSH2* (20 families) and *MSH6* (7 families). The distribution of age at cancer diagnosis for parents and children showed a consistent shift to early ages in children ($p=2.5 \times 10^{-7}$) (Figure 3). In all, cancer was diagnosed at an average of 12 years earlier in children (Table 2). Five out of 65 parents and 3 out of 30 children had been diagnosed with cancer (1^{st} LS-related cancer diagnosed) during the follow-up period or clinical surveillance of mutation carriers. Their exclusion from the analysis did not modify the results obtained (data not shown). The information available from polyp removal was incomplete and was not included in the study.

The relationship between telomere length and anticipation in Lynch syndrome was next investigated. Differences between parents and children resulted non-significant in cancer-affected *MMR* gene mutation carriers (p=0.867) (Figure 4). Looking into the association between telomere length and anticipation in more detail, we compared changes in telomere length in individual parent-child pairs (Table 3). In all instances, the child was diagnosed with cancer at earlier ages than his/her corresponding LS parent. Children showed shorter telomeres than their parents in 10 out of 21 (47.6%) cancer-affected carrier pairs, in 6 out of 9 (66.7%) cancer-free carrier pairs, and in 1 out of 3 (33.3%) cancer-free non-carriers evaluated. The results obtained showed no statistically significant differences among groups (p=0.562). Cancer-affected and unaffected individuals were not combined (e.g. affected parent vs. unaffected child) because, as shown in Figure 1, cancer affection itself affects telomere length measured in blood cells.

DISCUSSION

In this study we found that cancer-affected *MMR* gene mutation carriers have shorter telomeres and faster telomere attrition rate (accelerated telomere shortening with age) than unaffected mutation carriers and controls, and that telomere shortening is not associated with anticipation in the age of cancer onset in successive generations.

After evaluating the reported retrospective and prospective studies carried out on sporadic CRC risk and leukocyte telomere length, it has been recently suggested that the shorter telomere length observed in retrospectively collected samples from CRC patients is not a cancer risk factor, but arises as a result of disease (from treatment or disease burden) [14, 15]. However, a recent prospective study has shown that both extremely long and short telomeres increase CRC risk [16]. Our observation in hereditary cases, where shorter telomeres are identified in cancer-affected compared with cancer-free *MMR* gene mutation carriers, suggests either that the shortened telomere length is a result of the pathophysiological effects of the disease, as suggested for sporadic cases, or that shorter telomeres is an additional risk factor for LS patients, i.e., a modifier of risk. The role of shortened telomeres as cancer risk factor in LS is supported by our recent observation that a variant in the telomerase gene, *hTERT*, increases cancer risk in young LS patients (<45 years of age), being the this variant associated with shortened telomeres in cancer-affected *MMR* gene mutation carriers [11]. Further studies assessing telomere length before and after cancer diagnosis in CRC patients are required to provide a definitive answer to the suggested effect of cancer affection in the telomere length of blood cells.

Interestingly, cancer-free *MMR* gene mutation carriers showed a tendency to have longer telomeres than at-population-risk individuals ($p=0.07$). Jones et al. [15] proposed that risk genetic factors, in particular a variant in the telomerase RNA component, *TERC* (MIM 602322), cause longer telomeres. It assumes that this is mirrored in colorectal crypt stem cells and proposes that CRC might be influenced by longer telomeres, possibly through effects on stem cell numbers or longevity. If the tendency we observe in our LS series reaches statistical significance in larger samples, this model might be adapted to LS and therefore, further studies would be needed to explain the mechanism how *MMR* haploinsufficiency in the germline can induce a lengthening effect in telomeres. Another plausible explanation to this observation, and in line with the fact that short telomeres act as factors of increased cancer risk, is that longer telomeres in *MMR* gene mutation carriers protect them from developing cancer.

We also observed that leukocyte telomere dynamics with age is different in cancer-free mutation carriers than in cancer-affected carriers, who showed faster telomere shortening with age. Bozzao et al., studying a smaller LS sample (43 cancer-affected carriers, 31 unaffected carriers and 50 controls), found that affected carriers showed accelerated attrition rate compared to controls [24]. Based on previously published data on MMR-deficiency and telomeres [25-29], they suggested that MMR haploinsuffiency in LS individuals may lead somehow to tolerance to short telomeres causing faster telomere shortening in the course of life. However, according to our results, this increased shortening rate does not occur in cancer-free carriers, invalidating their hypothesis. Therefore, this effect may be either the result of a telomere shortening occurring in cancer-affected individuals, or a marker of increased cancer risk among *MMR* gene mutation carriers, supporting the idea of short telomere length or accelerated telomere attrition rate as a factor of increased cancer risk in LS. Knowing the dynamics of leukocyte telomere length with age in sporadic cases from retrospective and prospective studies may solve this question, but to our knowledge, this description has not been included in previous publications.

We next investigated whether the differences observed in telomere length in LS somehow translated in the anticipation in the age of onset observed in successive generations, but found no association. For years and despite the numerous reports identifying anticipation in LS, it was uncertain whether true genetic anticipation contributed to the young diagnosis observed in LS. Recently, a Bayesian method correcting for random effects, isolating the confounding effect of changes in secular trends, screening, and medical practices, and adjusting for changes in age-specific incidence across birth cohorts, confirmed the presence of anticipation in the age of onset between successive generations in LS families [10]. Nevertheless, so far the molecular mechanism underlying this observation has not been identified. Here, for the first time, we discard the role of telomere shortening in anticipation in LS. As previously suggested, other molecular mechanisms such as the accumulation of mismatch repair slippage events through generations or other genetic or environmental factors might explain age anticipation in successive generations in LS families [30, 31].

Van Vliet et al. recently suggested that CRC risks for carriers of mutations in *MMR* genes depend on the parent-of-origin of their mutation.

