UNIVERSITAT DE BARCELONA

Desarrollo de nuevas aproximaciones para el diagnóstico molecular de los síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema de reparación de apareamientos erróneos

María Isabel González Acosta

DESARROLLO DE NUEVAS APROXIMACIONES PARA EL DIAGNÓSTICO MOLECULAR DE LOS SÍNDROMES DE PREDISPOSICIÓN HEREDITARIA AL CÁNCER ASOCIADOS A DEFICIENCIA DEL SISTEMA DE REPARACIÓN DE APAREAMIENTOS ERRÓNEOS

María Isabel González Acosta

Barcelona, 2019

Ilustración de cubierta:

Diseñada por *pikisuperstar / Freepik*

Adaptada por *Àngel Sebastián Esparza*

DESARROLLO DE NUEVAS APROXIMACIONES PARA EL DIAGNÓSTICO MOLECULAR DE LOS SÍNDROMES DE PREDISPOSICIÓN HEREDITARIA AL CÁNCER ASOCIADOS A DEFICIENCIA DEL SISTEMA DE REPARACIÓN DE APAREAMIENTOS ERRÓNEOS

Memoria presentada por **María Isabel González Acosta** para optar al grado de Doctora por la Universidad de Barcelona

María Isabel González Acosta

Realizada en el Instituto Catalán de Oncología
del Instituto de Investigación Biomédica de Bellvitge (**ICO-IDIBELL**)
Codirigida por:

Dr. Gabriel Capellá Munar y **Dra. Marta Pineda Riu**

Tesis adscrita a la Facultad de Biología, Universidad de Barcelona (**UB**)
Programa de Doctorado en Genética
Bajo la tutoría de:

Dr. Daniel Grinberg Vaisman

Barcelona, 2019

A mi familia,
a mis amigos,
a Àngel.

Y a todos los nervios que he pasado.

RESUMEN

La función principal del sistema de reparación de apareamientos erróneos (*MisMatch Repair*, MMR) es corregir los errores principalmente introducidos por las DNA polimerasas durante la replicación del genoma. Las mutaciones germinales deletéreas en alguno de los cuatro genes principales del sistema MMR (*MLH1*, *MSH2*, *MSH6* y *PMS2*) son las responsables de los síndromes de predisposición hereditaria a cáncer asociados a deficiencia de este sistema: el síndrome de Lynch (SL), causado por mutaciones monoalélicas en estos genes, y el síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos (CMMRD), causado por mutaciones bialélicas. A consecuencia de la deficiencia de este tipo de reparación, los tumores asociados a estos dos síndromes exhiben pérdida de expresión de las proteínas MMR y/o inestabilidad de microsatélites (MSI) y, aunque en menor grado, estas características también se observan en tejido normal.

El diagnóstico de estos síndromes se basa en la identificación de mutaciones patogénicas en los genes MMR en línea germinal. Sin embargo, el diagnóstico no siempre es posible. La presencia de mutaciones crípticas, la identificación de variantes de significado desconocido (VUS) (que representan el 30% de las variantes que se encuentran en la rutina de diagnóstico) y la existencia de fenotipos intermedios o solapantes con otros síndromes, dificultan el diagnóstico, lo que impacta en el manejo clínico del paciente y sus familiares.

El objetivo de esta tesis doctoral es mejorar el diagnóstico de los síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema MMR. Con este objetivo global nos hemos planteado dos objetivos específicos. El primer objetivo es mejorar la evaluación de la patogenicidad de las variantes aplicando modelos multifactoriales, que integran múltiples líneas de evidencias tanto cualitativas como cuantitativas. También se ha estandarizado el ensayo *in vitro* de actividad reparadora, dirigido a testar la función más importante de una proteína MMR, con el fin de que pueda ser utilizado en la determinación de la patogenicidad de una VUS. El segundo objetivo es desarrollar una nueva metodología,

basada en la detección con alta sensibilidad de la MSI en los tejidos normales de los portadores de mutaciones MMR, para poder diagnosticar de estos síndromes a pesar de no encontrar mutación o de la presencia de VUS.

En relación al primer objetivo, se han reclasificado a patogénicas o benignas el 89% de las variantes estudiadas en esta tesis doctoral gracias a la integración del cálculo multifactorial de probabilidad, la frecuencia poblacional, las predicciones *in silico* y los ensayos funcionales a nivel de RNA y proteína en un nuevo algoritmo de clasificación, lo que apoya su utilidad. Además, el ensayo *in vitro* de actividad reparadora ha sido optimizado a nivel de reactivos y validado a nivel analítico demostrando robustez y reproducibilidad. A destacar, se han establecido protocolos estándar para su realización, lo que representa el primer paso para su implementación en el diagnóstico

En cuanto al segundo objetivo, la metodología desarrollada para la detección con alta sensibilidad de la MSI en sangre periférica discrimina con una sensibilidad y especificidad del 100% a los pacientes CMMRD del resto de grupos (pacientes SL y de otros síndromes con fenotipo solapante), aunque no ha demostrado suficiente sensibilidad para detectar MSI en los pacientes SL. Esta herramienta, por lo tanto, podría ser especialmente útil para el diagnóstico de CMMRD, sobre todo en los casos con ausencia de mutaciones patogénicas identificadas en los genes MMR. Será necesario testar otros tejidos, como la mucosa colónica normal, para profundizar en la dinámica de MSI en tejidos diana de pacientes afectos de SL antes del desarrollo de la neoplasia.

El trabajo realizado en esta tesis doctoral representa una mejora de las actuales estrategias para el diagnóstico molecular de los síndromes de síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema MMR.

ABSTRACT (English version)

Lynch syndrome (LS) and constitutional mismatch repair deficiency (CMMRD) are hereditary cancer syndromes associated with mismatch repair (MMR) deficiency. Both are characterized by tumours displaying loss of MMR protein expression and/or microsatellite instability (MSI), also reported at low levels in non-neoplastic tissues. Genetic diagnosis of these hereditary cancer syndromes requires identification of germline MMR gene pathogenic mutations; however, that is often hampered by the presence of variants of unknown significance (VUS) and overlapping phenotypes.

The aim of this work was to improve the diagnosis of these syndromes. For it, a new algorithm for VUS classification was proposed, that integrated clinico-pathological data, multifactorial likelihood calculations and functional analyses, allowing the reclassification of 89% of the variants studied in this work. Also, the *in vitro* MMR assay was validated by providing optimized protocols as a first step for meeting quality standards of diagnostic laboratories.

On the other hand, the performance of high-sensitivity MSI (hs-MSI) assessment for the identification of LS and CMMRD in non-neoplastic tissues was evaluated, showing a robustly discrimination between CMMRD and LS and other syndromes with overlapping phenotypes. There were no differences between LS patients and controls. Testing MSI in biopsies of normal colonic mucosa from LS will be necessary to clarify the MSI component in LS non-neoplastic tissues; however, the hs-MSI approach might result into a diagnostic tool for CMMRD diagnosis, especially in cases with a suggestive phenotype and in the absence of identified pathogenic MMR mutations.

In conclusion, the work presented here represents an improvement of the current strategies for the molecular diagnosis of hereditary cancer syndromes associated with MMR deficiency, critically important for the appropriate management of patients and their families.

ÍNDICE

Índice

Abreviaturas generales

Abreviatura	Nomenclatura oficial en inglés	Nomenclatura oficial en castellano
ACMG	*American College of Medical Genetics and Genomics*	-
ADP	*Adenosine DiPhosphate*	Adenosina difosfato
AFAP	*Attenuated Familial Adenomatous Polyposis*	Poliposis adenomatosa familiar atenuada
AMP	*Association for Molecular Pathology*	-
ASE	*Allele-Specific Expression*	Expresión específica de alelo
ATP	*Adenosine-5'-TriPhosphate*	Adenosina trifosfato
AUC	*Area Under the Curve*	Área bajo la curva
C4CMMRD	*Care for CMMRD*	-
CALM	*café-au-lait macules*	Manchas café con leche
CCFR	*Colon Cancer Family Registry*	-
cDNA	*complementary DNA*	ADN complementario
CIMRA	*Cell-free In vitro MmR Activity*	Actividad MMR in vitro libre de célula
CMMRD o BMMRD	*Constitutional (or Biallelic) MisMatch Repair Deficiency*	Deficiencia Constitucional (o Bialélica) de Reparación de Apareamientos Erróneos
DNA	*DeoxyriboNucleic Acid*	Ácido desoxirribonucleico (ADN)
EMAST	*Elevated Microsatellite Alterations at Selected Tetranucleotide repeats*	Elevada inestabilidad de microsatélites en repeticiones de tetranucleótidos
ENIGMA	*Evidence-based Network for the Interpretation of Germline Mutant Alleles*	-
evMSI	*ex vivo MSI*	ex vivo MSI
FAP	*Familial Adenomatous Polyposis*	Poliposis adenomatosa familiar
FFPE	*Formalin-Fixed Paraffin-Embedded*	Fijado en formalina e incrustado en parafina

gMSI	*germline MSI*	MSI germinal
gnomAD	*Genome Aggregation Database*	-
HNPCC	*Hereditary NonPolyposis Colorectal Cancer*	Cáncer colorrectal hereditario no polipósico
hs-MSI	*high sensitivity-MSI*	MSI de alta sensibilidad
HTH	*helix-turn-helix*	Hélice-vuelta-hélice
IARC	*International Agency for Research on Cancer*	-
IHC	*ImmunoHistoChemical*	Inmunohistoquímica
InSiGHT	*International Society for Gastrointestinal Hereditary Tumours*	-
LCL	*Lymphoblastoid Cell Lines*	Líneas linfoblastoides inmortalizadas
LOH	*Loss Of Heterozygosity*	Pérdida de heterocigosidad
LOVD	*Leiden Open Variation Database*	-
LR	*Likelihood Ratios*	Razones de verosimilitud
MLPA	*Multiplex Ligation-dependent Probe Amplification*	Amplificación dependiente de la ligación de sondas multiplexadas
MMR	*MisMatch Repair*	Reparación de apareamientos erróneos
MNNG	*MethylNitroNitrosoGuanidine*	Metilnitronitrosoguanidina
mRNA	*messenger RNA*	ARN mensajero
MSI	*MicroSatellite Instability*	Inestabilidad de microsatélites
MSI-H	*MicroSatellite Instability-High*	Alta inestabilidad de microsatélites
MSI-L	*MicroSatellite Instability-Low*	Baja inestabilidad de microsatélites
MS-MCA	*Methylation Specific-Melting Curve Analysis*	Análisis de la curva de fusión específica de metilación
MS-MLPA	*Methylation-Specific Multiplex Ligation-dependent Probe Amplification*	Amplificación dependiente de la ligación de sondas multiplexadas específicas de metilación
MSS	*MicroSatellite Stable*	Estabilidad de microsatélites

NaMe-PrO	*Nuclease-assisted Minor-allele enrichment with Probe-Overlap*	Enriquecimiento del alelo menor asistido por una nucleasa con una sonda solapante
NF1	*Neurofibromatosis type 1*	Neurofibromatosis tipo 1
NGS	*Next-Generation Sequencing*	Secuenciación de nueva generación
NLS	*Nuclear Localization Signal*	Señal de localización nuclear
NMD	*Nonsense-Mediated mRNA Decay*	Degradación del ARN mensajero mediada por mutaciones terminadoras
PCPE-PCR	*Probe Clamping Primer Extension-PCR*	PCR de extensión del cebador unido a la sonda
PCR	*Polymerase Chain Reaction*	Reacción en cadena de la polimerasa
PPAP	*Polymerase Proofreading-Associated Polyposis*	Poliposis asociada a polimerasas con prueba de lectura
PWWP	*Proline-Tryptophan-Tryptophan-Proline*	Prolina-Triptófano-Triptófano-Prolina
RNA	*RiboNucleic Acid*	Ácido ribonucleico (ARN)
ROC	*Receiver Operating Characteristic*	Característica operativa del receptor
RT-PCR	*Reverse Transcription PCR*	PCR con transcriptasa inversa
SL	*Lynch Syndrome (LS)*	Síndrome Lynch
SLL	*Lynch-Like Syndrome (LLS)*	Síndrome Lynch-Like
smMIP	*single-molecule Molecular Inversion Probes*	Sondas moleculares invertidas con molécula única
SNuPE	*Single Nucleotide Primer Extension*	Secuenciación de la reacción de extensión del nucleótido
SOP	*Standard Operating Procedures*	Procedimientos operativos estándar
SP-PCR	*Small-Pool PCR*	PCR de pequeña cantidad de DNA
TCGA	*The Cancer Genome Atlas program*	-
TIL	*Tumor-Infiltrating Lymphocytes*	Linfocitos infiltrantes de tumor
VUS	*Variant of Unknown Significance*	Variante de significado desconocido

Genes

Símbolo	Nombre HGNC (HUGO Gene Nomenclature Committee)
APC	*Adenomatous Polyposis Coli*
BRAF	*B-Raf proto-oncogene, serine/threonine kinase*
BRCA1	*BRCA1 DNA repair associated (previous name: Breast Cancer 1, early onset)*
BRCA2	*BRCA2 DNA repair associated (previous name: Breast Cancer 2, early onset)*
BUB1	*BUB1 mitotic checkpoint serine/threonine kinase*
CDH1	*Cadherin 1*
CTNNB1	*Catenin beta 1*
EPCAM	*Epithelial cell adhesion molecule (previous symbol: TACSTD1)*
EXO1	*Exonuclease 1*
FAN1	*FANCD2 and FANCI associated nuclease 1*
KRAS	*KRAS proto-oncogene, GTPase (previous name: Kirsten rat sarcoma 2 viral oncogene homolog)*
MCM9	*Minichromosome maintenance 9 homologous recombination repair factor*
MLH1	*mutL homolog 1*
MLH3	*mutL homolog 3*
MSH2	*mutS homolog 2*
MSH3	*mutS homolog 3*
MSH6	*mutS homolog 6*
MUTYH	*mutY DNA glycosylase*
NF1	*Neurofibromin 1*
PCNA	*Proliferating cell nuclear antigen*
PD-1 (PDCD1)	*Programmed cell death protein 1*
PD-L1 (CD274)	*Programmed Death-ligand 1 (CD274 molecule)*
PMS1	*PMS1 homolog 1, mismatch repair system component*

PMS2	*PMS1 homolog 2, mismatch repair system component*
PMS2CL	*PMS2 C-terminal like pseudogene*
POLD1	*DNA polymerase delta 1, catalytic subunit*
POLE	*DNA polymerase epsilon, catalytic subunit*
PTEN	*Phosphatase and tensin homolog*
RFC	*Replication factor C*
RPA	*Replication protein A*
SETD2	*SET domain containing 2, histone lysine methyltransferase*
SPRED1	*Sprouty related EVH1 domain containing 1*

Terminología en inglés

En esta tesis doctoral se ha decidido mantener la palabra original, sin traducir, de los siguientes conceptos listados:

Concepto	Descripción
Frameshift	Mutación con "*desplazamiento, desfase o cambio del marco de lectura*"
Hit	"*Golpe*", referido al hecho de que se produzca una mutación
Missense	Mutación con "*con cambio de sentido*"
Molecular barcodes	"*Código de barras molecular*"
Nonsense	Mutación "*sin sentido*" que provoca la aparición de un codón de terminación prematuro
Primer	"*Cebador*"
Proofreading	Capacidad "*prueba de lectura*" de algunas DNA polimerasas
Slippage	"*Patinaje*" de la polimerasa sobre la hebra de DNA
Splicing	"*corte y empalme*", referido al procesamiento del RNA mensajero
Two-hits	"*dos golpes*"
Wild-type	Referido al alelo "*salvaje*"

INTRODUCCIÓN

1. Sistema de reparación de apareamientos erróneos

1.1 Descripción del sistema de reparación de apareamientos erróneos

El sistema de reparación de apareamientos erróneos o *MMR* (de sus siglas en inglés *MisMatch Repair*) es un sistema de reparación postreplicativo del DNA que tiene la función de corregir los errores principalmente introducidos por las DNA polimerasas durante la replicación del genoma. Estos errores consisten en apareamientos erróneos de las bases nitrogenadas, causados por la incorporación accidental de una base errónea, y pequeñas inserciones o deleciones de nucleótidos debidas al deslizamiento o "*slippage*" de las polimerasas sobre secuencias cortas y repetitivas también llamadas microsatélites (Jiricny, 2013; Reyes et al., 2015). Por lo tanto, aunque generalmente el DNA es replicado con precisión gracias a la alta especificidad de las polimerasas y su capacidad para corregir sus propios errores o "*proofreading*", el sistema MMR representa el segundo mecanismo de control de calidad que posee la célula para asegurar la integridad de la información genética a lo largo de generaciones sucesivas (Kunkel, 2004; Kunkel & Erie, 2015). Asimismo, este sistema de reparación no solo corrige los errores que han escapado de la actividad *proofreading* de las polimerasas, sino que también corrige los errores que se forman durante la recombinación, las modificaciones químicas de las bases (por ejemplo a causa de agentes quimioterapéuticos). Cabe destacar que muchos de sus componentes intervienen en diferentes procesos celulares como el control del ciclo celular, la apoptosis, la hipermutación somática de los genes de las inmunoglobulinas o la expansión de tripletes de nucleótidos (Altieri et al., 2008; Harfe & Jinks-Robertson, 2000; Jun et al., 2006; Li, 2008; Peña-Diaz & Jiricny, 2012).