In particular, they suggested a maternally transmitted mechanism modifying cancer risk for male *MMR* gene mutation carriers [32]. Here we observe anticipation associated with shorter telomeres in all (5/5) mother-son cancer-affected *MMR* gene mutation carriers, compared to 33% (4/15) in the other combinations altogether (Table 3). Although the sample size of the parent-child pairs studied is very limited, this observation might suggest that the maternally transmitted risk mechanism is shortened telomeres. Studies assessing the whether telomere shortening has a parental control in humans, in this instance whether telomere shortening occurs in the female germline, as recently demonstrated in other mammals [33], together with the analysis of larger samples of LS parent-child pairs, are granted to verify this hypothesis.

In conclusion, our findings indicate that cancer-affected *MMR* gene mutation carriers show distinct features and dynamics in telomere length measured in blood DNA than controls and unaffected mutation carriers. Pending of additional studies, our observations suggest that telomere length may be a risk factor for Lynch syndrome and that it may explain the dependence of cancer risk on the parent-of-origin of *MMR* gene mutations previously observed. We discard telomere length attrition as the common cause of the anticipation in the age of cancer onset observed across generations in LS.

MATERIAL AND METHODS

Ethics statement

This study received the approval of the Ethics Committee of IDIBELL (ref. PR221/09).

Study participants

A total of 388 individuals, 244 *MMR* gene mutation carriers and 144 non-carriers, from 96 Lynch syndrome families were included in the study. They were assessed through the Hereditary Cancer Program of the Catalan Institute of Oncology, ICO-IDIBELL, from 1998 to 2012. *MMR* gene mutation analysis was performed on genomic DNA extracted from peripheral blood. Large genomic alterations in the genes were studied using MLPA (Multiplex Ligation-dependent Probe Amplification) (SALSA MLPA Kits, MRC-Holland, The Netherlands). Mutation screening was carried out by direct sequencing after PCR amplification (primers and conditions available upon request). Ninety cancer-free individuals were included as controls. All individuals are of Caucasian origin. Informed consent was obtained from all subjects, even though the data were analyzed anonymously. Table 1 shows a summary of the genetic and clinical characteristics of the groups studied.

Cases and controls were recruited from the same homogeneous population, and storage and DNA extraction from peripheral blood were performed at the same facility and using the same extraction methods. We have observed that blood DNA samples coming from different sources (extracted at different laboratories, using different extraction methods, etc.) show large non-genetic variation in telomere length (data not shown). Thus, only samples (cases and controls) that underwent the exact same DNA extraction protocol (Flexigene DNA kit, Qiagen, Hilden, Germany) performed by the personnel of the Unit of Molecular Diagnostics of the Hereditary Cancer Program at ICO-IDIBELL, 15/11/2012conditions, were included in the study, as recommended by Prescott et al. for retrospective studies [14].

Telomere length assessment

Telomere length quantification was performed using the monochrome multiplex quantitative PCR method previously described by Cawthon et al. [34]. It has been found to provide greater consistency than other methods used to measure telomere length [15]. The assays were performed using the Quantifast SybrGreen PCR Master Mix (Qiagen, Hamburg, Germany), beta-globin as single copy gene, and the LightCycler 480 real-time PCR detection system (Roche Diagnostics GmbH, Mannheim, Germany). All samples were assayed in triplicate. A standard curve with 7 concentrations spanning an 81-fold rage (60ng, 40ng, 20ng, 6.7ng, 4ng, 2.2ng and 0.74ng), also in triplicate, of an anonymous standard DNA (healthy and cancer-free 45 year-old individual) was included in every 384-well plate. Good replicates of the standard curves (for telomeres (T) and single copy gene (S)) and subsequent fitted linear regression lines were obtained among plates. Whenever possible, equal numbers of samples from different clinical groups were run in each 384-well plate. The relative telomere length (RTL) value for each sample was calculated according to the standard lines (for T and S) of the corresponding plate. Therefore, the value obtained was relative to the value of the anonymous standard DNA in the same plate (inter-run calibration). The average telomere length of each sample is expected to be proportional to its RTL value.

Statistical analyses

Shortening of telomere length is observed with increasing age in cancer-free individuals [35]. In our control set (controls plus non-carriers, n=234), relative telomere length (RTL) was inversely correlated with age (Pearson's

correlation r=-0.187; p=0.004). Therefore, in order to assess telomere length differences, RTL measurements were adjusted for age using the line of best fit for controls, as described previously [23]. Thus, the difference between the observed and the predicted value was calculated for each sample. There were no telomere length differences between sexes (data not shown), therefore, sex was not included as a confounding factor in the statistical tests.

Differences in age-adjusted telomere lengths were analyzed using the Wilcoxon rank sum test (Mann Whitney U). Pearson's test was used to assess the correlation between telomere length and age (telomere length dynamics). To measure the significance of the difference between two correlation's coefficients (Pearson's r), a Fisher r-to-z transformation was performed. Anticipation was represented by Kaplan-Meier curves and the differences between survival curves were studied using a log-rank test. Proportions were compared by Fisher's exact test (expected cell count <5). All tests were two-sided and p-values below 0.05 were considered statistically significant. The analyses were performed using R.

ACKNOWLEDGEMENTS

We thank all the people responsible for genetic counselling and genetic testing of the Hereditary Cancer Program ICO, IDIBELL.