El sistema de reparación MMR fue inicialmente identificado en bacterias, donde se observó que su inactivación comportaba un aumento de mutaciones espontáneas en *Escherichia coli* (Fazakerley et al., 1986; Modrich, 1991; Radman & Wagner, 1986). La

reacción básica de reparación, que consiste en la escisión de parte de la cadena que contiene el error y resíntesis de la misma, se ha conservado de bacterias a seres humanos (Groothuizen & Sixma, 2016) donde se han descrito proteínas homólogas a las que conforman el sistema MMR en *E. coli* **(Tabla 1)**. La relevancia de su función se demuestra por el grado de conservación del proceso de reparación entre las diferentes especies así como la especialización y/o superposición del funcionamiento de algunas proteínas a lo largo de la evolución.

Tabla 1. Relación de las proteínas del sistema MMR en *E. coli*, *S. cerevisiae* y *H. sapiens*. Adaptado de (Reyes et al., 2015).

E. coli	*S. cerevisiae*	*H. sapiens*	Descripción
MutS-MutS	MSH2-MSH6 (MutSα)	MSH2-MSH6 (MutSα)	Complejo de reconocimiento del apareamiento erróneo. Homodímero en *E. coli* y heterodímero en eucariotas. MutSα y MutSβ presentan cierta especificidad solapante en el reconocimiento del error y son parcialmente redundantes
	MSH2-MSH3 (MutSβ)	MSH2-MSH3 (MutSβ)	
MutL-MutL	MLH1-PMS1 (MutLα)	MLH1-PMS2 (MutLα)	Homodímero en *E. coli* y heterodímero en eucariotas. En *E. coli*, MutL promueve la reacción de corte de la cadena via MutH, mientras que en eucariotas MutLα tiene actividad endonucleasa intrínseca
	MLH1-MLH2 (MutLβ)	MLH1-PMS1 (MutLβ)	MutLβ es un factor accesorio para la reparación del apareamiento erróneo
	MLH1-MLH3 (MutLγ)	MLH1-MLH3 (MutLγ)	MutLγ puede substituir a MutLα en una pequeña fracción de apareamientos erróneos, pero principalmente actúa en la recombinación meiótica
Metilasa Dam	Absente	Absente	Promueve la metilación de los sitios d(GATC), sirve como señal de descriminación de cadena en *E. coli*
MutH	Absente	Absente	Endonucleada, corta la cadena hija usando los sitios d(GATC) hemi-metilados
Ninguno	EXO I	EXO I	5'-3' dsDNA exonucleasa, actúa en la reacción de escisión
RecJ, EXO VII	Ninguno	Ninguno	5'-3' dsDNA exonucleasa, actúa en la reacción de escisión
EXO I, EXO VII, EXO X	Ninguno	Ninguno	3'-5' dsDNA exonucleasa, actúa en la reacción de escisión
UvrD	Ninguno	Ninguno	DNA helicasa II, promueve reacción de escisión, activado por MutS
β-clamp	PCNA	PCNA	Factor de procesividad de la DNA polimerasa. En eucariotas estimula la actividad endonuclesa de MutLα
γ-Complejo	RFC	RFC	Se une a β-clamp / PCNA
SSB	RPA1-3	RPA1-3	Proteína de unión a ssDNA, interviene en la reacción de escisión y resíntesis
DNA Pol III	POLδ	POLδ, POLε	DNA polimerasa que resintetiza el fragmento escindido de la cadena de DNA
DNA ligasa	Desconocido	Ligasa I	Liga las cadenas después de la resíntesis del DNA

En la especie humana, el proceso de reparación MMR se estructura en diferentes fases secuenciales (Kunkel & Erie, 2015; Reyes et al., 2015).

Inicialmente, el error es reconocido por los heterodímeros formados por las proteínas MSH2-MSH6 o MSH2-MSH3 (también llamados MutSα o MutSβ, respectivamente, siendo ambos homólogos al MutS bacteriano). Aunque existe cierto solapamiento de funciones, el complejo MSH2-MSH6 se une preferentemente a bases desapareadas y a zonas de inserción/deleción de menos de 2 bases, mientras que MSH2-MSH3 se une a zonas de inserción/deleción de más de 2 bases y sólo reconoce un espectro limitado de apareamientos erróneos (Harrington & Kolodner, 2007; Srivatsan et al., 2014). El complejo MSH2-MSH6 es, en la especie humana, unas 10 veces más abundante que MSH2-MSH3 y representa el principal complejo de reconocimiento de errores. Cabe señalar que los defectos en MSH6 tienen un menor impacto funcional que los defectos en MSH2, probablemente a causa de la presencia de MSH2 en ambos complejos. Por otro lado, se ha descrito que MSH2-MSH3 juega un papel importante en la expansión de tripletes de nucleótidos, elemento característico de enfermedades hereditarias como el síndrome X Frágil o la enfermedad de Huntington (Flower et al., 2019; McMurray, 2010; Slean et al., 2008).

El complejo MSH2-MSH6 tiene, al igual que su homólogo bacteriano, una actividad intrínseca de unión e hidrólisis del ATP. Una vez unido a la región con el apareamiento erróneo, hidroliza el ATP y se produce un cambio conformacional que da lugar a una estructura en forma de pinza. Esta conformación permite que el complejo pueda moverse por la región con el error así como facilitar la interacción con otras proteínas (J. B. Lee et al., 2014). A continuación, al complejo activado de MSH2-MSH6, se le une el heterodímero formado por las proteínas MLH1-PMS2 (también llamado MutLα y homólogo al MutL bacteriano). MLH1-PMS2 tiene actividad endonucleasa, localizada en PMS2, y representa un elemento clave para la reparación del error. Aparte del complejo MLH1-PMS2, existen otros dos complejos homólogos al MutL bacteriano, MLH1-PMS1 (MutLβ) y MLH1-MLH3 (MutLγ).

MLH1-PMS1 carece de dominio endonucleasa y se ha propuesto como un factor accesorio a la reacción MMR (Campbell et al., 2014). En cambio MLH1-MLH3 conserva la actividad endonucleasa y se encuentra involucrado de manera principal en la recombinación meiótica (Manhart & Alani, 2016; Nishant et al., 2008). De hecho se ha descrito que MLH1-MLH3 puede llegar a reparar una pequeña proporción de inserciones/deleciones interactuando con MSH2-MSH3 (Flores-Rozas & Kolodner, 1998; Nishant et al., 2008).

Tras la unión de MLH1-PMS2 a MSH2-MSH6, PCNA (proliferating cell nuclear antigen) y RFC (replication factor C), ambos componentes de la maquinaria de replicación, son reclutados al complejo activando al función endonucleasa de MLH1-PMS2 gracias a un proceso dependiente de ATP. Seguidamente, MLH1-PMS2 introducirá un corte en la cadena donde se ha producido el error y la exonucleasa 1 (EXO1) será la encargada de degradar el fragmento de DNA que contiene el apareamiento erróneo. Mientras que EXO1 está actuando, las proteínas RPA (replication protein A) se unirán al DNA de cadena sencilla para protegerlo de las nucleasas. Existen otros dos mecanismos alternativos a la eliminación del error mediante EXO1: el primero consiste en la escisión de la cadena que contiene el error por el desplazamiento de la misma a causa de la síntesis de una nueva cadena de DNA por parte de las polimerasas (*strand-displacement synthesis*) (Kadyrov et al., 2009). El segundo, es la escisión la actividad exonucleasa 3'-5' de las polimerasas POL δ y POL ε, aunque este mecanismo aún no se ha demostrado *in vivo* y sigue siendo controvertido (D. Liu et al., 2017; McCulloch et al., 2004). Una vez eliminado el error, la cadena de DNA se resintetiza por acción de las polimerasas POL δ o POL ε en presencia de PCNA, RFC y RPA. Finalmente, la DNA Ligasa I unirá los dos extremos del fragmento sintetizado con la cadena de DNA **(Figura 1)**.

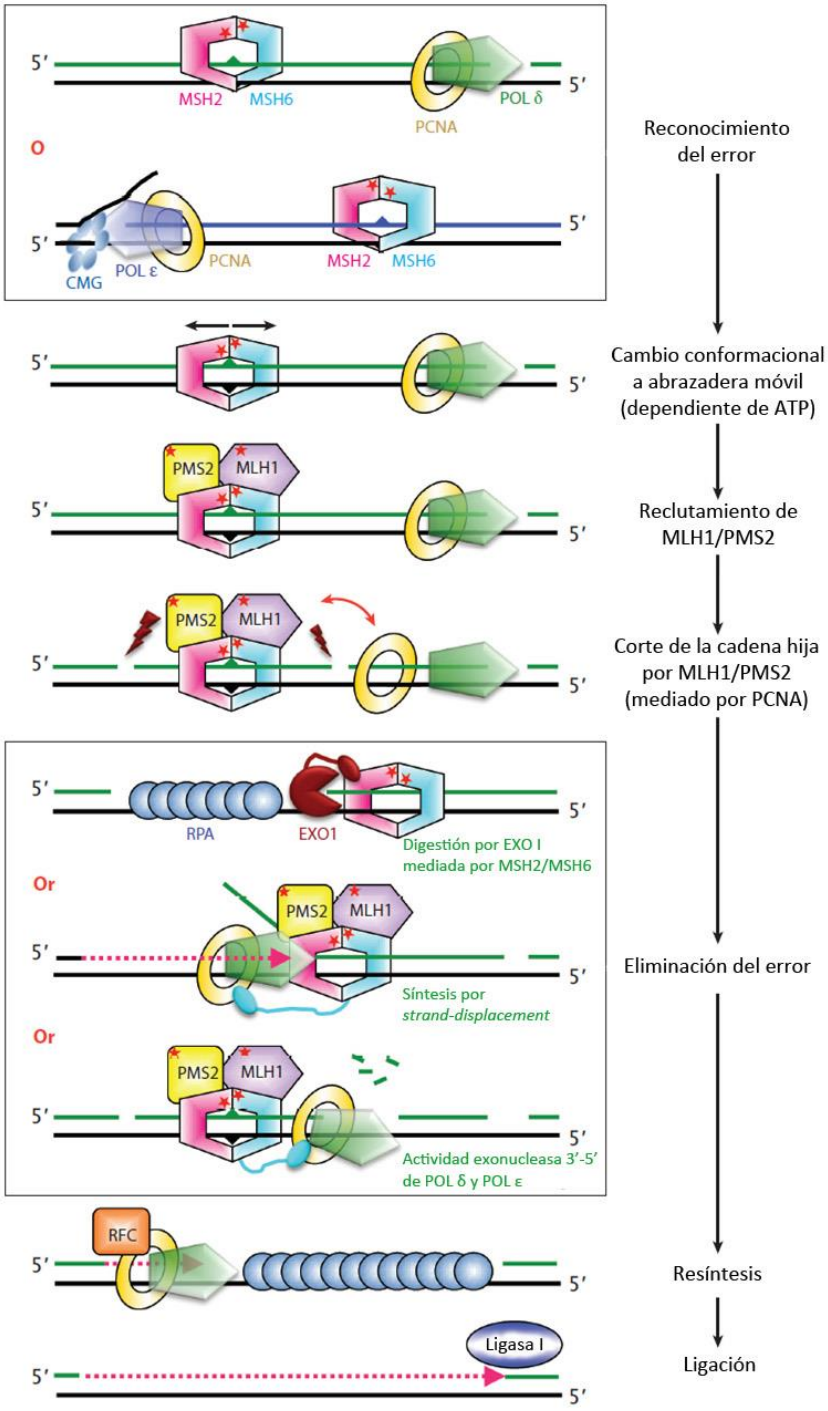


Figura 1. Sistema de reparación de apareamientos erróneos en eucariotas. Adaptado de (Kunkel & Erie, 2015).

Las mutaciones germinales deletéreas en alguno de los cuatro genes principales del sistema MMR (*MLH1*, *MSH2*, *MSH6* y *PMS2*) son la causa de los síndromes de predisposición hereditaria a cáncer asociados a deficiencia del sistema MMR, en concreto el síndrome de Lynch (SL) (OMIM #120435) y el síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos (CMMRD, de sus siglas en inglés *Constitutional MisMatch Repair Deficiency*) (OMIM #276300), que se describirán más adelante (Lynch et al., 2015; Sijmons & Hofstra, 2016; Wimmer et al., 2014). Recientemente, se ha descrito que mutaciones germinales bialélicas en *MSH3* y *MLH3* causan poliposis adenomatosa familiar (FAP, de sus siglas en inglés "*Familial Adenomatous Polyposis*") o su versión atenuada (AFAP) (Adam et al., 2016; Olkinuora et al., 2019).

1.2 Principales genes involucrados en el sistema MMR

MLH1

El gen *MLH1* de la especie humana (NM_00249.3) se localiza en el brazo corto del cromosoma 3 (3p21.3) y fue identificado por primera vez en 1994 gracias al análisis de ligamiento en familias con síndrome de Lynch (Bronner et al., 1994; Papadopoulos et al., 1994). Consta de 19 exones, comprende una región de 72557 pb y su tránscrito principal tiene una longitud de 2752 pb **(Figura 2A)**. La proteína resultante se compone de 756 aminoácidos y su peso molecular es de 84,6 kDa. Actúa como heterodímero uniéndose mayoritariamente a PMS2 y su función principal es reclutar otras proteínas y modular su actividad. En el dominio N-terminal se localiza el dominio de unión e hidrólisis del ATP, primordial para el cambio conformacional que ha de sufrir MLH1-PMS2 para poder realizar su función (Groothuizen & Sixma, 2016; Raschle et al., 2002), mientras que en el extremo C-terminal se encuentra el lugar de unión a PMS2, PMS1 y MLH3. MLH1 y PMS2, o alternativamente PMS1 o MLH3, dimerizan por su extremo C-terminal (Schmutte et al., 2001) y, en ambos casos, los respectivos dominios N- y C-terminal se encuentran unidos por una secuencia de enlace sin estructura (Groothuizen & Sixma, 2016; Kunkel & Erie, 2015)

(Figura 2B). Aparte de estos dominios, MLH1 cuenta con un dominio central de interacción con MSH2, MSH6 y MSH3, y otro para EXO1. Por último, también tiene señales de localización nuclear (NLS) ya que su función la realiza en el núcleo (Brieger et al., 2005; Leong et al., 2009) **(Figura 2A)**.

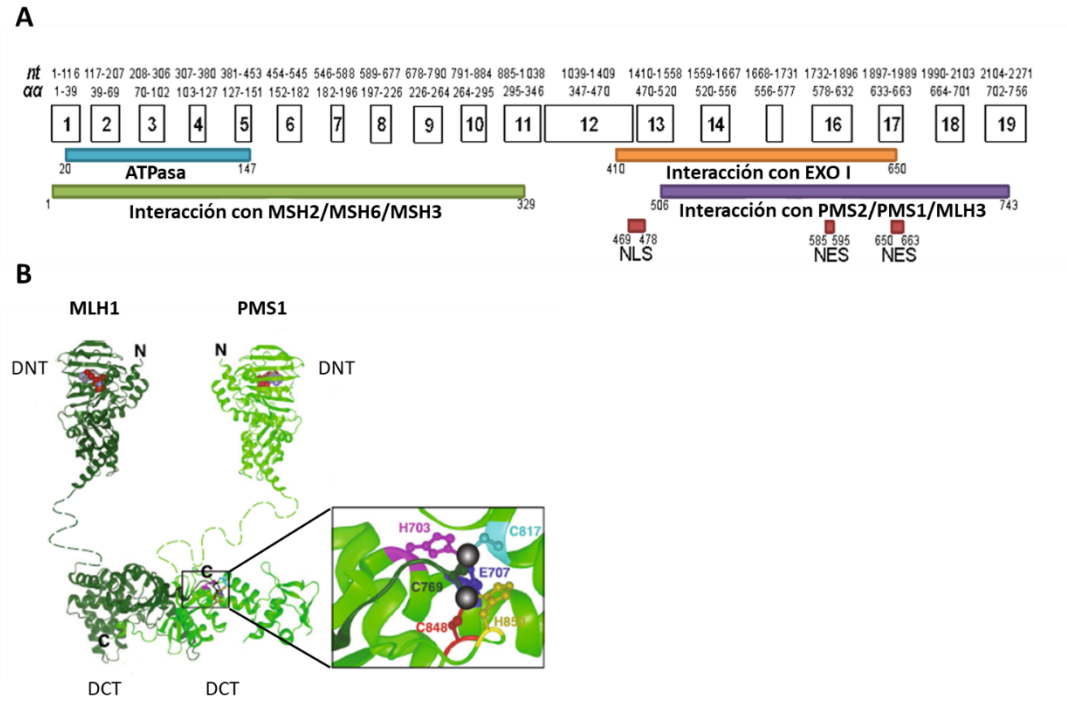


Figura 2. Representación del gen y proteína MLH1. A) Representación gráfica del gen y la proteína MLH1 humana. Adaptado de las guías "*MMR gene variant classification criteria*" de InSiGHT (https://www.insight-group.org/criteria/). **B)** Diagrama de la estructura de MLH1/PMS1 (hPMS2) en *S. cerevisiae*. Recuadrado en negro se muestra la vista ampliada de los aminoácidos conservados que componen el sitio con actividad endonucleasa, localizado en PMS1 (hPMS2). DNT, dominio N-terminal; DCT, dominio C-terminal. Adaptado de (Reyes et al., 2015).