REFERENCES

1. Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, et al. (1993) Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology 104: 1535-1549.
2. Scott RJ, McPhillips M, Meldrum CJ, Fitzgerald PE, Adams K, et al. (2001) Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. Am J Hum Genet 68: 118-127.
3. Vasen HF, Taal BG, Griffioen G, Nagengast FM, Cats A, et al. (1994) Clinical heterogeneity of familial colorectal cancer and its influence on screening protocols. Gut 35: 1262-1266.
4. Rodriguez-Bigas MA, Lee PH, O'Malley L, Weber TK, Suh O, et al. (1996) Establishment of a hereditary nonpolyposis colorectal cancer registry. Dis Colon Rectum 39: 649-653.
5. Westphalen AA, Russell AM, Buser M, Berthod CR, Hutter P, et al. (2005) Evidence for genetic anticipation in hereditary non-polyposis colorectal cancer. Hum Genet 116: 461-465.
6. Stella A, Surdo NC, Lastella P, Barana D, Oliani C, et al. (2007) Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. Clin Genet 71: 130-139.
7. Nilbert M, Timshel S, Bernstein I, Larsen K (2009) Role for genetic anticipation in Lynch syndrome. J Clin Oncol 27: 360-364.
8. Larsen K, Petersen J, Bernstein I, Nilbert M (2009) A parametric model for analyzing anticipation in genetically predisposed families. Stat Appl Genet Mol Biol 8: Article26.
9. Boonstra PS, Gruber SB, Raymond VM, Huang SC, Timshel S, et al. (2010) A review of statistical methods for testing genetic anticipation: looking for an answer in Lynch syndrome. Genet Epidemiol 34: 756-768.
10. Boonstra PS, Mukherjee B, Taylor JM, Nilbert M, Moreno V, et al. (2011) Bayesian modeling for genetic anticipation in presence of mutational heterogeneity: a case study in Lynch syndrome. Biometrics 67: 1627-1637.
11. Bellido F, Guino E, Jagmohan-Changur S, Segui N, Pineda M, et al. (2012) Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome. Eur J Hum Genet.
12. Londono-Vallejo JA (2004) Telomere length heterogeneity and chromosome instability. Cancer Lett 212: 135-144.
13. Meeker AK, Hicks JL, Iacobuzio-Donahue CA, Montgomery EA, Westra WH, et al. (2004) Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. Clin Cancer Res 10: 3317-3326.
14. Prescott J, Wentzensen IM, Savage SA, De Vivo I (2012) Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. Mutat Res 730: 75-84.
15. Jones AM, Beggs AD, Carvajal-Carmona L, Farrington S, Tenesa A, et al. (2012) TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. Gut 61: 248-254.
16. Cui Y, Cai Q, Qu S, Chow WH, Wen W, et al. (2012) Association of Leukocyte Telomere Length with Colorectal Cancer Risk: Nested Case-Control Findings from the Shanghai Women's Health Study. Cancer Epidemiol Biomarkers Prev 21: 1807-1813.
17. Vulliamy T, Marrone A, Szydlo R, Walne A, Mason PJ, et al. (2004) Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. Nat Genet 36: 447-449.
18. Marrone A, Walne A, Dokal I (2005) Dyskeratosis congenita: telomerase, telomeres and anticipation. Curr Opin Genet Dev 15: 249-257.
19. Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, et al. (2005) Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita.

Proc Natl Acad Sci U S A 102: 15960-15964.
20. Tabori U, Nanda S, Druker H, Lees J, Malkin D (2007) Younger age of cancer initiation is associated with shorter telomere length in Li-Fraumeni syndrome. Cancer Res 67: 1415-1418.
21. Trkova M, Prochazkova K, Krutilkova V, Sumerauer D, Sedlacek Z (2007) Telomere length in peripheral blood cells of germline TP53 mutation carriers is shorter than that of normal individuals of corresponding age. Cancer 110: 694-702.
22. Gramatges MM, Telli ML, Balise R, Ford JM (2010) Longer relative telomere length in blood from women with sporadic and familial breast cancer compared with healthy controls. Cancer Epidemiol Biomarkers Prev 19: 605-613.
23. Martinez-Delgado B, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M, et al. (2011) Genetic anticipation is associated with telomere shortening in hereditary breast cancer. PLoS Genet 7: e1002182.
24. Bozzao C, Lastella P, Ponz de Leon M, Pedroni M, Di Gregorio C, et al. (2011) Analysis of telomere dynamics in peripheral blood cells from patients with Lynch syndrome. Cancer 117: 4325-4335.
25. Pickett HA, Baird DM, Hoff-Olsen P, Meling GI, Rognum TO, et al. (2004) Telomere instability detected in sporadic colon cancers, some showing mutations in a mismatch repair gene. Oncogene 23: 3434-3443.
26. Mendez-Bermudez A, Hills M, Pickett HA, Phan AT, Mergny JL, et al. (2009) Human telomeres that contain (CTAGGG)n repeats show replication dependent instability in somatic cells and the male germline. Nucleic Acids Res 37: 6225-6238.
27. Martinez P, Siegl-Cachedenier I, Flores JM, Blasco MA (2009) MSH2 deficiency abolishes the anticancer and pro-aging activity of short telomeres. Aging Cell 8: 2-17.
28. Rampazzo E, Bertorelle R, Serra L, Terrin L, Candiotto C, et al. (2010) Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers. Br J Cancer 102: 1300-1305.
29. Mendez-Bermudez A, Royle NJ (2011) Deficiency in DNA mismatch repair increases the rate of telomere shortening in normal human cells. Hum Mutat 32: 939-946.
30. Gruber SB, Mukherjee B (2009) Anticipation in lynch syndrome: still waiting for the answer. J Clin Oncol 27: 326-327.
31. Coolbaugh-Murphy MI, Xu JP, Ramagli LS, Ramagli BC, Brown BW, et al. (2010) Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. Hum Mutat 31: 317-324.
32. Van Vliet CM, Dowty JG, van Vliet JL, Smith L, Mead LJ, et al. (2011) Dependence of colorectal cancer risk on the parent-of-origin of mutations in DNA mismatch repair genes. Hum Mutat 32: 207-212.
33. Bender HS, Murchison EP, Pickett HA, Deakin JE, Strong MA, et al. (2012) Extreme telomere length dimorphism in the tasmanian devil and related marsupials suggests parental control of telomere length. PLoS One 7: e46195.
34. Cawthon RM (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res 37: e21.
35. Iwama H, Ohyashiki K, Ohyashiki JH, Hayashi S, Yahata N, et al. (1998) Telomeric length and telomerase activity vary with age in peripheral blood cells obtained from normal individuals. Hum Genet 102: 397-402.