PMS2

El gen *PMS2* (NM_000535.5) se localiza en el brazo corto del cromosoma 7 (7p22.2) y fue identificado en 1994 por Nicolaides y colaboradores (Nicolaides et al., 1994) gracias a la conservación de su secuencia respecto a su homólogo en levadura. Está formado por 15 exones, comprende una región de 35886 pb y su tránscrito principal tiene una longitud de

2855 pb **(Figura 3)**. La proteína codificada consta de 862 aminoácidos y tiene un peso de 95,8 KDa. Actúa como heterodímero uniéndose a MLH1 por su extremo C-terminal y de esta unión depende su localización nuclear y su estabilidad (Brieger et al., 2005; Chang et al., 2000). En su dominio C-terminal también se encuentra localizado el dominio de actividad endonucleasa, primordial para su función dentro del sistema de reparación MMR (Kadyrov et al., 2006) **(Figura 2B)**. Finalmente, en la región N-terminal, se encuentra el dominio de unión e hidrólisis del ATP **(Figura 3)**.

El análisis del gen *PMS2* es especialmente complejo debido a la presencia de múltiples pseudogenes. Catorce de ellos comparten una elevada homología con el extremo 5' del gen, mientras que otro, el llamado PMS2CL, es homólogo a los exones 9 y 11-15 (Blount & Prakash, 2018) **(Figura 3)**. Para poder realizar el análisis mutacional del gen *PMS2*, se han diseñado estrategias concretas para evitar la amplificación de sus pseudogenes (Clendenning et al., 2006; Etzler et al., 2008; Vaughn et al., 2010; Wimmer & Wernstedt, 2014).

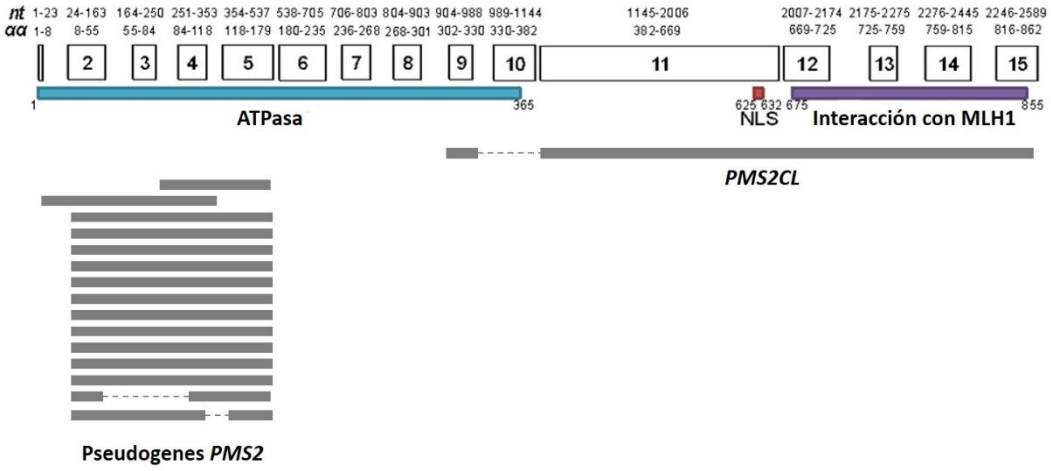


Figura 3. Representación gráfica del gen y la proteína PMS2 humana y sus pseudogenes. Adaptado de las guías "*MMR gene variant classification criteria*" de InSiGHT (https://www.insight-group.org/criteria/) y de (Sugano et al., 2016).

MSH2

El gen *MSH2* (NM_000251.2) se localiza en el brazo corto del cromosoma 2 (2p21) y fue identificado en 1993 por Richard Fishel y Fredrick Leach (Fishel et al., 1993; Leach et al., 1993). Está compuesto por 16 exones, comprende una región de 80259 pb y su tránscrito principal tiene una longitud de 3307 pb **(Figura 4A)**. La proteína codificada consta de 934 aminoácidos y tiene un peso de 104,74 kDa. MSH2 actúa como heterodímero uniéndose mayoritariamente a MSH6 por su extremo C-terminal y en esta región se encuentra el dominio de interacción con MSH6 y MSH3, el dominio ATPasa y un dominio HTH (*helix-turn-helix*). Estos dos dominios, el ATPasa y el HTH, también se ven involucrado en la dimerización con MSH6 y, cuando ésta se produce, los dominios ATPasa de ambas proteínas quedan parcialmente entrelazados **(Figura 4B)**. En función de su unión a ATP o ADP, estos dos dominios modulan la conformación de todo el heterodímero (Groothuizen & Sixma, 2016; Obmolova et al., 2000). En el extremo más N-terminal, en cambio, se localiza el dominio de interacción con el DNA. A continuación, se encuentra el dominio conector, que une el dominio de unión al DNA con el resto de la proteína y, además, se encarga de las interacciones intramoleculares y las señales alostéricas producidas entre los diferentes dominios (Warren et al., 2007). Seguidamente, un dominio palanca o *lever* que une a un dominio abrazadera o *clamp* y a otro dominio palanca. Cada dominio palanca se encarga de conectar respectivamente el dominio de unión al DNA y el dominio ATPasa con el dominio abrazadera y, seguidamente, el dominio abrazadera es el encargado de enviar señales entre ambas zonas de la proteína (Gammie et al., 2007). Por otro lado, *MSH2* también cuenta con un dominio de interacción con MLH1-PMS2 y otro para interaccionar y estabilizar EXO1 **(Figura 4A)**.

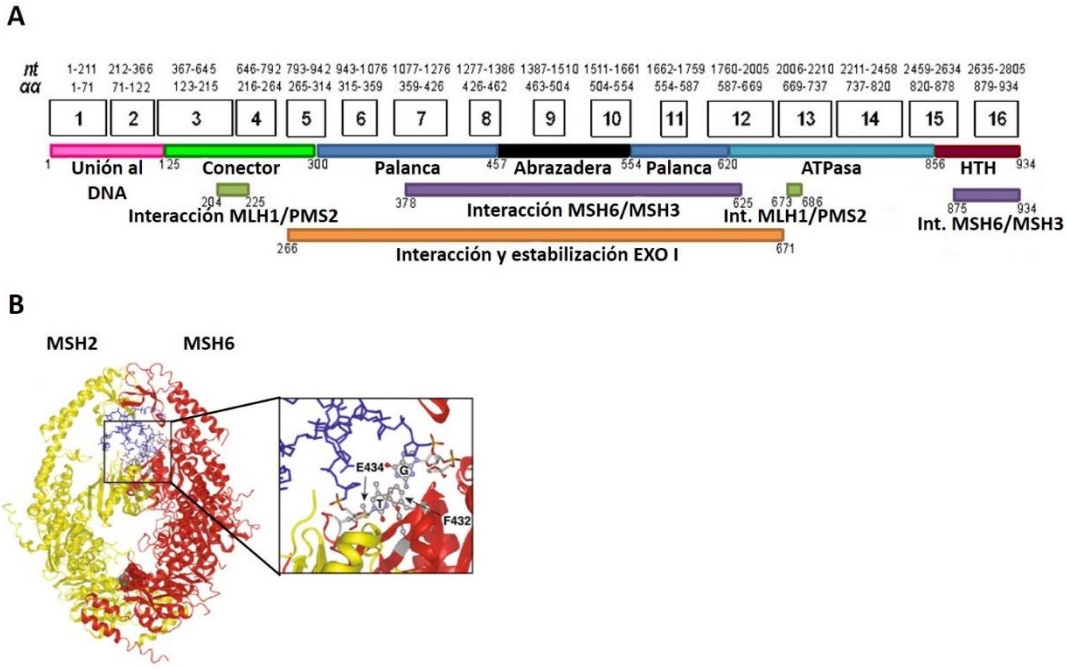


Figura 4. Representación del gen y proteína MSH2. A) Representación gráfica del gen y la proteína MSH2 humana. Adaptado de las guías "*MMR gene variant classification criteria*" de InSiGHT (https://www.insight-group.org/criteria/). **B)** Diagrama de la estructura de MSH2/MSH6 en humanos. Recuadrado en negro se muestra la vista ampliada de la región que interactúa con las bases mal emparejadas, localizada en MSH6. Se indican los residuos fenilalanina 432 (F432) y ácido glutámico 432 (E432), así como las bases mal emparejadas guanina (G) y timina (T). Adaptado de (Reyes et al., 2015).

MSH6

El gen *MSH6* (NM_000179.2) se localiza también el brazo corto del cromosoma 2 (2p16) y fue identificado en 1997 gracias a la cosegregación de una mutación deléterea en este gen en una familia con SL y múltiples individuos afectos de cáncer (Miyaki et al., 1997). Está formado por 10 exones, comprende una región de 23871 pb y su tránscrito principal tiene una longitud de 4328 pb **(Figura 5)**. La proteína codificada tiene 1360 aminoácidos y un peso molecular de 152,79 kDa. MSH6 actúa junto a MSH2 formando un heterodímero y esta unión contribuye a la estabilidad de MSH6. La dimerización de MSH2 y MSH6 forma un complejo asimétrico en el que MSH6 es el encargado de reconocer el error en el DNA y entrar en contacto con él a través de la fenilalanina 432 y el ácido glutámico 434, ambos situados

en el dominio conector de MSH6 (Groothuizen & Sixma, 2016; Jiricny, 2013; Reyes et al., 2015; Warren et al., 2007) **(Figura 4B)**. Aparte del dominio de unión al apareamiento erróneo y el dominio conector, MSH6 consta de un dominio de interacción con PCNA y un dominio PWWP, que es el responsable de que la cromatina pueda reclutar el complejo MSH2-MSH6 (Laguri et al., 2008), además de otros dominios que cumplen las mismas funciones que los dominios que llevan el mismo nombre en la proteína MSH2: el dominio de unión al DNA, el dominio de interacción con MSH2, los dos dominios palanca, el dominio abrazadera y el dominio ATPasa. Por último, MSH6 también cuenta con señales de localización nuclear NLS (Gassman et al., 2011) **(Figura 5)**.

Figura 5. Representación gráfica del gen y la proteína MSH6 humana. Adaptado de las guías "*MMR gene variant classification criteria*" de InSiGHT (https://www.insight-group.org/criteria/).

2. Síndromes hereditarios de predisposición a cáncer asociados a deficiencia del sistema de reparación de apareamientos erróneos

2.1 Síndrome de Lynch

2.1.1 Historia del síndrome de Lynch

El síndrome de Lynch (SL) (OMIM #120435), previamente denominado cáncer colorrectal hereditario no polipósico (HNPCC, de sus siglas en inglés *Hereditary NonPolyposis Colorectal Cancer*), debe su nombre al Dr. Henry Lynch, que en la década de los 60 describió varias familias con una alta agregación de tumores malignos y acuñó por primera vez el concepto de "síndrome familiar de predisposición a cáncer" (Boland & Lynch, 2013; Lynch et al., 2015). Éstas fueron las llamadas Familias N y M (Lynch et al., 1966). No obstante, la primera descripción que se conoce del SL se remite a casi un siglo antes, en 1895, cuando el Dr. Aldred Scott Warthin, patólogo de la Universidad de Michigan, empezó a recopilar datos de una familia con una agregación desproporcionada de tumores gástricos, intestinales y endometriales, a edades muy jóvenes y en diversas generaciones. A esta familia la llamó Familia G y en sus publicaciones sobre el caso ya intuyó que el fenotipo de cáncer era compatible con una herencia mendeliana autosómica dominante (Warthin, 1913). Décadas después, y tras describir a las familias N y M, H. Lynch recuperó la Familia G y contactó con los descendientes, confirmado la predisposición a cáncer de la familia (Lynch & Krush, 1971). En 1984, el SL pasó a denominarse HNPCC para enfatizar sus diferencias con la poliposis adenomatosa familiar; sin embargo, dado que el término HPNCC sólo hacía referencia al cáncer colorrectal y el síndrome predispone a tipos tumorales, esta terminología cayó en desuso y por consenso se decidió recuperar el término "síndrome de Lynch" para este tipo de predisposición a cáncer (Boland, 2005; Jass, 2006).

2.1.2 Características genéticas

El SL es una condición autosómica dominante con penetrancia incompleta causada mayoritariamente por mutaciones patogénicas germinales en los genes *MLH1*, *MSH2*, *MSH6* o *PMS2*, y, en una pequeña proporción de casos, por deleciones en el extremo 3' de *EPCAM* (también conocido como *TACSTD1*), que comporta el silenciamiento epigenético de *MSH2* (Lynch et al., 2015; Yurgelun & Hampel, 2018). También se han descrito epimutaciones constitucionales en el gen *MLH1* como causa del SL, que comportan la metilación hemialélica del promotor del gen en todos los tejidos del organismo. Entre éstas, las epimutaciones primarias (sin mutación genética asociada detectada) son de origen desconocido y no siguen un patrón de herencia mendeliano (Damaso et al., 2018; Hitchins et al., 2007; M. Morak et al., 2008; Suter et al., 2004; Ward et al., 2013).

Para entender cómo el hecho de ser portador de una mutación patogénica predispone a desarrollar cáncer, hay que remitirse al modelo propuesto por Knudson en el año 1985 y que fue llamado modelo "*two-hits*" (Knudson, 1985). En él, la pérdida del alelo salvaje o *wild-type* (no mutado) en las células somáticas comportaría la pérdida total del sistema MMR en dichas células y el establecimiento de un fenotipo mutador que acabaría promoviendo la tumorogénesis **(Figura 6)**. Esta inactivación del alelo *wild-type* se da habitualmente por mutaciones puntuales, deleciones del gen o pérdida de heterozigosidad o LOH (de sus siglas en inglés *Loss Of Heterozygosity*) (de la Chapelle, 2004; Hemminki et al., 1994).

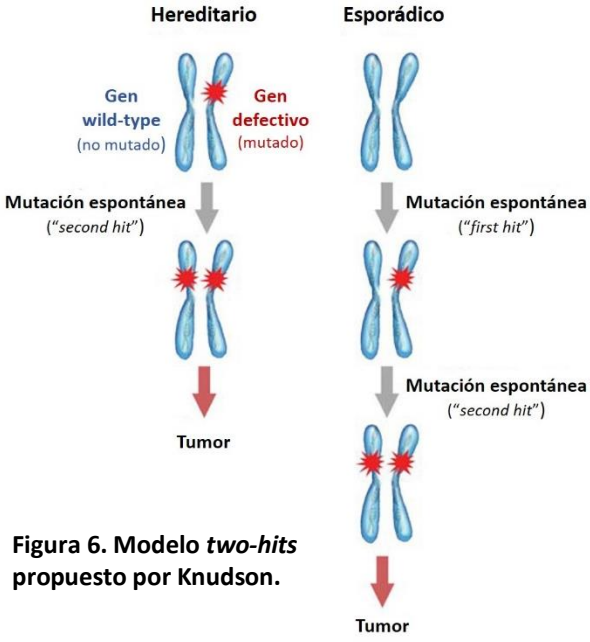


Figura 6. Modelo *two-hits* propuesto por Knudson.

Casi en el 90% de las familias que cumplen criterios Amsterdam con tumores MSI se detectan mutaciones en los genes MMR (Peltomaki, 2016; Elena M. Stoffel et al., 2018). Según los datos recogidos en 2018 de la *International Society for Gastrointestinal Hereditary Tumours (InSiGHT) database* (Plazzer et al., 2013) (https://www.insight-database.org/) y de la base de datos de ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), dos repositorios de variantes MMR, el 35% de las variantes patogénicas y probablemente patogénicas reportadas se localizaron en el gen *MSH2* y el 34% en *MLH1*, mientras que el 21% fue en *MSH6* y sólo 9% en *PMS2*, aunque es probable que esta distribución se deba a los criterios clínicos de derivación de los pacientes SL utilizados hasta el momento (Soto & Castillejo, 2019). Según el tipo de mutación y el cambio predicho a nivel de proteína, la mayoría (entre el 51% al 84% según el gen) de las variantes patogénicas reportadas fueron mutaciones truncantes (*nonsense* y *frameshift*). En una proporción mucho menor, se encontraron las mutaciones que implican el cambio del aminoácido codificado (*missense*) (6-14%) o las que afectan al sitio consenso de *splicing* del RNA (5-16%) **(Figura 7)**. Además, los grandes reordenamientos representaron entre un 3-20% del total de las alteraciones patogénicas (Soto & Castillejo, 2019).

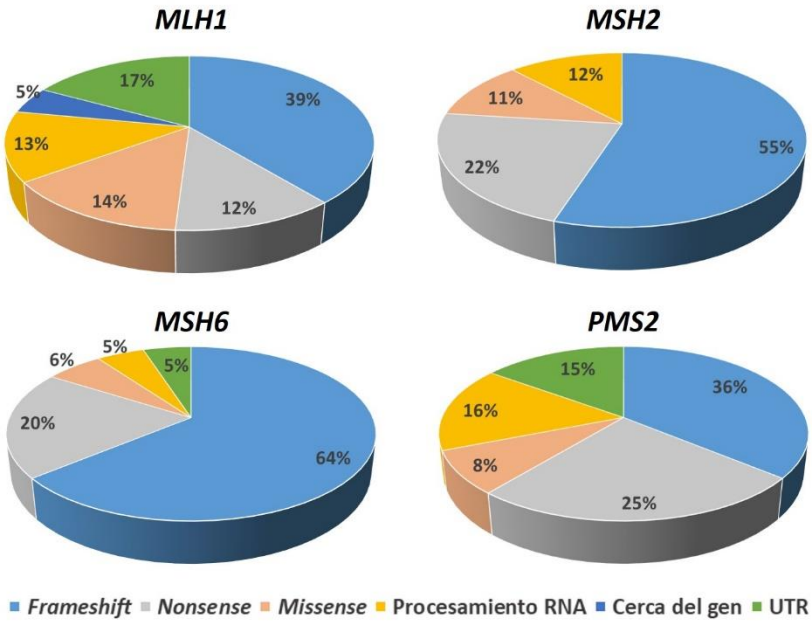


Figura 7. Distribución de las variantes patogénicas según el gen MMR afectado y la naturaleza de la mutación. Adaptado de (Soto & Castillejo, 2019)

2.1.3 Epidemiologia

Tradicionalmente, se ha calculado la prevalencia del SL a partir de datos de pacientes con historial clínico de cáncer. De esta manera, el SL representa el 3% de todos los casos con cáncer colorrectal y el 1,8% de los de endometrio (Biller et al., 2019). Sin embargo, estudios en la población general postulan que el SL es más frecuente y menos penetrante de lo que se había pensado. En un estudio reciente realizado por Win y colaboradores, se utilizaron datos del *Colon Cancer Family Registry* (CCFR) (CCFR: http://coloncfr.org) (Newcomb et al., 2007) para estimar la prevalencia del SL en la población general y observaron que ésta es del 0,36% o 1:279 (IC 95% 1:192-1:402) (Win et al., 2017). Curiosamente, las mutaciones patogénicas en *PMS2* fueron las más prevalentes, presentando una frecuencia del 0,14% (1:714), seguidas de las mutaciones en *MSH6* con una frecuencia del 0,13% (1:758). En contraposición, *MLH1* y *MSH2* presentaron unas frecuencias de 0,051 (1:1946) y 0.035 (1:2841) respectivamente (Win et al., 2017). Estas diferencias en la prevalencia son un reflejo de la variable penetrancia del síndrome en función del gen afectado, siendo PMS2, por ejemplo, el gen que confiere un riesgo más moderado a cáncer de los 4 involucrados en el síndrome (Biller et al., 2019; Moller et al., 2018; Yurgelun et al., 2017).