FIGURES

Figure 1. Age-adjusted relative telomere length in subjects belonging to LS families: *MMR* gene mutation carriers affected with cancer (median: -0.131), *MMR* gene mutation carriers with no diagnosed cancer (median: -0.079) and cancer-free non-carriers (median: -0.091); and to cancer-free controls (median: -0.097). The boxes represent the interquartile range of distributions (25th and 75th percentile); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles.

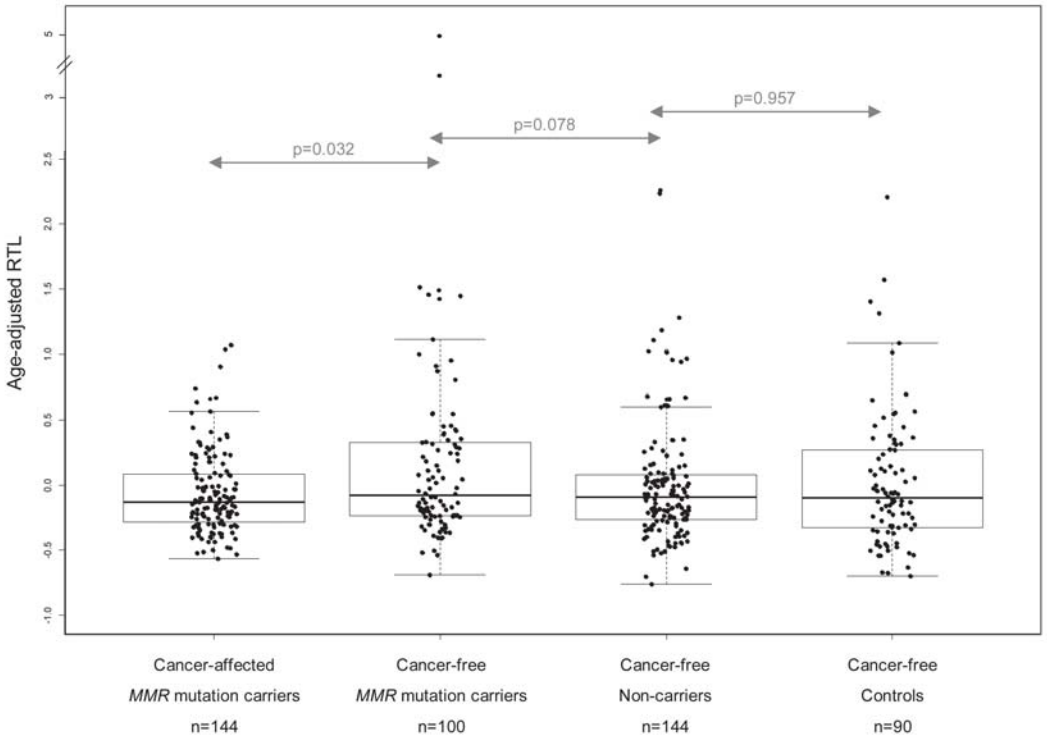


Figure 2. Correlation of relative telomere length (RTL) with age at blood collection. (A) RTL distribution as a function of age for controls (n=234; white circles) and all *MMR* gene mutation carriers (n=244; black circles). (B) RTL distribution as a function of age for cancer-affected *MMR* gene mutation carriers (n=144; black circles) and cancer-free *MMR* gene mutation carriers (n=100; crossed circles). r: Pearson's correlation coefficient.

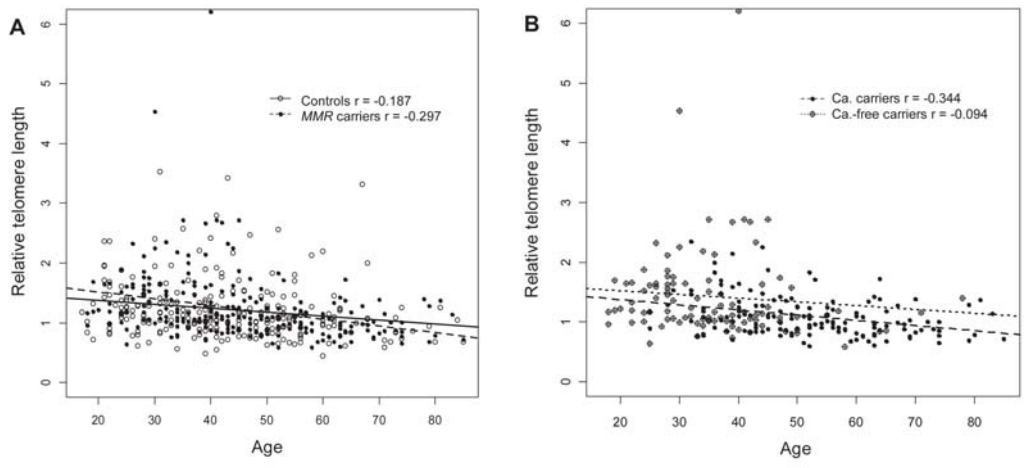


Figure 3. Anticipation in the age of cancer onset in LS families. Kaplan-Meier curves and associated p-values showing the differences in age of cancer onset between parents and children, all of them cancer-affected *MMR* gene mutation carriers.