Por otro lado, la prevalencia del SL puede variar entre poblaciones e incluso aumentar debido al efecto de las mutaciones fundadoras, como es el caso de la población islandesa, la franco-canadiense o la afroamericana, entre otras (Castellsague et al., 2015; Guindalini et al., 2015; Haraldsdottir et al., 2017; Ponti et al., 2015). En el caso de Islandia concretamente, la prevalencia del SL es del 0,44% (1:226) como consecuencia de 3 mutaciones fundadoras (2 en *MSH6* y 1 en *PMS2*), una cifra bastante superior a la observada en el CCFR. No obstante, el riesgo a sufrir cáncer colorrectal en esta población no se ve aumentado, seguramente debido a la naturaleza menos penetrante de *MSH6* y *PMS2* (Haraldsdottir et al., 2017).

2.1.4 Características clínicas

A nivel clínico, el SL se caracteriza por la predisposición a desarrollar tumores a edades jóvenes, principalmente en colon y endometrio. La edad media de aparición del cáncer colorrectal en estos pacientes es de 45 años, en contraposición de los 65 años del cáncer de colon esporádico (Hampel et al., 2005; Lynch et al., 2015). Los tumores se localizan de forma preferente en el colon derecho y acostumbran a darse múltiples neoplasias sincrónicas (diagnosticadas a la vez) o metacrónicas (diagnosticadas más de 6 meses después de la cirugía). Además, a nivel histológico, estos tumores suelen ser poco diferenciados, con rasgos mucinosos, presencia de células en anillo de sello y linfocitos infiltrantes de tumor (TILs), alta infiltración linfocitaria general (reacción de Crohn) y crecimiento medular (Risio et al., 1996). Por otro lado, estos pacientes presentan una carcinogénesis acelerada, siendo el proceso adenoma-carcinoma inferior a 3 años, en contraposición a los 10 o 15 años observados en los tumores esporádicos (Lynch et al., 2015).

Además de tumores colorrectales, los individuos con SL también presentan un riesgo incrementado a desarrollar tumores extracolónicos como cáncer de endometrio, ovario, gástrico, intestino delgado, tracto biliar, páncreas y de las vías urinarias (Moller et al., 2018; Vasen et al., 2013). El cáncer de endometrio representa el tipo de tumoral más frecuente entre las mujeres con SL y acostumbra a ser de tipo endometriode debutando, en promedio, a los 50 años (Moller et al., 2018). Es más, la prevalencia del cáncer de endometrio en la población con síndrome de Lynch afecta de cáncer varía según el gen, siendo las portadoras de mutaciones en *MSH6* las que presentan un mayor riesgo a desarrollar este tipo de tumores (Hendriks et al., 2004; Lynch et al., 2015).

El riesgo para cada tipo de neoplasia varía en función del gen MMR mutado. Así, los portadores en seguimiento de mutaciones en *MLH1* y *MSH2* presentan un riesgo acumulado a los 70 años de sufrir cáncer colorrectal del 40,1% (95% CI 33,5% - 46,7%) y 40,8% (95% CI 31,6% - 50,1%) respectivamente, mientras que los portadores de mutaciones en MSH6 presentan un riesgo del 15,0% (95% CI 3,3% - 26,6%) (Moller et al., 2018). Para el cáncer de

endometrio, el riesgo acumulado a los 70% para los genes *MLH1*, *MSH2* y *MSH6* fue del 40,3% (95% CI 31,5% - 49,1%), 52,7% (95% CI 38,7% - 66,8%) y 46,3% (95% CI 27,3% - 65,0%) respectivamente. En este estudio, los riesgos acumulados para el gen *PMS2* eran negiblibles pero el número limitado de casos no permitió derivar conclusiones. No obstante, en un estudio realizado en 284 familias con mutaciones en *PMS2*, el riesgo acumulado a los 80 años de sufrir cáncer colorrectal o de endometrio fue del 13% para ambos casos (Ten Broeke et al., 2018). En otro estudio realizado en 1108 pacientes SL afectos de cáncer colorrectal, se observó que el riesgo a sufrir este tipo de tumor en pacientes con mutaciones en *PMS2* era del 25,9% (95% IC 7% - 71%), que podría ser incluso superior al riesgo de los portadores de mutaciones en *MSH6* (6,3%; 95% IC 0% - 12,8%) (Sanchez et al., 2017). En este estudio el número de portadores de PMS2 era muy limitado. En la **Tabla 2** se resume el riesgo acumulado para cáncer colorrectal y de endometrio observado en estos diferentes estudios.

Tabla 2. Riesgo acumulado de cáncer colorrectal y de endometrio a los 70 años. RA, riesgo acumulado; CCR, cáncer colorrectal; IC, intervalo de confianza; H, hombres; M, mujeres; CE, cáncer de endometrio. Adaptado de (Biller et al., 2019).

Referencia	Número de familias o pacientes	Genes estudiados (número de pacientes)	% RA de CCR (95% IC)*	% RA de CE (95% IC)*
Moller, Seppala et al. 2018	3119 pacientes SL	*MLH1* (n=1473)	*MLH1*: 40,1% (33,5-46,7)	*MLH1*: 40,3% (31,5-49,1)
		MSH2 (n=1060)	*MSH2*: 40,8% (31,6-50,1)	*MSH2*: 52,7% (38,7-66,8)
		MSH6 (n=462)	*MSH6*: 15,0% (3,3-26-6)	*MSH6*: 46,3% (27,3-65)
		PMS2 (n=124)	*PMS2*: 0	*PMS2*: 26,4% (0,8-51,9)
Ten Broeke, van der Klift et al. 2018	284 familias *PMS2*	*PMS2* (n=797)	*PMS2**: (H): 13% (7,9-22) (M): 13% (7,0,24)	*PMS2**: 13% (7,0-24)
Sanchez, Navarro et al. 2017	1108 pacientes SL	*MLH1* (n=449)	*MLH1*: 25,6% (13,2-38,2)	
		MSH2 (n=371)	*MSH2*: 22,1% (11,3-35,1)	
		MSH6 (n=197)	*MSH6*: 6,3% (0-12,8)	
		PMS2 (n=68)	*PMS2*: 25,9% (7-71)	

* La serie de Ten Broeke reporta el RA a los 80 años, no a los 70.

2.1.5 Características moleculares de los tumores asociados a SL

Como se ha comentado anteriormente, la inactivación somática del alelo *wild-type* conlleva la pérdida completa de la función del sistema MMR en el SL. Esta pérdida de la

función reparadora puede llevar, en algunas células, a la acumulación de errores en secuencias repetitivas del DNA dando lugar a inestabilidad de microsatélites (MSI, de sus siglas en inglés *MicroSatellite Instability*) y, bajo estas condiciones, la tasa de mutación se incrementa entre 100 y 1000 veces, aumentando la probabilidad de que otros genes supresores de tumores y oncogenes se vean afectados y se promueva la tumorogénesis (Pinheiro et al., 2015; The Cancer Genome Atlas et al., 2012). Este proceso mutacional somático subyace a la signatura mutacional 6 definida por Alexandrov y colaboradores, caracterizada por un largo número de substituciones y pequeños indels (principalmente de 1 base) en regiones microsatélites y observada en tumores colorrectales, uterinos y de páncreas, entre otros (Alexandrov et al., 2013). Por otro lado, también se ha descrito que algunas mutaciones en los microsatélites se dan de forma recurrente en los diferentes tipos tumorales (Giannakis et al., 2016; Hause et al., 2016).

Así, la gran mayoría de los tumores del espectro desarrollados en portadores de SL se caracterizan por la pérdida de la expresión de las proteínas reparadoras y/o presencia de MSI en el tejido tumoral. No obstante, la deficiencia reparadora también se ha observado en el 10-15% de los tumores colorrectales esporádicos, principalmente causada por la metilación somática del promotor de *MLH1* y que se ha visto estrechamente relacionada con la mutación V600E en el gen *BRAF* (Gausachs et al., 2012; Yamamoto & Imai, 2015). Por este motivo, la hipermetilación de *MLH1* y la mutación en BRAF V600E, se utilizan para diferenciar los tumores colorrectales MSI y/o MMR deficientes posiblemente asociados a LS de los tumores esporádicos (más información en el apartado 3.1).

Actualmente es motivo de debate la secuencia de lesiones que conducen a la aparición de tumores colorrectales del SL. Tradicionalmente, se había propuesto la deficiencia MMR como un evento tardío en el desarrollo del tumor, es decir, el pólipo aparecería siguiendo el mismo mecanismo que los pólipos esporádicos (mediado por la inactivación de APC) (The Cancer Genome Atlas et al., 2012), tras el cual se produciría la inactivación del sistema MMR que comportaría la acumulación de mutaciones somáticas (p.

ej. MSI) y que, a su vez, aceleraría el proceso carcinogénico (Yurgelun et al., 2012). Sin embargo, estudios recientes postulan que existen dos rutas alternativas a ésta en las que la deficiencia del sistema MMR ocurriría al inicio del proceso. En ellas, la presencia de una cripta deficiente en reparación en la mucosa intestinal, histológicamente normal y sin evidencias de malignidad, derivaría en un adenoma deficiente en reparación que evolucionaría a un tumor o, alternativamente, pasaría de forma directa a un crecimiento invasivo sin pasar por la fase de pólipo gracias a mutaciones en el exón 3 del gen *CTNNB1*, que codifica para la proteína beta-catenina, y que se asocian a un peor pronóstico (Ahadova et al., 2018; Akiyama, 2000) **(Figura 8)**. Este nuevo modelo de carcinogénesis, con 3 posibles vías de las que aún se desconoce la contribución relativa de cada una de ellas al proceso, podría explicar la aparición de los llamados tumores colorrectales de intervalo, diagnosticados entre colonoscopias de seguimiento (Biller et al., 2019).

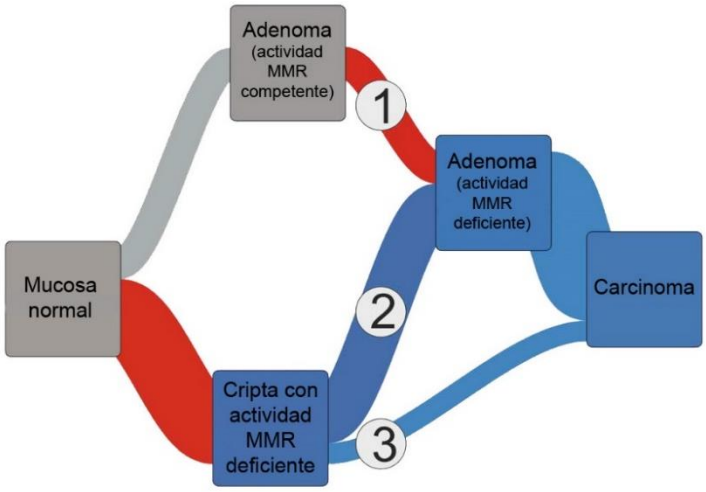


Figura 8. Modelo de la carcinogénesis colorrectal en el síndrome de Lynch. El desarrollo de los tumores colorrectales en el SL sigue 3 posibles rutas: (1) formación de un adenoma con actividad MMR que adquiere la deficiencia de reparación de forma secundaria; o una inactivación inicial del sistema MMR en la mucosa colónica que conduce a (2) formación de un adenoma deficiente en reparación o (3) invasión directa de la pared colónica y desarrollo de un carcinoma. Pintado en rojo se indica el momento en el que se produce la inactivación del sistema MMR y, en azul, cuando la vía MMR ya está inactivada. Adaptado de (Ahadova et al., 2018).

2.1.6 Deficiencia de reparación en tejido no neoplásico

Además de la pérdida de expresión de las proteínas reparadoras y presencia de MSI en los tumores del espectro del SL, en los últimos años se ha reportado que estas características también se pueden detectar, pero a muy baja frecuencia, en tejido no

neoplásico de pacientes con SL. Alazzouzi *et al.* Analizaron el marcador de inestabilidad BAT26 en dos familias con SL mediante técnicas de *clonal sequencing* (clonación del producto de PCR de interés dentro de una bacteria para su posterior análisis) y observaron que, en linfocitos de sangre periférica, la frecuencia de los alelos inestables de BAT26 era de promedio el 5,6% (102/1814), mientras que en los controles negativos (no portadores de mutación en los genes MMR) no se detectaba ningún alelo inestable (Alazzouzi et al., 2005). Del mismo modo, en Coolbaugh-Murphy *et al.* también se detectaron niveles bajos de MSI en sangre periférica de pacientes con SL (11,8% de alelos inestables) utilizando otra técnica llamada *small-pool* PCR (SP-PCR), que consistía en diluir el DNA del paciente a equivalentes de genoma, amplificar los microsatélites dinucleótidos D2S123, D5S346 y D17S518, y posteriormente secuenciarlos (M. I. Coolbaugh-Murphy et al., 2010). Utilizando la misma técnica, también se observó que la MSI incrementaba con la edad en individuos control (M. I. Coolbaugh-Murphy et al., 2005) y que era posible detectar niveles bajos de MSI en saliva de individuos con SL (Hu et al., 2011).

En línea con la teoría de las 3 posibles vías para la carcinogénesis colorrectal en el SL se han identificado criptas deficientes en reparación (presentando pérdida de expresión de proteínas MMR y MSI) en mucosa colónica aparentemente normal de individuos con SL (Kloor et al., 2012; Pai et al., 2018; Shia et al., 2015; Staffa et al., 2015). Estas criptas se distribuyen a lo largo de todo el epitelio intestinal con una frecuencia de 1 por cada 1063 criptas, correpondiente a 1 cripta deficiente por cada 8,6 mm^2 de mucosa colónica (Pai et al., 2018), y no han sido descritas en individuos con cáncer colorectal esporádico o controles sanos.

Recientemente, también se ha descrito pérdida de la expresión de las proteínas MMR en endometrio normal o con hiperplasia simple o compleja, con y sin atípia, en mujeres SL (Nieminen et al., 2009; Niskakoski et al., 2018).

2.2 Síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos

2.2.1 Historia del síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos

El síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos (CMMRD, de sus siglas en inglés *Constitutional MisMatch Repair Deficiency*) (OMIM #276300) fue descrito por primera vez en el año 1999 en dos familias consanguíneas con SL y portadores de una mutación patogénica en *MLH1*. Parte de la descendencia había heredado los 2 alelos mutados por parte de sus progenitores, siendo en consecuencia homocigotos para la mutación patogénica. Los portadores homozigotos presentaban un fenotipo de cáncer mucho más agresivo que el descrito en el SL desarrollando tumores hematológicos y cerebrales a edades infantiles (en un rango de los 14 meses a los 6 años) y, además, presentaban lesiones no neoplásicas que recordaban a las de la neurofibromatosis tipo 1 (NF1) (Ricciardone et al., 1999; Wang et al., 1999).

Como se puede sospechar, la identificación y manejo de los pacientes con CMMRD es complicada y a lo largo de la última década se han puesto en marcha iniciativas como el consorcio europeo "*Care for CMMRD*" (C4CMMRD), el consorcio internacional para la deficiencia bialélica del sistema de reparación de apareamientos erróneos (BMMRD, de sus siglas en inglés *Biallelic MisMatch Repair Deficiency*) y la *European Reference Network* para síndromes raros con riesgo a cáncer (ERN-GENTURIS), que han ayudado a mejorar la comprensión del síndrome y ofrecido guías para su manejo (Durno et al., 2017; Suerink, Ripperger, et al., 2018; Tabori et al., 2017; Vasen et al., 2014; Wimmer et al., 2014).

2.2.2 Características genéticas

El síndrome CMMRD es una condición autosómica recesiva de predisposición a cáncer causada por mutaciones patogénicas bialélicas en alguno de los genes MMR (*MLH1*,

MSH2, *MSH6* y *PMS2*) y, a diferencia del SL, no requiere de un segundo *hit* somático para iniciar el proceso tumorogénico. Su penetrancia es prácticamente completa, ya que es extremadamente raro que un paciente no desarrolle algún tipo de neoplasia antes de la tercera década de vida (Bakry et al., 2014; Lavoine et al., 2015; Wimmer et al., 2014). Esto convierte el síndrome CMMRD en uno de los más penetrantes entre todos los síndromes de predisposición a cáncer infantil o del adulto joven.