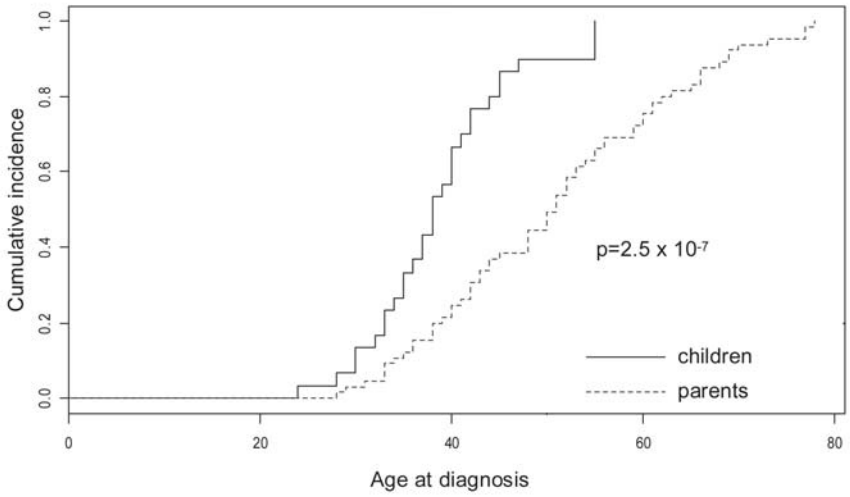
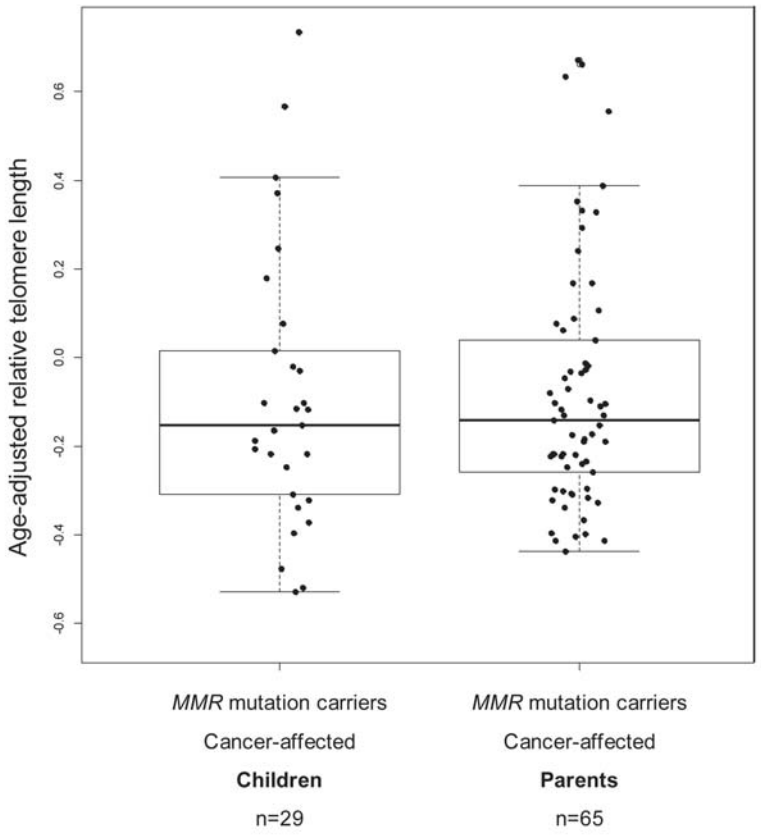


Figure 4. Changes in age-adjusted RTL between parents and children. Box plots representing age-adjusted RTL distributions in cancer-affected parents and children with *MMR* gene mutations. The boxes represent the interquartile range of distributions (25th and 75th percentile); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles.



TABLES

Table 1. Genetic and clinical characteristics of the studied groups.

	Lynch syndrome families (n=96)			**Controls**
	MMR gene mutation carriers		*Non-carriers*	
	Cancer	Cancer-free		
N	144	100	144	90
MMR gene mutated: n (%)	*MLH1*: 87 (60.4) *MSH2*: 42 (29.2) *MSH6*: 9 (6.3) *PMS2*: 5 (3.5) *EpCAM*: 1 (0.7)	*MLH1*: 62 (62.0) *MSH2*: 29 (29.0) *MSH6*: 9 (9.0) *PMS2*: 0 *EpCAM*: 0	-	-
Mean age at blood extraction (± SD)	51.8 (±13.4)	35.7 (±11.5)	43.4 (±14.9)	43.6 (±15.1)
Sex: n (%)	M: 74 (51.4) F: 70 (48.6)	M: 40 (40.0) F: 60 (60.0)	M: 64 (44.4) F: 80 (55.6)	M: 28 (31.1) F: 62 (68.9)
Mean age at cancer diagnosis (± SD)	45.3 (±12.9)	-	-	-

N: number of subjects; SD: standard deviation; M: male; F: female

Table 2. Differences in the average age of onset of cancer (1st LS-related tumor diagnosed) between parents and children in Lynch syndrome.

	Parents		**Children**				
	N	Mean age of onset (±SD)	N	Mean age of onset (±SD)	Mean age difference	t-statistic	p-value
All	65	50.9 (±12.8)	30	38.9 (±7.6)	12.0	4.74	7.5×10^{-6}
MLH1	36	50.9 (±11.7)	20	38.0 (±6.8)	12.9	4.53	3.3×10^{-5}
MSH2	21	51.4 (±15.4)	7	41.4 (±10.4)	10	1.59	0.124
MSH6	8	49.6 (±11.5)	2	37.5 (±6.4)	12.1	1.40	0.200

N: number of subjects; SD: standard deviation.

Table 3. Age-adjusted RTL and anticipation in the age of cancer diagnosis in parent-child pairs.

		PARENT				CHILD				
Family	*MMR* gene	Tumor	Age at diagnosis	Gender	Age-adjusted RTL	Tumor	Age at diagnosis	Gender	Age-adjusted RTL	*Tel_{child} < Tel_{parent}
Cancer-affected carriers										
Fam-1	*MSH2*	CRC	73	F	-0.172	CRC	45	M	-0.216	yes
Fam-2	*MSH2*	EC	78	F	0.167	CRC	55	F	-0.339	yes
Fam-2	*MSH2*	CRC	55	F	-0.339	CRC	33	F	-0.188	no
Fam-2	*MSH2*	EC	78	F	0.167	CRC	55	F	-0.217	yes
Fam-3	*MSH2*	CRC	60	F	-0.098	CRC	35	F	0.407	no
Fam-4	*MLH1*	CRC	53	F	-0.143	CRC	40 (in surv)	F	0.372	no
Fam-4	*MLH1*	CRC	53	F	-0.143	CRC	32	M	-0.521	yes
Fam-5	*MLH1*	CRC	48	M	0.328	CRC	35	M	0.567	no
Fam-6	*MLH1*	CRC	52	F	0.635	CRC	34	F	0.736	no
Fam-7	*MLH1*	CRC	50	F	0.388	CRC	55 (in surv)	F	-0.103	yes
Fam-8	*MLH1*	CRC	51	M	-0.240	CRC	28	F	-0.022	no
Fam-9	*MLH1*	CRC	51	M	0.332	CRC	45	F	0.178	yes
Fam-10	*MLH1*	CRC	33	M	-0.398	CRC	24	F	-0.154	no
Fam-11	*MLH1*	CRC	69 (in surv)	M	-0.298	CRC	37	F	-0.206	no
Fam-11	*MLH1*	CRC	69 (in surv)	M	-0.298	CRC	41 (in surv)	F	-0.247	no
Fam-12	*MLH1*	CRC	66	F	-0.014	CRC	47	M	-0.373	yes
Fam-12	*MLH1*	CRC	66	F	-0.014	CRC	39	M	-0.322	yes
Fam-12	*MLH1*	CRC	66	F	-0.014	CRC	40	M	-0.103	yes
Fam-13	*MLH1*	CRC	60	F	-0.081	CRC	38	F	-0.030	no
Fam-14	*MLH1*	CRC	77	F	-0.235	CRC	38	F	-0.076	no
Fam-15	*MSH6*	UC	69	M	-0.032	CRC	33	M	-0.478	yes
Cancer-free carriers			Age at blood extraction				Age at blood extraction			
Fam-16	*MLH1*	-	48	M	-0,347	-	24	M	-0,327	no
Fam-16	*MLH1*	-	48	M	-0,347	-	25	F	-0,688	yes
Fam-17	*MLH1*	-	47	M	-0,121	-	27	F	-0,387	yes
Fam-18	*MLH1*	-	64	F	-0,244	-	36	M	-0,095	no
Fam-18	*MLH1*	-	64	F	-0,244	-	39	M	-0,134	no
Fam-19	*MLH1*	-	45	F	0,048	-	18	M	-0,202	yes
Fam-20	*MLH1*	-	78	M	0,399	-	37	F	-0,186	yes
Fam-21	*MSH2*	-	47	F	0,552	-	22	M	-0,021	yes
Fam-22	*MSH6*	-	65	F	-0,230	-	30	M	-0,300	yes
***Cancer-free non-carriers*										
Fam-23	-	-	71	F	-0,336	-	45	M	0,000	no
Fam-20	-	-	74	F	-0,284	-	50	M	-0,181	no
Fam-24	-	-	85	F	-0,268	-	53	F	-0,432	yes

MMR: DNA mismatch repair; RTL: relative telomere length; CRC: colorectal cancer; EC: endometrial cancer; UC: urinary tract cancer; M: male; F: female; in surv: cancer diagnosed as a consequence of the extensive clinical surveillance of LS families.

*Tel_{child} < Tel_{parent}, age-adjusted telomere length is shorter in the child than in the parent. "Yes" indicates that age-adjusted RTL is smaller in the child than in the parent, indicating therefore association with anticipation when cancer-affected carriers are studied.

**Cancer-free non-carriers are individuals who belong to Lynch syndrome families but do not carry the pathogenic *MMR* gene mutation.

ANNEX II: Informes dels directors

INFORME DELS DIRECTORS SOBRE EL FACTOR D'IMPACTE DELS ARTICLES PUBLICATS

Com a directors de la tesi doctoral de l'Ester Borràs Flores, titulada "Estudi de variants de significat desconegut en la síndrome de Lynch", fem constar que la doctoranda ha participat activament en el disseny i realització experimental dels treballs que s'inclouen en aquesta tesi, l'anàlisi de resultats, la seva discussió i obtenció de conclusions, i en la preparació dels manuscrits finals. Les contribucions concretes de cada treball s'especifiquen a continuació, conjuntament amb els factors d'impacte dels articles a la data de la seva publicació.

ARTICLES PUBLICATS

ARTICLE 1:

MLH1 Founder Mutations with Moderate Penetrance in Spanish Lynch Syndrome Families

Ester Borràs(*), Marta Pineda(*), Ignacio Blanco, Ethan M. Jewett, Fei Wang, Àlex Teulé, Trinidad Caldés, Miguel Urioste, Cristina Martínez-Bouzas, Joan Brunet, Judith Balmaña, Asunción Torres, Teresa Ramón y Cajal, Judit Sanz, Lucía Pérez-Cabornero, Sergi Castellví-Bel, Ángel Alonso, Ángel Lanas, Sara González, Víctor Moreno, Stephen B. Gruber, Noah A. Rosenberg, Bhramar Mukherjee, Conxi Lázaro, i Gabriel Capellá. (*) Aquests autors han contribuït equitativament en aquest treball.