Hasta la fecha sólo se han reportado alrededor de 200 casos CMMRD en todo el mundo. Wimmer *et al.*, recogieron datos de 146 pacientes de 91 familias y se observó que el 58% de los casos correspondía a mutaciones bialélicas en *PMS2* mientras que el 20% a MSH6 y el porcentaje restante, un 22%, a mutaciones en *MLH1* o *MSH2* (Wimmer et al., 2014). Esto contrasta notablemente con lo observado en el SL, dónde la mayoría de casos son portadores de mutaciones en *MLH1* o *MSH2*, y se especula que podría ser un reflejo de la baja penetrancia de las mutaciones patogénicas en *PMS2* cuando se dan de forma monoalélicas y de su elevada frecuencia en la población normal en comparación con el resto de genes MMR (Wimmer et al., 2014; Win et al., 2017). Además, también está en discusión la posible letalidad de las mutaciones bialélicas de pérdida de función (nulas) para *MLH1* y *MSH2*, que contribuiría a explicar la distribución polarizada de las mutaciones (Tabori et al., 2017). Por otro lado, la mayoría de mutaciones que se dan en el síndrome CMMRD son de naturaleza truncante, aunque alrededor del 30% de las mutaciones identificadas son mutaciones *missense*, más frecuentes en *MLH1* y *MSH2* que en *MSH6* y *PMS2*, y algunas pueden retener parte de la expresión de la proteína (Colas et al., 2018; Lavoine et al., 2015; Wimmer et al., 2014).

2.2.3 Epidemiologia

La incidencia exacta del síndrome CMMRD sigue siendo desconocida ya que sólo se han reportado alrededor de 200 casos en todo el mundo. No obstante, se cree que es un síndrome infradiagnosticado a causa del desconocimiento médico, su presentación clínica variable y su fenotipo solapante con otros síndromes. Actualmente se propone que su

frecuencia en la población general debe depender de la frecuencia de portadores de mutaciones monoalélicas en los genes MMR y, basándose en los datos publicados en Win *et al.*, se ha estimado que la incidencia del síndrome CMMRD en niños de padres no emparentados debe ser 1:1.000.000 **(Figura 9)**, aunque esta incidencia puede aumentar considerablemente en poblaciones con mutaciones fundadoras o alta consanguinidad (Suerink, Ripperger, et al., 2018; Win et al., 2017).

El síndrome CMMRD comparte características no neoplásicas con otros síndromes como NF1 o el síndrome de Legius, causados por mutaciones patogénicas en los genes *NF1* y *SPRED1* respectivamente, y se cree que un pequeño porcentaje de los pacientes con sospecha de estos dos síndromes pero sin mutaciones identificadas podría presentar CMMRD. Teniendo en cuenta la frecuencia estimada del síndrome CMMRD, el porcentaje de mutaciones detectadas en *NF1*/*SPRED1* en los casos con sospecha de NF1 o síndrome de Legius y la incidencia de mutaciones *de novo* en estos síndromes, se ha estimado que aproximadamente el 0,4% (1/258) de estos casos sin mutación identificada correspondería a casos CMMRD **(Figura 9)** (Suerink, Ripperger, et al., 2018).

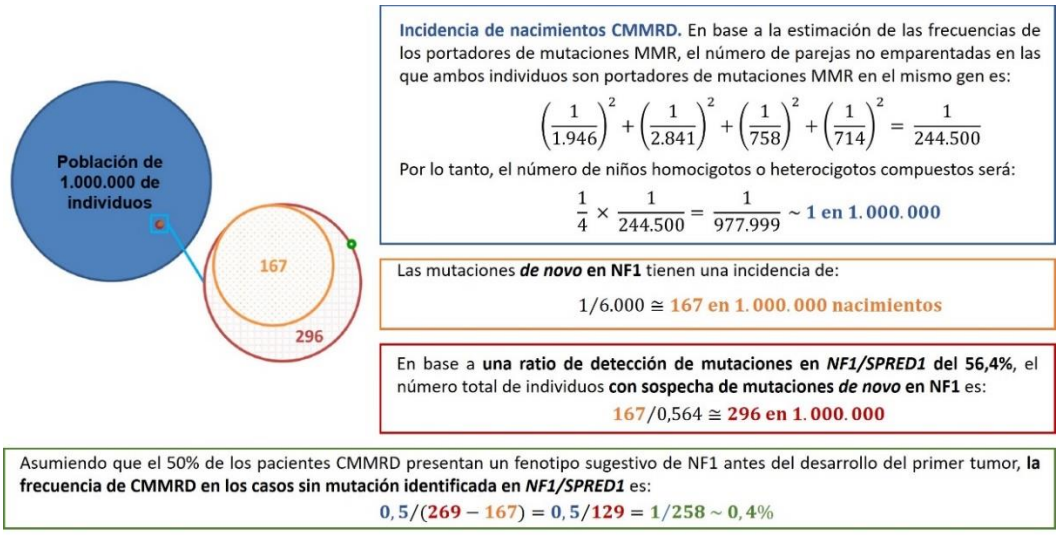


Figura 9. Frecuencia estimada de CMMRD en los casos con sospecha de NF1 o síndrome de Legius sin mutación identificada en *NF1/SPRED1*. CMMRD, deficiencia constitucional de reparación de apareamientos erróneos; NF1, neurofibromatosis tipo 1. Adaptado de (Suerink, Ripperger, et al., 2018).

2.2.4 Características clínicas

A diferencia del SL, el síndrome CMMRD se caracteriza por la predisposición a un amplio abanico de neoplasias a edades muy jóvenes, normalmente durante la infancia, que se pueden agrupar en 3 grupos principales: tumores hematológicos, tumores cerebrales y tumores asociados a SL, predominando los tumores del tracto digestivo por encima de los de endometrio o de las vías urinarias. La edad media de aparición del primer tumor es a los 7,5 años, en un rango que va de los 4 meses a los 39 años (Wimmer et al., 2014). La edad de diagnóstico y la prevalencia del tipo tumoral varían en función del gen afectado; así, los portadores de mutaciones en *MLH1* o *MSH2* debutan, en promedio, a los 5 años y presentan una incidencia mayor de tumores hematológicos que los portadores de mutaciones en *PMS2* o *MSH6*, que debutan en promedio a los 10 y 8 años, respectivamente, y en los que los tumores cerebrales suelen ser más frecuentes que en los bialélicos por mutaciones en *MLH1* o *MSH2* (Wimmer et al., 2014). La supervivencia promedio después del diagnóstico del primer tumor es inferior a los 30 meses, por lo que la mayoría de pacientes no logra alcanzar la edad adulta, aunque los bialélicos de *PMS2* presentan una mayor supervivencia al primer tumor y, por tanto, mayor riesgo a desarrollar tumores metacrónicos (Lavoine et al., 2015; Wimmer et al., 2014). En la **Tabla 3** se recoge la penetrancia estimada para cada tipo de neoplasia en relación al conjunto de casos CMMRD reportados hasta el año 2017 (Aronson et al., 2016; Herkert et al., 2011; Lavoine et al., 2015; Vasen et al., 2014; Wimmer et al., 2014).

Tabla 3. Penetrancia estimada de neoplasias malignas y premalignas en CMMRD. Adaptado de (Durno et al., 2017).

Órgano	Penetrancia estimada, %	Edad media de diagnóstico (rango), años	Referencia
Adenomas del intestino delgado[a]	50	12 (10-20)	Aronson, Gallinger et al. 2016, Herkert, Niessen et al. 2011
Adenomas colorrectales[a]	>90	9 (6-15)	Aronson, Gallinger et al. 2016, Herkert, Niessen et al. 2011
Cáncer de intestino delgado	10	28 (11-42)	Aronson, Gallinger et al. 2016, Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014
Cáncer colorrectal[b]	70	16 (8-48)	Aronson, Gallinger et al. 2016, Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014
Tumores cerebrales de alto grado[c]	70	9 (2-40)	Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014
Linfoma	20-40	5 (0,4-30)	Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014
Leucemia	10-40	8 (2-21)	Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014
Cáncer de endometrio	<10	(19-44)	Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014
Cáncer de la vías urinarias	<10	(10-22)	Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014
Otros[d]	<10	(1-35)	Lavoine, Colas et al. 2015, Vasen, Ghorbanoghli et al. 2014, Wimmer, Kratz et al. 2014, Herkert, Niessen et al. 2011

[a]Adenomas de bajo y alto grado con una rápida progresión probablemente.

[b]Los pacientes se someten a una colectomía subtotal y anastomosis ileoanal, lo que disminuye el riesgo a cáncer colorrectal.

[c]Glioma de alto grado, meduloblastoma y tumores neuroectodérmicos primitivos.

[d]Se han informado menos de 5 casos de las siguientes neoplasias: neuroblastoma, tumor de Wilms, rabdomiosarcoma, osteosarcoma, cáncer de mama, melanoma, tumor neuroectodérmico primitivo de ovario, pilomatricoma y adenoma hepático.

Aparte de las neoplasias descritas, los individuos con CMMRD también presentan características no tumorales que recuerdan a las observadas en NF1 y el síndrome de Legius, como se ha mencionado anteriormente. La principal de ellas es la presencia de máculas *café au lait* o CALMs. Casi el 80% de los individuos CMMRD presentan 2 o más CALMS y, aunque éstas normalmente presentan un borde más irregular que las clásicas CALMS reportadas en NF1, también se han identificado individuos con CALMs clásicas. Además, también se han reportado otros signos típicos de NF1 como neurofibromas o nódulos de Lisch. Por otro lado, prácticamente todos los pacientes con CMMRD acaban desarrollando múltiples adenomas colorrectales sincrónicos en la tercera década de vida, que recuerdan a la versión atenuada de FAP, y también se han reportado casos CMMRD con áreas de piel hipopigmentada, anomalías venosas, agénesis del cuerpo calloso y descenso de los niveles de

inmunoglobulinas, entre otras **(Tabla 4)** (Aronson et al., 2016; Tabori et al., 2017; Wimmer et al., 2014).

Tabla 4. Características no neoplásicas de CMMRD. Adaptado de (Wimmer et al., 2014).

Características no tumorales	Nº de individuos*
Máculas *café au lait* o áreas de piel hiperpigmentadas	91
Máculas *café au lait* y otros signos reminiscentes a NF1	27
Áreas de piel hipopigmentadas	9
Descenso o absencia de las inmunoglobulinas IgG2, IgG4 o IgA y/o incremento de IgM	12
Agénesis del cuerpo calloso con o sin heterotopia de la materia gris	4
Hemangioma cavernoso cerebral	3
Hemangiomas capilar	2
Deformaciones congénitas (asplenia, isomerismo izquierdo, defecto del tabique ventricular)	1
Lupus eritematoso	2

2.2.5 Características moleculares de los tumores

Igual que en el SL, muchos de los tumores del síndrome CMMRD se caracterizan por pérdida de la expresión de las proteínas reparadoras y presencia de MSI. Es más, está deficiencia del sistema MMR también se observa en tejido normal, como se explicará más adelante (apartado 2.2.6). No obstante, se han descrito casos de mutaciones *missense* que retienen parte de la expresión de la proteína que se detecta al analizar el tumor (Wimmer et al., 2014). Además, en algunos tipos tumorales, como es el caso de los hematopoyéticos, es sumamente complicado realizar el análisis de expresión de las proteínas MMR sobre el tumor.

En cuanto a la detección de MSI, la mayoría de los tumores gastrointestinales en pacientes CMMRD, así como el resto de tumores asociados a SL y la mayoría de las neoplasias hematológicas, presentan MSI al utilizarse las técnicas convencionales de detección. Sin embargo, estas técnicas fallan a la hora de analizar tumores cerebrales y otras neoplasias, por lo que estos tumores suelen clasificarse como estables aunque se ha visto

que pueden llegar a tener cambios sutiles en el número de repeticiones del microsatélite (Bakry et al., 2014; Giunti et al., 2009; Wimmer et al., 2014).

Por otro lado, en los últimos 5 años se ha demostrado que la mayoría de los tumores cerebrales asociados a CMMRD son ultra-hipermutados (>250 mutaciones / Mb). Esta tasa excepcionalmente alta de mutaciones somáticas está asociada a mutaciones somáticas que ocurren al inicio del proceso tumorogénico en los genes que codifican para las polimerasas replicativas (*POLE* o *POLD1*) y que inactivan la capacidad *proofreading* de éstas. La combinación de la deficiencia MMR con la deficiencia de la actividad *proofreading* de las polimerasas provoca el incremento de la tasa mutacional y confiere esta ultra-hipermutabilidad, característica de este tipo de tumores (Shlien et al., 2015; Waterfall & Meltzer, 2015).

2.2.6 Deficiencia de reparación en tejido no neoplásico

Como hemos mencionado anteriormente, la deficiencia constitucional en MMR también se observa en tejido no neoplásico de los pacientes con CMMRD de manera que en todos los tejidos normales de los individuos CMMRD se puede observar pérdida de la expresión de las proteínas MMR mediante inmunohistoquímica (IHC, de sus siglas en inglés ImmunoHistoChemical) **(Figura 10)**. Así, la concordancia de la pérdida de expresión entre el tumor y el tejido normal se considera un rasgo diagnóstico del síndrome (Wimmer J Med Genet 2014, Bakry Eur J Cancer 2014).

En cuanto a la detección de MSI en tejidos normales, se requiere de técnicas altamente sensibles para su detección. En 2013, Ingham y colaboradores (Ingham et al., 2013) desarrollaron un método sencillo y mucho más sensible que las PCR convencionales para detectar MSI germinal (gMSI) en sangre periférica. El método consistía en el análisis de 3 microsatélites dinucleótidos (D2S123, D17S250 y D17S791) mediante electroforesis capilar y cuantificación del pico principal que presentaba el microsatélite y los picos típicamente asociados a artefactos de PCR.

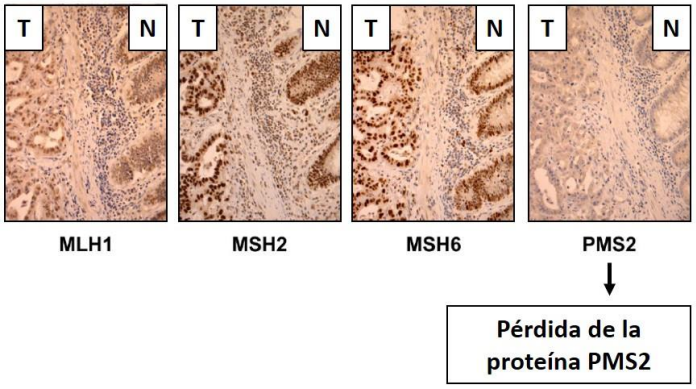


Figura 10. Análisis de expresión de las proteínas MMR mediante inmunohistoquímica en un paciente CMMRD portador de una mutación bialélica en *PMS2*. En las imágenes se muestra la tinción IHC para cada proteína MMR en colon normales (N) y en un adenocarcinoma invasivo (T). La expresión de las proteínas MMR se identifica como una tinción marrón oscuro de los núcleos de las células. La contratinción de los núcleos está hecha con hematoxilina, en azul. Adaptado de (L. Li et al., 2015).

En Ingham *et al.* observaron que la proporción de estos últimos picos en relación al pico principal era significativamente mayor en muestras CMMRD y que se podía establecer un punto de corte para diferenciar los pacientes CMMRD portadores de mutaciones en *MLH1*, *MSH2* y *PMS2*, de los controles. No obstante, la técnica no tenía suficiente sensibilidad para detectar los bialélicos de *MSH6*, posiblemente porque *MSH6* se encuentra principalmente involucrado en la reparación de repeticiones mononucleótidas (de una única base). Con intención de resolver esta limitación, otros autores desarrollaron métodos más complejos basados, por ejemplo, en la obtención de líneas linfoblastoides inmortalizadas (LCL, de sus siglas en inglés *Lymphoblastoid Cell Lines*) a partir de linfocitos del paciente. Estas líneas inmortalizadas se cultivan durante 3 o 4 meses para dar tiempo a que se acumulen errores y, pasado este tiempo, se analiza la llamada MSI *ex vivo* (evMSI). Para complementar los resultados de la evMSI, se somete las células a un ensayo de tolerancia a la citotoxicidad de agentes metilantes como MNNG o la 6-tioguanina, ya que las células con un sistema MMR deficiente presentan resistencia a estos agentes (Bodo et al., 2015). La combinación de la evMSI con el ensayo de tolerancia a agentes metilantes presenta una sensibilidad mayor que el ensayo gMSI de *Ingham et al.* y es capaz de detectar a los portadores de mutaciones de *MSH6*. No obstante, su implementación en la mayoría de

laboratorios de diagnóstico molecular es difícil y el largo tiempo requerido hasta el diagnóstico es una limitación, ya que a veces un resultado rápido es crucial para el paciente.

2.3 Otros síndromes asociados a deficiencia del sistema MMR

Aparte de los síndromes previamente descritos y asociados a deficiencia constitucional de alguno de los genes principales del sistema MMR (*MLH1*, *MSH2*, *MSH6* y *PMS2*), recientemente se han descrito otros síndromes vinculados a defectos en alguno de los otros genes del sistema, como *MSH3* o *MLH3*.

La proteína MSH3 forma actúa como heterodímero uniéndose a MSH2 y se encuentra involucrada en la reparación de inserciones o deleciones de más de 2 bases (Harrington & Kolodner, 2007; Srivatsan et al., 2014). La pérdida de la función de *MSH3* ya se había descrito previamente en tumores colorrectales esporádicos, causando inestabilidad en las regiones microsatélite tetranucleótidas o EMAST (de sus siglas en inglés *Elevated Microsatellite Alterations at Selected Tetranucleotide repeats*). La EMAST se asocia a un mal pronóstico de los tumores típicamente definidos como estables o con inestabilidad baja, ya que la clasificación en tumor estable o inestable viene definida por la presencia de inestabilidad en marcadores mono- y -dinucleótidos pero no en tetranucleótidos (Carethers, 2017). Sin embargo, recientemente se ha reportado que mutaciones germinales bialélicas en *MSH3* causan AFAP y pólipos duodenales. El fenotipo de estos pacientes bialélicos es consistente con la pérdida de la función de *MSH3* ya que el tejido tumoral presenta EMAST y, al igual que en los casos con CMMRD, se puede observar pérdida de la expresión de la proteína MSH3 tanto en tumores como en el tejido normal. Por otro lado, mutaciones de *MSH3* en heterocigosis con el wild-type no parecen aumentar la predisposicón **(Tabla 5)** (Adam et al., 2016).

En cuanto a *MLH3*, actúa cómo heterodímero uniéndose a MLH1 y, aunque puede substituir al complejo MLH1-PMS2 en una pequeña fracción de apareamientos erróneos,

principalmente actúa en la recombinación meiótica (Flores-Rozas & Kolodner, 1998; Nishant et al., 2008). Se ha reportado que mutaciones bialélicas germinales en este gen causan FAP y AFAP, aunque en este caso los tumores son estables y lo que presentan es inestabilidad cromosómica **(Tabla 5)** (Olkinuora et al., 2019).

Taula 5. Fenotipo asociado a mutaciones en los genes del sistema MMR según su dosis génica. SL, síndrome de Lynch; CMMRD, deficiencia constitucional de reparación de apareamientos erróneos; FAP, poliposis adenomatosa familiar; MSI, inestabilidad de microsatélites; MSS, estabilidad de microsatélites; EMAST, inestabilidad en regiones microsatélite tetranucleótidas .

Gen	Dosis génica	Fenotipo asociado	Características en tumor y tejido normal
MLH1	Monoalélico	SL	Pérdida de expresión en tejido tumoral, MSI
	Bialélico	CMMRD	Pérdida de expresión en tejido normal y tumor, MSI
MSH2	Monoalélico	SL	Pérdida de expresión en tejido tumoral, MSI
	Bialélico	CMMRD	Pérdida de expresión en tejido normal y tumor, MSI
MSH6	Monoalélico	SL	Pérdida de expresión en tejido tumoral, MSI
	Bialélico	CMMRD	Pérdida de expresión en tejido normal y tumor, MSI
PMS2	Monoalélico	SL	Pérdida de expresión en tejido tumoral, MSI
	Bialélico	CMMRD	Pérdida de expresión en tejido normal y tumor, MSI
MSH3	Monoalélico	Sin fenotipo asociado	-
	Bialélico	Poliposis	Pérdida de expresión en tejido normal y tumor, EMAST
MLH3	Monoalélico	Sin fenotipo asociado	-
	Bialélico	Poliposis	Pérdida de expresión en tejido normal y tumor, MSS e inestabilidad cromosómica

2.4 Fenotipos intermedios entre SL y CMMRD y síndromes con fenotipo solapante

Como se ha ido comentando a lo largo de los apartados anteriores, aunque a nivel molecular cada síndrome se encuentra bien definido, su diagnóstico clínico, esencial para guiar el posterior análisis molecular, no siempre es fácil a causa de la existencia de fenotipos intermedios entre SL y CMMRD o fenotipos solapantes con otros síndromes no relacionados con la deficiencia del sistema MMR.

Fenotipos intermedios entre SL y CMMRD

Existe cierto solapamiento fenotípico entre el SL y el síndrome CMMRD (Bougeard et al., 2014; Carethers & Stoffel, 2015; Maletzki et al., 2017). En la literatura se pueden encontrar reportados casos atenuados de CMMRD cuya edad de debut del primer cáncer sobrepasa los 30 años, típicamente asociados a mutaciones hipomórficas (con penetrancia reducida), mientras que también hay casos extremadamente agresivos de SL con un primer cáncer diagnosticado antes de los 15 años (Ahn et al., 2016; Aronson et al., 2016; Bodas et al., 2008; C. A. Durno et al., 2015; Kets et al., 2009; L. Li et al., 2015). Todos estos datos apoyan la teoría sobre la existencia de un continuo fenotipo clínico que va desde los fenotipos menos severos de CMMRD, que mimetizan con el SL, hasta los más agresivos de SL que se confunden con el síndrome CMMRD (Bodo et al., 2015; Fernandez-Rozadilla et al., 2019; Lavoine et al., 2015; L. Li et al., 2015).

Solapamientos fenotípicos entre CMMRD y otros síndromes

El síndrome CMMRD también presenta solapamiento fenotípico con otros síndromes, como NF1 y el síndrome de Legius, a causa de las características no neoplásicas que comparten (apartado 2.2.4). Además, el síndrome CMMRD también se puede confundir con FAP o AFAP por la presencia de pólipos colorrectales y con el síndrome Li-Fraumeni al tener ambos síndromes un espectro tumoral similar compuesto de tumores hematológicos y cerebrales a edades jóvenes (Aronson et al., 2016; Michaeli & Tabori, 2018; Shuen et al., 2019; Wimmer et al., 2014). Recientemente, también se han reportado mimetismos entre CMMRD y la poliposis asociada a la actividad reparadora de las polimerasas o (PPAP, de sus siglas en inglés *Polymerase Proofreading-Associated Polyposis*), causada por mutaciones patogénicas heterocigotas en los genes *POLE* y *POLD1*. Este mimetismo entre CMMRD y PPAP se debe, principalmente, al fenotipo ultra-hipermutado de algunos tumores cerebrales CMMRD y a la presencia de CALMs u otras características CMMRD en alguno de los pacientes con PPAP (Wimmer et al., 2017).

Solapamientos fenotípicos entre SL y otros síndromes

Cuando un paciente cumple criterios de SL, es decir, presenta un tumor del espectro del SL con pérdida de expresión de las proteínas reparadoras y/o MSI (en absencia de metilación del promotor de *MLH1*), pero no se le encuentra mutación patogénica causal en los genes MMR, ya sea porque no se ha detectado ninguna variante o porque lo encontrado es una variante de significado desconocido (VUS, de sus siglas en inglés *Variant of Unknown Significance*), pasa a denominarse síndrome *Lynch-Like* (SLL) (Rodriguez-Soler et al., 2013). Aproximadamente el 50% de los pacientes con tumores colorrectales deficientes en el sistema MMR entrarían dentro de esta categoría. Pertenecer a este grupo impide un manejo clínico apropiado así como una estimación del riesgo en el paciente y sus familiares, ya que se considera que tienen un riesgo intermedio entre el LS y la población general a desarrollar cáncer (Buchanan et al., 2014).

Existen ciertos síndromes y condiciones que pueden llegar a mimetizar el fenotipo de SL al provocar, en última instancia, deficiencia del sistema MMR. Estos síndromes solapantes explicarían parte de los casos SLL en los que no se encuentra ninguna variante en los genes MMR. Las mutaciones bialélicas en *MUTYH*, un gen asociado a AFAP, llegan a representar del 1 y al 3% de todos los casos SLL (Castillejo et al., 2014; M. Morak et al., 2014). Así, los defectos en la reparación mediada por *MUTYH* pueden llegar a causar mutaciones somáticas en los genes MMR y mimetizar las características tumorales del SL. De forma similar, recientemente se ha reportado que las mutaciones patogénicas en los genes *POLE* y *POLD1* en línea germinal pueden asociarse a tumores con deficiencia MMR, probablemente a causa de mutar somáticamente los genes MMR, y explicar algunos casos SLL (Elsayed et al., 2015; Jansen et al., 2016).

Por otro lado, gracias a las mejoras en la secuenciación masiva o NGS (de sus siglas en inglés "*Next-Generation Sequencing*"), se han identificado mutaciones germinales en heterocigosis en otros genes como *FAN1*, *MCM9*, *BUB1*, *SETD2* e incluso *BRCA1* y *BRCA2*, en pacientes SLL, por lo que estos genes han empezado a emerger como posibles candidatos

para explicar los casos con fenotipo de SL pero sin mutación en los genes MMR (de Voer et al., 2013; Q. Liu, Hesson, et al., 2016; Vargas-Parra et al., 2017; Yurgelun et al., 2015).

Por último, en una proporción variable de tumores esporádicos se han observado dobles hits somáticos en los genes MMR, lo que confiere al tumor características de SL. Dentro del grupo de pacientes SLL, estos casos representarían del 30 al 82% de los casos, aunque debido a su naturaleza somática, su implicación en una predisposición genética al cáncer es poco probable (Vargas-Parra Int J Cancer 2017, Geurts-Giele J Pathol 2014, Mensenkamp Gastroenterology 2014, Sourrouille Fam Cancer 2013).

3. Diagnóstico molecular y manejo clínico de los síndromes asociados a deficiencia del sistema MMR

3.1 Diagnóstico y manejo del síndrome de Lynch

3.1.1 Criterios de selección y algoritmo diagnóstico del síndrome de Lynch

Con el objetivo de poder identificar aquellas familias candidatas a tener un diagnóstico de SL, en 1991 se consensuaron unos criterios exclusivamente clínicos para ello basados en la historia personal y familiar de cáncer colorrectal del paciente. Fueron los llamados criterios de Amsterdam I **(Tabla 6)** (Vasen et al., 1991). Más tarde, en 1999, estos criterios se revisaron para incluir también los tumores extracolónicos asociados a SL, definiéndose así los criterios de Amsterdam II **(Tabla 6)** (Vasen et al., 1999). Aunque estos criterios eran altamente específicos, a la vez resultaron demasiado restrictivos, así que, debido al tamaño reducido de algunas familias o la falta de historia familiar, se elaboraron en paralelo los criterios de Bethesda (Rodriguez-Bigas et al., 1997), cuyo objetivo era incrementar la sensibilidad en la detección del síndrome incluyendo el análisis de MSI. En 2004, estos criterios también fueron revisados, pasándose a llamar criterios de Bethesda revisados **(Tabla 6)** (Umar et al., 2004).

Con el objetivo de incrementar la tasa de detección del SL, en 2009 se propuso el cribado universal de todos los tumores colorrectales y de endometrio mediante el estudio de la expresión de proteínas reparadoras con IHC y/o análisis de MSI (Hampel, 2010; Hampel et al., 2008; Hampel et al., 2005). Moreira y colaboradores determinaron que la estrategia más coste-efectiva era realizar el cribado en todos aquellos individuos con tumores de cólon a una edad inferior a los 70 años o que cumplieran criterios de Bethesda (criterios de Jerusalen revisados) (Moreira et al., 2012). A día de hoy, el cribado universal o el de los menores de 70 años es el que realiza la mayoría de hospitales. Con todo, debido a la aparición de la NGS, actualmente se encuentra en discusión si la secuenciación de los

tumores mediante NGS resultaría más efectiva que el cribado universal tradicional a la hora de identificar a los pacientes con SL (Biller et al., 2019; Hampel et al., 2018).

Tabla 6. Criterios de Amsterdam I, Amsterda II y criterios de Bethesda revisados. Adaptado de (Lynch et al., 2015).

CRITERIOS DE AMSTERDAM I
Para la selección de una familia candidata a sufrir SL, los criterios de Amsterdam I requieren al menos tres familiares con cáncer colorrectal verificado histológicamente y que se cumplan los siguientes criterios:
1. Uno es pariente de primer grado de los otros dos.
2. Al menos dos generaciones sucesivas están afectadas.
3. Al menos uno de los familiares ha habido de ser diagnosticado de cáncer colorrectal antes de los 50 años.
4. FAP ha sido excluida.
CRITERIOS DE AMSTERDAM II
Para la selección de una familia candidata a sufrir SL, los Criterios de Amsterdam II requieren al menos tres familiares con un tumor asociado a SL (cáncer colorrectal, endometrio, ovario, estómago, intestino delgado, tracto biliar, tracto urinario y sistema nervioso central) y que se cumplan los siguientes criterios:
1. Uno es pariente de primer grado de los otros dos.
2. Al menos dos generaciones sucesivas están afectadas.
3. Al menos uno de los familiares ha habido de ser diagnosticado de un tumor asociado a SL antes de los 50 años.
4. FAP ha sido excluida en los casos con cáncer colorrectal.
5. Los tumores se han de verificar histológicamente siempre que sea posible.
CRITERIOS DE BETHESDA REVISADOS
Un individuo es candidato a sufrir SL si cumple al menos uno de los siguientes criterios:
1. Un diagnóstico de cáncer colorrectal antes de los 50 años.
2. Presencia de tumores colorrectales sincrónicos o metacrónicos u otros tumores asociados a LS, independientemente de la edad.
3. Cáncer colorrectal con histología de MSI-H (presencia de TILs, reacción linfocitaria similar a la de la enfermedad de Crohn, diferenciación mucinosa/presencia de células en anillo de sello o crecimiento medular) diagnosticado antes
4. Un diagnóstico de cáncer colorrectal u otro tumor asociado a SL antes de los 50 años en al menos un familiar de primer grado.
5. Diagnóstico de cáncer colorrectal u otro tumor asociado a SL a cualquier edad en dos familiares de primer o segundo grado.

Por otro lado, a raíz de los criterios clínicos anteriormente comentados, se han desarrollado modelos predictivos para calcular el riesgo que tiene un individuo a sufrir SL. Los modelos más destacados son el MMRPro (Chen et al., 2006), PREMM1, 2, 6 (Kastrinos et al., 2011) y PREMM5 (Kastrinos et al., 2017), donde se recomienda analizar los genes MMR

cuando la probabilidad de riesgo supera el 5%. PREMM5 es el único que incluye *PMS2* y *EPCAM* en sus algoritmos y se recomienda testar estos genes si el riesgo supera el 2,5%.

Dicho esto, el algoritmo diagnóstico para la identificación de individuos SL consiste en una primera selección de los candidatos (cribado poblacional o criterios clínicos (Amsterdam o Bethesda)) seguido del estudio de MSI y/o IHC sobre el tumor, el análisis de *BRAF* y/o la hipermetilación del promotor de *MLH1*. En el caso la proteína MLH1 no se exprese si el gen BRAF está mutado o en ausencia de metilación de *MLH1*, se realiza el estudio genético **(Figura 11)**. Cabe que recordar que, en caso de que el paciente presente hipermetilación del promotor de *MLH1* pero el debut del tumor haya sido a edad joven o presente múltiples tumores, sería aconsejable analizar la metilación germinal de *MLH1* para confirmar o descartar un posible caso de Lynch causado por metilación constitucional de *MLH1* (Hitchins & Ward, 2009).

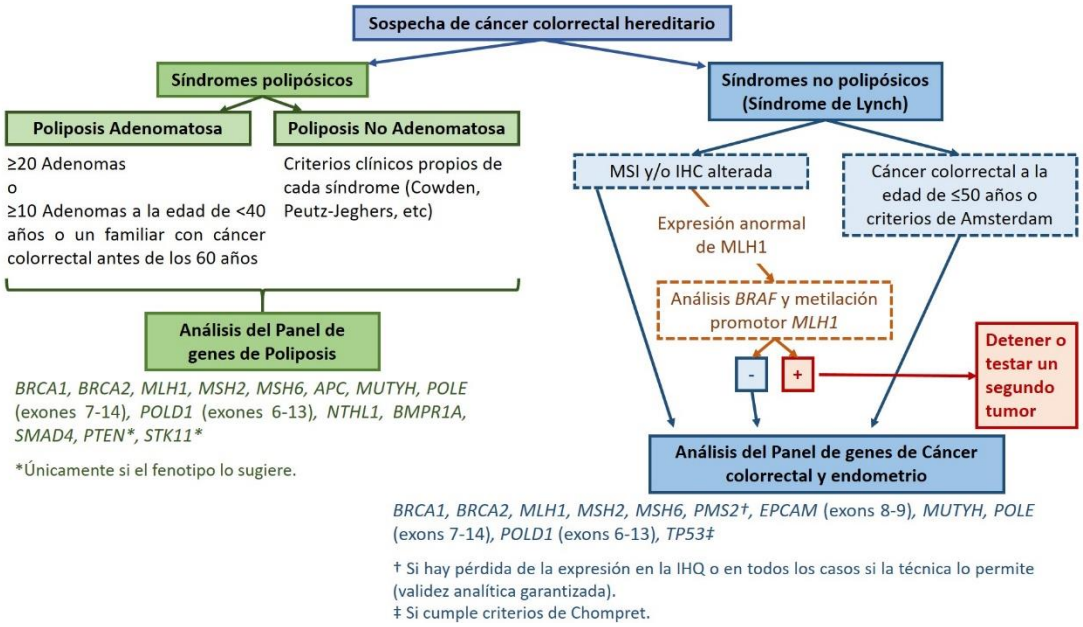


Figura 11. Algoritmo diagnóstico de los casos con sospecha de cáncer colorrectal hereditario. Algortimo basado en el consenso catalán (2019) sobre el uso de paneles de genes en el diagnóstico de cáncer hereditario.