Cancer Research: 70(19); 7379–91 (2010)

Factor d'impacte (2009 JCR Science Edition): 7.543

Contribució de la doctoranda: Posada a punt de la metodologia i cribratge de les dues mutacions en una sèrie cas-control. Anàlisis de la pèrdua de LOH en el tumor, estudi *in silico* a nivell de RNA i proteïna, cultiu curt de limfòcits dels pacients amb la mutació, disseny i anàlisi dels estudis de processament i estabilitat a nivell de RNA, modificació d'un plàsmid d'expressió en cèl·lules humanes per introduir al mutació a estudiar, transfecció i estudis d'expressió i estabilitat a nivell de proteïna. Realització de l'estudi d'haplotips en els diferents individus portadors, així com dels familiars disponibles i d'una sèrie d'individus control. Anàlisi i interpretació dels resultats obtinguts, així com l'escriptura del primer esborrany de l'article i preparació de la versió final del manuscrit. Marta Pineda, amb qui la doctoranda comparteix primera autoria, va aportar els resultats preliminars de l'estudi d'haplotips, i va participar en l'escriptura i revisió del manuscrit, així com en la supervisió de tot el treball, juntament amb el Dr. Gabriel Capellá.

ARTICLE 2:

Comprehensive Functional Assessment of MLH1 Variants of Unknown Significance

Ester Borràs, Marta Pineda, Angela Brieger, Inga Hinrichsen, Carolina Gómez, Matilde Navarro, Judit Balmaña, Teresa Ramón y Cajal, Asunción Torres, Joan Brunet, Ignacio Blanco, Guido Plotz, Conxi Lázaro i Gabriel Capellá

Human Mutation: 33(11): 1576-88 (2012).

Factor d'impacte (2011 JCR Science Edition): 5.686

Contribució de la doctoranda: Posada a punt de la metodologia i anàlisi de la freqüència de les VSD identificades en una sèrie control, estudi *in silico* a nivell de RNA i proteïna, cultiu curt de limfòcits amb i sense puromicina, disseny i posada a punt de l'anàlisi del processament i l'estabilitat del RNA, modificació d'un plàsmid d'expressió específic per cada variant i la seva posterior transfecció, extracció de proteïna, estudi de l'activitat reparadora, anàlisi d'expressió de les variants i localització subcel·lular. Aquests últims tres punts es van dur a terme a la Universitat Johann Wolfgang Goethe, Frankfurt. Finalment, escriptura del primer esborrany de l'article i preparació de la versió final del manuscrit.

MANUSCRIT EN PREPARACIÓ

ARTICLE 3:

Refining the role of PMS2 in Lynch syndrome: germline mutation alanalysis improved by comprehensive assessment of variants

Ester Borràs, Marta Pineda, Juan Cadiñanos, Jesús del Valle, Angela Brieger, Inga Hinrichsen, Rubén Cabanillas, Matilde Navarro, Joan Brunet, Helen van der Klift, Conxi Lázaro, Guido Plotz, Ignacio Blanco i Gabriel Capellá.

Manuscrit sotmès a publicació

Contribució de la doctoranda: Posada a punt al laboratori de l'estudi mutacional del gen *PMS2* mitjançant long-range PCR (tant a nivell de DNA com de RNA), estudi de la freqüència en una sèrie de controls de totes les VSD detectades, disseny i anàlisi de l'estudi de la fase al·lèlica per dues VSD identificades en un mateix pacient, anàlisi bioinformàtica a nivell de RNA i proteïna, cultiu de limfòcits tractats amb i sense puromicina, extracció de RNA, disseny i anàlisi del processament i estabilitat del RNA, modificació del plàsmid d'expressió per l'estudi de les VSD a nivell de proteïna, transfecció de les VSD en una línia cel·lular estable, extracció de proteïna, estudi de l'activitat reparadora, anàlisi d'expressió de les variants i posada a punt i estudi de la localització subcel·lular. Part dels últims tres punts, es van dur a terme a la Universitat Johann Wolfgang Goethe, Frankfurt. Finalment, preparació i escriptura del manuscrit.

ARTICLES PUBLICATS (ANNEXES)

ARTICLE 4:

Allele-specific expression of APC in Adenomatous polyposis families

Ester Castellsagué, Sara González, Elisabet Guinó, Kristen N. Stevens, **Ester Borràs**, Victoria M. Raymond, Conxi Lázaro, Ignacio Blanco, Stephen B. Gruber Gabriel Capellá.

Gastroenterology: 139 (2): 439-47, 447.e1 (2010)

Factor d'impacte (2009 JCR Science Edition): 12.899

Contribució de la doctoranda: Extracció del RNA, RT-PCR, estudi del cDNA del gen *APC* i mesura de l'expressió específica d'al·lel a partir d'un polimorfisme conegut amb la tècnica del SNuPE, en una sèrie de limfòcits control. Participació en la redacció del manuscrit.

ARTICLE 5:

Characterization of new founderAlu-mediated rearrangements in MSH2 gene associated with a Lynch syndrome phenotype

Lucia Pérez-Cabornero, **Ester Borràs Flores**, Mar Infante Sanz, Eladio Velasco Sampedro, Alberto Acedo Becares, Enrique Lastra Aras, Jorge Cuevas González, Marta Pineda Riu, Teresa Ramón y Cajal Asensio, Gabriel Capellá Munar, Cristina Miner Pino, and Mercedes Durán Domínguez.

Cancer Prevention Research: 4(10):1546-55 (2012).

Factor d'impacte (2011 JCR Science Edition): 4,908

Contribució de la doctoranda: Caracterització del punt de ruptura de la deleció dels exons 4-8 del gen *MSH2* mitjançant LR-PCR i posterior seqüenciació. Disseny d'una PCR múltiplex que permet amplificar específicament l'al·lel mutat i l'al·lel salvatge. Així, com també va participar en la revisió crítica del manuscrit.

ARTICLE 6:

MLH1 methylation screening is effective in identifying epimutation carriers

Marta Pineda(*), Pilar Mur(*), María Dolores Iniesta, **Ester Borràs**, Olga Campos, Gardenia Vargas, Sílvia Iglesias, Anna Fernández, Stephen B Gruber, Conxi Lázaro, Joan Brunet, Matilde Navarro, Ignacio Blanco and Gabriel Capellá.

(*) Aquests autors han contribuït equitativament en aquest treball.

European Journal of Human Genetics: doi: 10.1038/ejhg.2012.136 (2012)

Factor d'impacte (2011 JCR Science Edition): 4.400

Contribució de la doctoranda: Participació en l'anàlisi d'expressió específica d'al·lel i suport en l'estudi de seqüenciació clonal. Participació en l'anàlisi i interpretació dels resultats i revisió crítica del manuscrit.

MANUSCRITS EN PREPARACIÓ (ANNEXES)

ARTICLE 7:

Identification of a founder EPCAM deletion in Spanish Lynch syndrome families

Pilar Mur*, Marta Pineda*, Atocha Romero, Jesús del Valle, **Ester Borràs**, Carla Guarinos, Alicia Canales, Matilde Navarro, Joan Brunet, Daniel Rueda, Teresa Ramón y Cajal, Víctor Moreno, Conxi Lázaro, José Luis Soto, Trini Caldés Ignacio Blanco, Gabriel Capellá.

(*) Aquests autors han contribuït equitativament en aquest treball.

Manuscrit sotmès a publicació

Contribució de la doctoranda: Participació en l'anàlisi d'haplotips en individus control i portadors de la mutació. Diagnòstic predictiu de la mutació fundadora mitjançant PCR múltiplex. Participació en l'anàlisi i interpretació dels resultats, així com en la revisió crítica del manuscrit.

ARTICLE 8:

Telomere length in Lynch syndrome: dynamics and association with genetic anticipation

Nuria Seguí, Marta Pineda, Elisabet Guinó, **Ester Borràs**, Matilde Navarro, Fernando Bellido, Victor Moreno, Conxi Lázaro, Ignacio Blanco, Gabriel Capellá, Laura Valle.

Manuscrit sotmès a publicació

Contribució de la doctoranda: Participació en la selecció de parelles pares-fills. Preparació dels DNAs d'individus control i parelles pares-fills amb síndrome de Lynch. Participació en l'anàlisi i interpretació dels resultats, així com en la revisió crítica del manuscrit.

Atentament,

Dr. Gabriel Capellá Munar

Director Programa de Càncer Hereditari ICO
Laboratori de RecercaTranslacional (LRT)
Institut Català d'Oncologia (ICO)-IDIBELL
gcapella@iconcologia.net

Dra. Marta Pineda Riu

Laboratori de Recerca Translacional (LRT)
Institut Català d'Oncologia (ICO)-IDIBELL
mpineda@iconcologia.net

INFORME DELS DIRECTORS SOBRE COAUTORIA D'UN DELS ARTICLES PUBLICATS

Com a directors de la tesi doctoral de l'Ester Borràs Flores, titulada "Estudi de variants de significat desconegut en la síndrome de Lynch", fem constar que la Dra. Marta Pineda, coautora de l'article *MLH1 Founder Mutations with Moderate Penetrance in Spanish Lynch Syndrome Families* presentat en aquesta tesi, no n'ha utilitzat, ni implícitament ni explícitament, els resultats per a la realització d'una tesi doctoral. Seguidament confirmem la participació de la doctoranda en el treball.

MLH1 Founder Mutations with Moderate Penetrance in Spanish Lynch Syndrome Families

Ester Borràs(*), Marta Pineda(*), Ignacio Blanco, Ethan M. Jewett, Fei Wang, Àlex Teulé, Trinidad Caldés, Miguel Urioste, Cristina Martínez-Bouzas, Joan Brunet, Judith Balmaña, Asunción Torres, Teresa Ramón y Cajal, Judit Sanz, Lucía Pérez-Cabornero, Sergi Castellví-Bel, Ángel Alonso, Ángel Lanas, Sara González, Víctor Moreno, Stephen B. Gruber, Noah A. Rosenberg, Bhramar Mukherjee, Conxi Lázaro, i Gabriel Capellá.*(*) Aquests autors han contribuït equitativament en aquest treball.*

Cancer Research: 70(19); 7379–91 (2010)

Factor d'impacte (2009 JCR Science Edition): 7.543

Contribució de la doctoranda: Posada a punt de la metodologia i cribratge de les dues mutacions en una sèrie cas-control. Anàlisis de la pèrdua de LOH en el tumor, estudi *in silico* a nivell de RNA i proteïna, cultiu curt de limfòcits dels pacients amb la mutació, disseny i anàlisi dels estudis de processament i estabilitat a nivell de RNA, modificació d'un plàsmid d'expressió en cèl·lules humanes per introduir al mutació a estudiar, transfecció i estudis d'expressió i estabilitat a nivell de proteïna. Realització de l'estudi d'haplotips en els diferents individus portadors, així com dels familiars disponibles i d'una sèrie d'individus control. Escriptura del primer esborrany de l'article i preparació de la versió final del manuscrit. Marta Pineda, amb qui la doctoranda comparteix primera autoria, va aportar els resultats preliminars de l'estudi d'haplotips,i va participar en l'escriptura i revisió del manuscrit, així com en la supervisió de tot el treball, juntament amb el Dr. Gabriel Capellá.

Atentament,

Dr. Gabriel Capellá Munar

Director Programa de Càncer Hereditari ICO
Laboratori de RecercaTranslacional (LRT)
Institut Català d'Oncologia (ICO)-IDIBELL
gcapella@iconcologia.net

Dra. Marta Pineda Riu

Laboratori de Recerca Translacional (LRT)
Institut Català d'Oncologia (ICO)-IDIBELL
mpineda@iconcologia.net