3.1.2 Asesoramiento genético del síndrome de Lynch

La identificación de individuos en riesgo de padecer SL, así como cualquier otro síndrome de predisposición hereditaria a cáncer, es fundamental para reducir la mortalidad y morbilidad de la enfermedad. El asesoramiento genético es el proceso que ayuda al paciente a entender la contribución de la predisposición genética a la aparición del cáncer y que facilita la comprensión y la aceptación de las implicaciones médicas, psicológicas y familiares que conlleva esta predisposición genética (National Society of Genetic Counselors' Definition Task et al., 2006). Es importante que el asesoramiento empiece antes del diagnóstico genético, ya que durante el proceso de consejo genético se obtiene la información clínica personal y familiar del paciente sobre los antecedentes de cáncer, se construye y evalúa el árbol genealógico, se determina el riesgo de susceptibilidad hereditaria a cáncer y se informa al paciente de las posibles implicaciones de ello. Si el diagnóstico genético resulta ser positivo para el SL, será la unidad de consejo genético quién informe al paciente y sus familiares del riesgo asociado a padecer los diferentes tumores del espectro del SL así como las estrategias de prevención y seguimiento. Además, responderá de forma personalizada cualquier duda que tenga la familia sobre su condición hereditaria, asegurando el soporte psicosocial en caso de ser necesario (Giardiello et al., 2014; Hampel, 2016). Por lo tanto, el asesoramiento genético se ha de considerar parte integral del proceso de diagnóstico y ha de ser realizado por un equipo de profesionales especializados debidamente capacitados para el consejo genético (Rantanen et al., 2008; Rolnick et al., 2011).

3.1.3 Estrategias de prevención y seguimiento

Dado que los tumores del SL se desarrollan a edades jóvenes y presentan una carcinogénesis acelerada, tanto la prevención como la detección precoz de dichos tumores suponen una mejora de la calidad de vida y una mayor supervivencia (Moller et al., 2018). La mayoría de guías para el manejo clínico del SL recomiendan la realización de colonoscopias cada 1-2 años a partir de los 20-25 años y control ginecológico anual a partir de los 30-35 años, que incluye un exámen pélvico, el análisis de una biopsia de endometrio

y una ecografía transvaginal. Además, se recomienda la realización de una histerectomía y una salpingooforectomía bilateral profiláctica en todas aquellas mujeres mayores de 40 años que hayan cumplido sus deseos reproductivos (Giardiello et al., 2014; Valle et al., 2019; Vasen et al., 2013; Yurgelun & Hampel, 2018). Por otro lado, se ha demostrado que el consumo diario de aspirina tiene un efecto quimiopreventivo en los pacientes con SL y reduce el riesgo a desarrollar tumores colorrectales (Burn et al., 2011). Por último, debido a que la deficiencia en la reparación comporta la acumulación de mutaciones *frameshift* en los microsatélites y que se pueden predecir tanto estas mutaciones como los neoptidos inmunogénicos asociados a ellas, se han empezado a realizar ensayos clínicos para determinar la viabilidad de una prevención del cáncer mediante inmunoterapia, por ejemplo utilizando vacunas con estos neopéptidos (Biller et al., 2019). En la **Tabla 7** se resumen las principales estrategias de seguimiento para los diferentes tumores asociados a SL.

Tabla 7. Principales estrategias de seguimiento en el síndrome de Lynch. Adaptado de (Giardiello et al., 2014).

Tipo de cáncer	Inicio del seguimiento	Procedimiento
Cáncer colorrectal	A partir de los 20 - 25 años	Colonoscopias cada 1-2 años hasta los 40 años. A partir de entonces, colonoscopias anuales.
Cáncer de endometrio	A partir de los 30 - 35 años	Examen pélvico anual y análisis de biopsia de endometrio.
Cáncer de ovario	A partir de los 30 - 35 años	Ecografía transvaginal anual.
Cáncer gástrico	A partir de los 30 - 35 años	Esofagogastroduodenoscopia con biopsia del antro cada 2 - 3 años y tratamiento de la infección de *Helicobacter pylori* si se detecta.
Cáncer de intestino delgado	-	No existen datos suficientes que sugieran el beneficio del seguimiento.
Cáncer del tracto urinario	A partir de los 30 - 35 años	Urianálisis anual.
Cáncer del tracto biliar	-	No existen datos suficientes que sugieran el beneficio del seguimiento.
Tumor cerebral	-	No existen datos suficientes que sugieran el beneficio del seguimiento.

3.2 Diagnóstico y manejo del síndrome CMMRD

3.2.1 Criterios de selección y algoritmo diagnóstico del síndrome CMMRD

Dada la agresividad del síndrome CMMRD, es necesario efectuar el diagnóstico rápido de estos pacientes para poder adaptar el tratamiento del cáncer a la deficiencia de reparación e iniciar los protocolos de seguimiento adecuados, sobre todo teniendo en cuenta el alto riesgo a desarrollar una segunda neoplasia. La diversidad clínica y el hecho de que el historial clínico familiar suele ser no informativo, ya que los progenitores acostumbran a ser demasiado jóvenes como para haber desarrollado ya tumores del espectro SL, no es fácil identificar criterios de selección claros. El consorcio europeo C4CMMRD propuso en 2014 un sistema de puntuaciones para identificar los pacientes candidatos a padecer CMMRD basado en unas pocas características altamente asociadas al síndrome. Este sistema asigna una cantidad variable de puntos (de 1 a 3) a cada característica sugestiva de CMMRD y, si se alcanzan los 3 puntos, existe una alta probabilidad de padecer el síndrome y se recomienda iniciar el consejo genético y los análisis genéticos pertinentes para confirmarlo **(Tabla 8)** (Wimmer et al., 2014). Además, se recomienda aplicar este sistema de puntuaciones sobre todos aquellos pacientes con cáncer colorrectal infantil y a todos los casos con tumores hematológicos de células T y gliomas malignos, en especial a los que provengan de regiones o etnias asociadas a alta consanguinidad (Ripperger & Schlegelberger, 2016; Tabori et al., 2017; Wimmer et al., 2014).

La identificación de mutaciones bialélicas patogénicas en alguno de los cuatro genes MMR es la única manera de confirmar y validar el diagnóstico de CMMRD; sin embargo, dado que muchas veces se necesita un diagnóstico rápido para efectuar el tratamiento, que el análisis de ciertos genes como *PMS2* puede llegar a ser complicado o que se identifican variantes de significado desconocido que imposibilitan el diagnóstico, se han propuesto aproximaciones alternativas para confirmar con alta fiabilidad la sospecha de CMMRD. Estos análisis son la evaluación de la expresión de las proteínas reparadoras en el tejido normal y

tumoral del paciente mediante IHC, el análisis de MSI y, más recientemente, la determinación de la tasa de mutación de los tumores si se trata de un tumor cerebral. La pérdida de expresión de alguna de las proteínas reparadoras y/o la presencia de MSI en el tejido normal y tumoral, así como una tasa de mutación superior a 100 mutaciones/Mb en tumores cerebrales, son patognomónicas. Es más, realizar estos análisis antes del análisis de los genes MMR permite, en algunos casos, dirigir el estudio genético y, así, ahorrar tiempo y recursos (Bakry et al., 2014; C. Durno et al., 2017; Tabori et al., 2017; Wimmer et al., 2014).

Tabla 8. Sistema de puntuaciones para la detección de individuos con sospecha clínica de padecer CMMRD. Adaptado de (Wimmer et al., 2014).

Criterios para testar la condición CMMRD en un paciente enfermo de cáncer	≥3 puntos
Neoplasias y lesiones premalignas: una es obligatoria; si hay más de una en el paciente, se agregan los puntos	
Tumor del espectro del SL a una edad de <25 años	3 puntos
Múltiples adenomas intestinales a la edad de <25 años y ausencia de mutaciones en *APC* / *MUTYH* o un solo adenoma con displasia de alto grado a la edad de <25 años	3 puntos
Glioma de grado III o IV (según la Organización Mundial de la Salud) a la edad de <25 años	2 puntos
Linfoma no Hodgkin de células T o tumor neuroectodérmico primitivo supratentorial a la edad de <18 años	2 puntos
Cualquier neoplasia antes de los 18 años	1 punto
Características adicionales: opcional; si hay más de una de las siguientes, se agregan los puntos	
Características reminiscentes a NF1 y/o ≥2 alteraciones de la piel hiper- o hipopigmentadas de Ø>1 cm en el paciente	2 puntos
Diagnóstico de SL en un pariente de primer o segundo grado	2 puntos
Tumor del espectro del SL antes de los 60 años en un pariente de primer, segundo o tercer grado	1 punto
Un hermano con un tumor del espectro del SL, un glioma de alto grado, un linfoma no Hodgkin de células T o un tumor neuroectodérmico primitivo supratentorial	2 puntos
Un hermano con cualquier tipo de cáncer infantil	1 punto
Presencia de múltiples pilomatricomas en el paciente	2 puntos
Presencia de un pilomatricoma en el paciente	1 punto
Agénesis del cuerpo calloso	1 punto
Padres consanguíneos	1 punto
Descenso o ausencia de las inmunoglobulinas IgG2, IgG4 o IgA	1 punto

3.2.2 Asesoramiento genético del síndrome CMMRD

El diagnóstico confirmado del síndrome CMMRD en un individuo supone el diagnóstico del síndrome de Lynch en los progenitores y, en caso de que haya más descendencia, ésta tendrá un 25% de probabilidades de sufrir también el síndrome CMMRD y un 50% de padecer SL, por lo que es sumamente importante que el asesoramiento genético se inicie antes que las pruebas diagnósticas y acompañe a la familia a lo largo de todo el proceso para asegurar que ésta comprenda todas las implicaciones del diagnóstico. Además, teniendo en cuenta el impacto de este síndrome, se recomienda que el apoyo psicológico se ofrezca de forma sistemática a las familias (Durno et al., 2017; Wimmer et al., 2014).

Sin embargo, el problema ético aparece cuando se trata de diagnosticar a un niño sano, ya sea porque tiene un hermano con CMMRD, con características de NF1 o que sume 3 o más en el sistema de puntuación (Suerink, Ripperger, et al., 2018). Entre los beneficios de un diagnóstico precoz se encuentra la oportunidad de iniciar los protocolos de seguimiento y prevención antes del desarrollo del tumor, pudiendo detectarlo en estadios muy iniciales y mejorando el tratamiento; el diagnóstico del SL en los progenitores y el resto de familiares, iniciando también el seguimiento en ellos, y la posibilidad de tomar medidas reproductivas en caso de que los padres deseen tener más hijos. Por el contrario, los potenciales daños de realizar un estudio genético en un niño sano son los riesgos asociados a un intenso seguimiento cuando su eficacia no ha sido comprobada, además de la existencia de fenotipos atenuados; el riesgo de identificar una variante de significado desconocido, que imposibilita el manejo clínico y puede llegar a inducir estrés y ansiedad en el paciente y sus familiares, y, por último, la posibilidad de diagnosticar SL en un menor, creándole una carga psicológica innecesaria a su edad (Suerink, Ripperger, et al., 2018). Será la unidad de consejo genético quién determine cuál es el mejor procedimiento a seguir (Suerink, Potjer, et al., 2018; Suerink, Ripperger, et al., 2018).

3.2.3 Estrategias de prevención y seguimiento

Las estrategias de seguimiento actuales se basan en los datos disponibles sobre la frecuencia de los tumores según la edad extraídos de los pocos casos CMMRD que hay reportados en la literatura, por lo que su eficacia es desconocida y requiere de estudios prospectivos para su evaluación. Teniendo esto en cuenta, se recomienda empezar el seguimiento a partir de los 2 años realizando una resonancia magnética cada 6 o 12 meses para el cribado de los tumores cerebrales, mientras que se aconsejan colonoscopias anuales a partir de los 8 años para el cribado del cáncer colorrectal. Para los tumores hematológicos, la recomendación es contaje de células sanguíneas cada 6 meses a partir del año de vida (Bakry et al., 2014; Durno et al., 2017; Tabori et al., 2017; Vasen et al., 2014). En la **Tabla 9** se resumen las principales medidas de seguimiento para cada tipo de tumor.

Tabla 9. Principales estrategias de seguimiento en el síndrome CMMRD. Adaptado de (Vasen et al., 2014)**.**

Tipo de cáncer	Inicio del seguimiento	Procedimiento
Tumor cerebral	A partir de los 2 años	Imagen por resonancia magnética, 1 cada 6 - 12 meses.
Tumores del tracto digestivo		
Cáncer de intestino delgado	A partir de los 10 años	Cápsula endoscopia, esofagogastroduodenoscopia anual. Se realiza a la vez que la colonoscopia y bajo anestesia general.
Cáncer colorrectal	A partir de los 8 años	Ileocolonoscopia anual
Tumores hematológicos		
Linfoma no Hodgkin y otros linfomas	A partir de 1 año	Examen clínico cada 6 meses. Opcional: ecografía abdominales cada 6 meses.
Leucemia	A partir de 1 año	Contaje de células de la sangre cada 6 meses.
Tumores asociados a SL	A partir de los 20 años	Anual: examen ginecológico, ecografía transvaginal, biopsia de endometrio, citología de la orina y urianálisis.
Todos los tumores	Se aconseja a los padres y pacientes que contacten con su médico ante cualquier signo o síntoma inusual, independientemente de cuando se hizo la última revisión.	

Por otro lado, recientemente se ha empezado a debatir si, al igual que ocurre en el SL, la aspirina podría tener un efecto quimipreventivo en los paciente con CMMRD. No

obstante, aún no se disponen de suficientes datos para su recomendación (Leenders et al., 2018).

3.3 Técnicas de diagnóstico

3.3.1 Cribado molecular de los tumores

La deficiencia reparadora es la característica molecular de los tumores asociados a los síndromes de predisposición a cáncer debidos a defectos en el sistema MMR. Así, para la identificación de la deficiencia MMR se realiza, como ya se ha ido indicando en los apartados anteriores, el análisis de MSI y de la expresión de las proteínas reparadoras mediante IHC sobre el tumor, y, para descartar que se trate de un caso esporádico, se analiza la metilación del promotor de *MLH1* y/o las mutaciones en *BRAF* en caso de pérdida de la proteína MLH1.

Análisis de la inestabilidad de microsatélites

La MSI se define como cambios en el patrón de repeticiones que presenta un microsatélite al analizarlo en el tumor frente al patrón que presenta en el tejido normal del mismo paciente **(Figura 12**) (Yurgelun & Hampel, 2018). Históricamente, la inestabilidad se ha determinado mediante el análisis por PCR convencional de un panel de microsatélites llamado panel de Bethesda. Este panel está formado por 5 microsatélites: 2 mononucleótidos (BAT26 y BAT25) y 3 dinucleótidos (D2S123, D5S346 y D17S250). Cuando más de 2 microsatélites muestran inestabilidad, el tumor es clasificado como altamente inestable (MSI-H), mientras que si solo un marcador da positivo para la inestabilidad, el tumor es clasificado con baja inestabilidad (MSI-L). Por el contrario, si ningún marcador ha resultado ser inestable, el tumor es definido como estable (MSS, de sus siglas en inglés *MicroSatellite Stable*) (Boland et al., 1998; Umar et al., 2004). Se ha demostrado que los marcadores dinucleótidos muestran una sensibilidad y especificidad menor que los

mononucleótidos, así que posteriormente se propuso un panel de 5 microsatélites mononucleótidos cuasimonomórficos para analizar la MSI en tumores. Este panel está formado por los marcadores BAT26, BAT25, NR21, NR22 y NR24, y, aparte de mostrar mayor sensibilidad para detectar MSI (especialmente en tumores de portadores de mutaciones en *MSH6*), permite el estudio de los tumores colorrectales y endometriales sin necesidad de tener tejido normal aperado (Suraweera et al., 2002).

Más recientemente, gracias al continuo desarrollo de las tecnologías de NGS, se han descrito algunos estudios que determinan la inestabilidad de los tumores mediante secuenciación de exomas y/o paneles NGS de microsatélites (Niu et al., 2014; Salipante et al., 2014; Zhu et al., 2018).

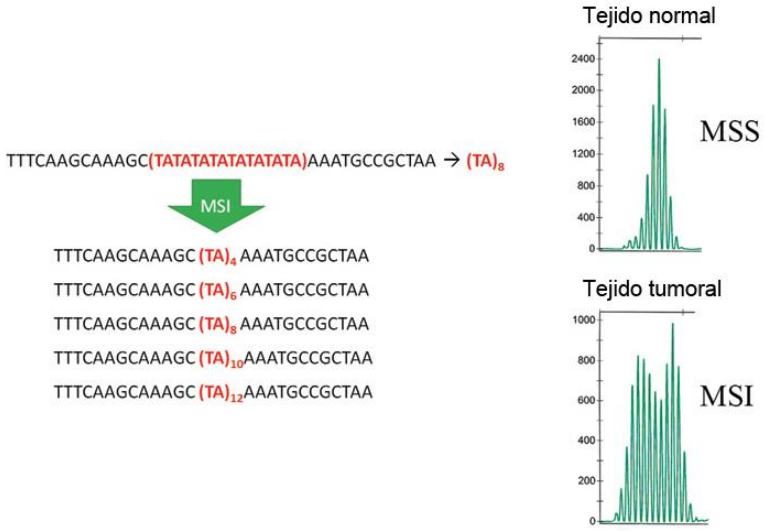


Figura 12. Esquema de la inestabilidad de microsatélites. Izquierda, ilustración de la variación de la longitud de un microsatélite dinucleótido de timina y adenima a causa de la MSI. Derecha, electroferograma de un microsatélite en tejido normal y tumoral del mismo individuo. MSI, inestabilidad de microsatélites; MSS, estabilidad de microsatélites. Adaptado de (Lázaro et al., 2018).

Inmunohistoquímica de las proteínas MMR

El análisis de la expresión de las proteínas MMR consiste en la tinción inmunohistoquímica con anticuerpos anti MLH1, MSH2, MSH6 y PMS2, en la muestra de tejido tumoral y normal adyacente, normalmente conservado en parafina (FFPE, de sus

siglas en inglés *Formalin-Fixed Paraffin-Embedded*), para comprobar la posible pérdida de expresión. Además de proporcionar evidencia de SL o CMMRD, la pérdida de expresión de alguna de las proteínas reparadoras servirá de guía para los posteriores estudios genéticos en línea germinal, ya que el que patrón de tinción es característico de cada alteración molecular subyacente. Así, ante la pérdida de expresión de MLH1 y PMS2, el defecto suele encontrarse en *MLH1*, ya que PMS2 requiere de MLH1 para estabilizarse y sin él no es estable; por el contra, si sólo se observa pérdida de PMS2, se puede suponer que es en este gen dónde se encuentra la mutación. Gracias a la redundancia del sistema MMR, en ausencia de PMS2 MLH1 se puede unir a otras proteínas y permanecer estable. Lo mismo se observa con MSH2 y MSH6 (Gruber, 2006).

Tanto el análisis de MSI como la IHC de las proteínas MMR presentan una sensibilidad y especificidad similar, aunque la IHC puede dar en algunas ocasiones falsos positivos a causa de que el anticuerpo hibride con un fragmento de la proteína truncada (Vasen et al., 2007) y de que no todas las mutaciones patogénicas causan pérdida de la expresión proteica, especialmente las que producen cambios *missense* (Pineda et al., 2010). Por lo tanto, ambos enfoques son complementarios y necesarios para determinar la sospecha de SL o CMMRD.

En relación con los pacientes CMMRD, cabe recordar que también se observa pérdida de la expresión de las proteínas reparadoras en tejido normal y que dicha pérdida en tejido no tumoral no se ha de interpretar como un fallo de la tinción. Es más, en aquellos tumores en los que es prácticamente imposible evaluar la expresión mediante IHC, como es el caso de los tumores hematológicos, se recomienda testar directamente la expresión proteica mediante IHC en tejido normal para determinar si se trata de un caso con sospecha de CMMRD (Wimmer et al., 2014).

Análisis del promotor de *MLH1* y mutaciones en *BRAF*

Debido a que un porcentaje considerable de tumores esporádicos presenta pérdida de la expresión proteica de MLH1 y PMS2 a causa de la hipermetilación somática del promotor de *MLH1*, analizar el estado de metilación de este promotor representa un buen método de preselección para descartar los casos esporádicos y así reducir el coste del análisis genético (Cenin et al., 2018; Gausachs et al., 2012; Perez-Carbonell et al., 2010). Existen diferentes técnicas para evaluar la metilación, como la pirosecuenciación, el análisis de la curva de fusión específica de metilación (o MS-MCA, de sus siglas en inglés *Methylation Specific-Melting Curve Analysis*) o la amplificación dependiente de la ligación de sondas multiplexadas específicas de metilación (MS-MLPA, de sus siglas en inglés *Methylation-Specific Multiplex Ligation-dependent Probe Amplification*), entre otras, y algunas requieren de un paso previo de desaminación de las citosinas no metiladas, convirtiéndolas en uracilos, como es el caso de la pirosecuenciación o el MS-MCA (Kurdyukov & Bullock, 2016). En la literatura se han reportado diferentes rangos de especificidad según la técnica, el punto de corte y los criterios utilizados para la selección de los casos (Moreira et al., 2015; Newton et al., 2014). Alternativamente, en los tumores colorrectales se ha propuesto el cribado de la mutación *BRAF* V600E como un método válido para detectar los casos esporádicos, ya que se encuentra presente en el 69-78% de los casos con metilación del promotor de *MLH1* y únicamente de forma ocasional en tumores Lynch (Adar et al., 2017; Deng et al., 2004; Gausachs et al., 2012; Palomaki et al., 2009). Es importante recordar que, aunque se da de forma poco frecuente, la presencia de hipermetilación constitucional del promotor de *MLH1* también debe considerarse como una posible causa de SL.

3.3.2 Técnicas de diagnóstico molecular

Tras analizar el tumor, aquellos pacientes con MSI y/o pérdida de expresión de alguna de las proteínas reparadoras en el tumor, y en los que se haya descartado que se trate de un caso esporádico, pasarán a analizarse a nivel genético en la línea germinal para los genes MMR. Existen diversas metodologías para analizar estos genes, desde la

secuenciación Sanger de hace unos años hasta la secuenciación NGS de determinados genes o de todo el exoma o genoma, técnicas que ya se están implementando en los laboratorios de diagnóstico genético. La identificación de una o dos mutaciones patogénicas germinales en los genes MMR conducirá al diagnóstico de SL o CMMRD, respectivamente.

Detección de mutaciones puntuales mediante secuenciación Sanger

La secuenciación Sanger todavía está considerada como la técnica de referencia para el análisis mutacional de cualquier gen. Dicho análisis debe incluir el estudio de toda la región codificante del gen, sus límites intrón-exón y, aunque no de forma obligada, también se recomienda secuenciar la región del promotor ya que se han descrito mutaciones patogénicas causales (Q. Liu, Thompson, et al., 2016).

Cabe remarcar que la secuenciación de *PMS2* es más compleja debido a la presencia de pseudogenes y que un análisis convencional no podrá distinguir entre las variantes localizadas en el gen o en el pseudogen. Para resolver este problema, una de las estrategias más utilizada es la realización de una PCR *long-range* para amplificar de manera específica el gen antes de la secuenciación sanger (van der Klift et al., 2010; Vaughn et al., 2010).

Detección de grandes reordenamientos

Aparte de las mutaciones puntuales, el defecto en el gen también puede deberse a deleciones o duplicaciones de uno o varios exones del gen. Por ejemplo, cuando se trata del análisis de *MSH2*, es importante realizar también un estudio de grandes reordenamientos en *EPCAM*, ya que se ubica a 5’ del promotor de *MSH2* y deleciones de los últimos exones de *EPCAM* provocan la hipermetilación del promotor de *MSH2*, provocando un efecto clínico similar al de las mutaciones patogénicas en *MSH2* (Kovacs et al., 2009; Ligtenberg et al., 2009).

Dado que por secuenciación Sanger no se pueden detectar estos grandes reordenamientos, se han desarrollado diversos métodos para ello. El más comúnmente

utilizado es la amplificación dependiente de ligación de sondas multiplexadas (MLPA, de sus siglas en inglés *Multiplex Ligation-dependent Probe Amplification*), en el que solo se amplifican las regiones del DNA hibridadas con la sondas. Su visualización en un electroferograma permitirá su cuantificación relativa. Alternativamente a este método, las variaciones a nivel de número de copia del DNA también se pueden estudiar por PCR cuantitativa a tiempo real o por PCR múltiple cuantitativa de fragmentos fluorescentes cortos. Con todo, este tipo de mutaciones representan alrededor del 10% de todas las mutaciones reportadas en los genes MMR (Lázaro et al., 2018).

Secuenciación por *Next-Generation Sequencing*

La secuenciación mediante NGS ha ido substituyendo a la secuenciación Sanger a lo largo de los últimos años gracias a sus precios cada vez más competitivos y por ofrecer un mejor rendimiento sin perder la calidad del proceso (Feliubadalo et al., 2013). La ventaja que presenta la NGS por encima de las otras metodologías es que permite secuenciar simultáneamente centenares o miles de genes a una mayor profundidad y a un precio y plazo de tiempo competitivos (Rohlin et al., 2017; E. M. Stoffel et al., 2018; Susswein et al., 2016; Yurgelun et al., 2015). Según la región que se quiera analizar, la secuenciación mediante NGS se divide en 3 categorías diferentes: secuenciación de todo el genoma del individuo (*whole-genome sequencing*), secuenciación de todo el exoma del individuo (*whole-exome sequencing*) o secuenciación de un grupo concreto de genes o regiones del genoma. A estos conjuntos de genes o regiones se les denomina “paneles de genes” y presentan la ventaja de que es el propio investigador, o la casa comercial, quién diseña y escoge qué genes o regiones del genoma se integrarán en el panel. Por esta razón, los paneles de genes son la opción más utilizada en la práctica clínica ya que permiten escoger y analizar a la vez todos los genes asociados a un fenotipo clínico concreto (Lázaro et al., 2018).

En la actualidad existen multitud de metodologías diferentes para realizar la NGS, cada una de ellas con un protocolo de realización y un proceso químico propios y una

plataforma de secuenciación específica. Las plataformas más utilizadas hoy en día son las desarrolladas por Illumina, en las que la secuenciación se realiza mediante la síntesis de DNA utilizando dideoxinucleótidos terminadores reversibles fluorescentes (Bentley et al., 2008), y las plataformas de Ion Torrent, en la que los nucleótidos naturales se van añadiendo de forma secuencial y lo que se detecta es el protón que se libera al unirse dos nucleótidos (Rothberg et al., 2011).

En cuanto a los diferentes métodos para realizar la NGS, en general se pueden distinguir dos métodos conceptuales básicos para capturar la región de interés del DNA y enriquecerla: los métodos basados en captura por hibridación (en inglés *capture hybridization-based method*) o los basados en amplicones (en inglés *amplicon-based method*). Los **métodos basados en captura por hibridación** normalmente empiezan con la fragmentación del DNA por sonicación, seguido de la captura de la región de interés mediante la hibridación con una sonda complementaria. Ejemplos de esta aproximación serían las tecnologías *SureSelect* (Agilent Technologies, Inc.), *SeqCap* (Roche NimbleGen, Inc.) o *Nextera* (Illumina, Inc.). Por otro lado, **los métodos basados en amplicones** utilizan oligonucleótidos complementarios a los extremos 5' y 3' de la región de interés como cebadores o *primers* de PCR y, así, amplifican únicamente la región que se quiere analizar. En el caso de la tecnología *Haloplex* (Agilent Technologies, Inc.), primero hay un paso de fragmentación enzimática antes de hibridar con los oligonucleótidos, que serán complementarios a los extremos de los fragmentos generados; en cambio, en la tecnología *Ion AmpliSeq* (Ion Torrent, Thermo Fisher Scientific), no existe este primer paso de fragmentación **(Figura 13)**. Cada método presenta unas ventajas y limitaciones diferentes, por lo que un método será más adecuado que otro, en función de lo que se quiera analizar y los requerimientos propios del laboratorio (Samorodnitsky et al., 2015).

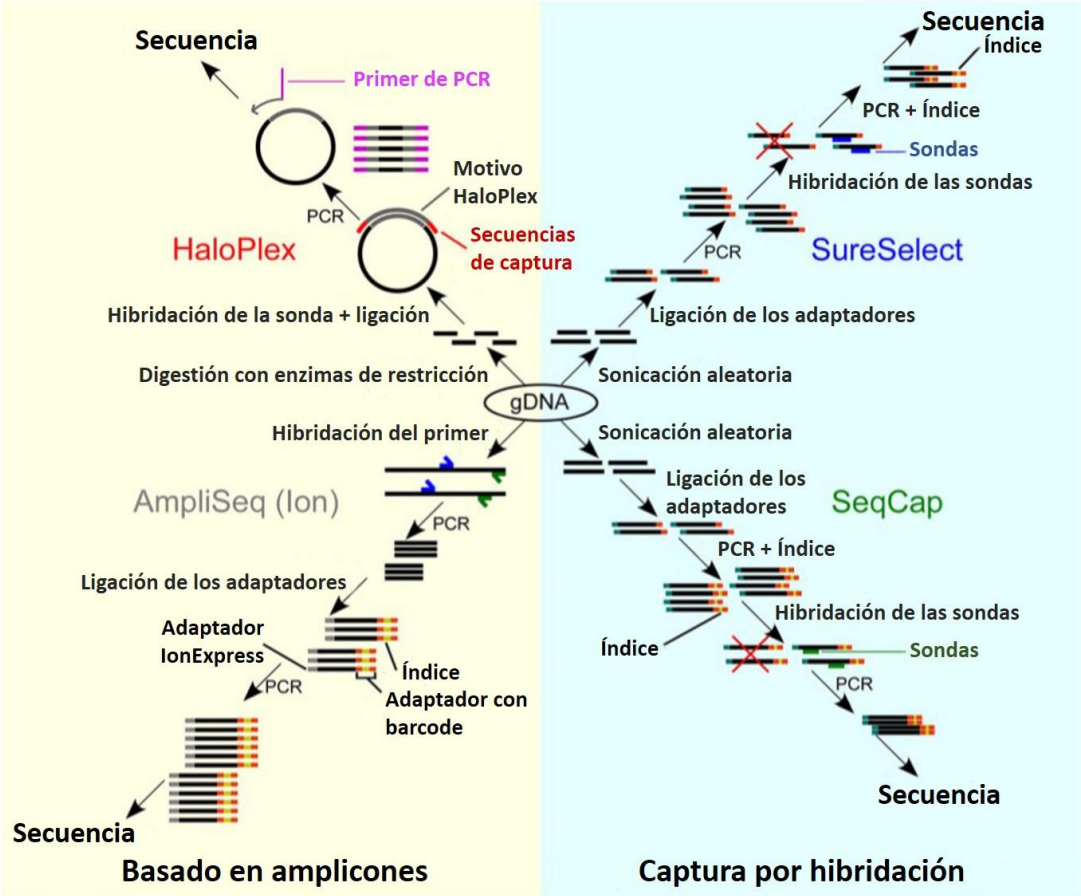


Figura 13. Esquema de diferentes ejemplos de metologías para realizar la NGS. SureSelect y SeqCap se clasifican como métodos basados en la captura por hibridación debido a que fragmentan el DNA mediante sonicación y utilizan oligonucleótidos para hibridar y capturar las regiones de interés. Por el contrario, HaloPlex y AmpliSeq se clasifican como métodos basados en amplicones porque usan los oligonucleótidos como *primers* de PCR para generar los amplicones. Adaptado de (Samorodnitsky et al., 2015).

En los últimos años se ha empezado a incluir en todas las diferentes tecnologías de NGS el uso de los llamados *molecular barcodes* - secuencias de oligonucleótidos degenerados que se unen a moléculas individuales de DNA - que permiten identificar post-amplificación el origen de cada nueva molécula y con ello minimizar el impacto de los errores de PCR y secuenciación **(Figura 14)** (Peng et al., 2015; Schmitt et al., 2012). Gracias a su utilización, se ha incrementado la sensibilidad y especificidad de la NGS, así como la

capacidad de detectar variantes ultra-raras a muy baja frecuencia (Kou et al., 2016; MacConaill et al., 2018; Salk et al., 2018).

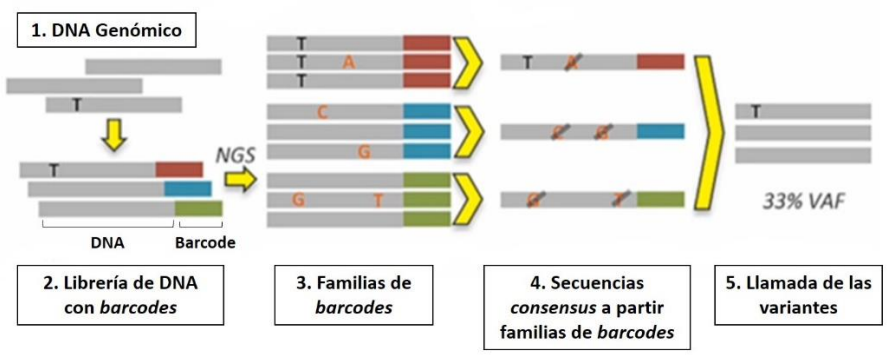


Figura 14. Funcionamiento de los molecular barcodes para minimizar el impacto de los errores de PCR y secuenciación. Adaptado del seminario online de Agilent Technologies, Inc., "Ultra High Sensitivity Sequencing using Targeted Molecular Barcodes" impartido por el Dr. Eric Duncavage.

A pesar de las múltiples ventajas que presenta la NGS, también tiene limitaciones entre las que cabe destacar: (i) la necesidad de validar los hallazgos mediante secuenciación Sanger, sobre todo cuando el resultado es diagnóstico; (ii) las dificultades a la hora de capturar ciertas regiones del DNA, por ejemplo regiones ricas en GC, como sería el caso de promotores y primeros exones de algunos genes; y (iii) el análisis de las secuencias repetitivas como los microsatélites o, en el caso de los genes MMR, el gen *PMS2* al no poder diferenciarlo fácilmente de sus pseudogenes (Lázaro et al., 2018).

4. Clasificación de variantes en genes MMR

4.1 Tipos de evidencias utilizadas para la clasificación de variantes

La identificación de mutaciones patogénicas en alguno de los genes reparadores permite el diagnóstico de los diferentes síndromes asociados a deficiencia del sistema MMR y el manejo clínico de portadores y familiares. Alrededor del 30% de las variantes que se encuentran en la rutina de diagnóstico son VUS, lo que impide el diagnóstico de portadores y familiares (Aceto et al., 2009; Peltomaki, 2016; Thompson et al., 2014). Como consecuencia de la implementación del cribado universal de los tumores colorrectales y de endometrio para detectar SL y el uso de paneles NGS multigénicos en la rutina diagnóstica, está aumentado considerablemente la detección de variantes de este tipo (Howarth et al., 2015; Rohlin et al., 2017; Susswein et al., 2016; Yurgelun et al., 2015).

Para poder categorizar una variante, ya sea como patogénica o como benigna, es necesario integrar de forma rigurosa múltiples líneas de evidencias tanto cualitativas como cuantitativas. Dichas evidencias pueden ser de tres tipos: evidencias basadas en la secuencia de DNA, evidencias clínico-moleculares y evidencias funcionales. Además, estas evidencias se pueden integrar en cálculos multifactoriales que darán un valor cuantitativo de patogenicidad que puede ayudar a objetivar el peso relativo de las diferentes evidencias.

4.1.1 Evidencias basadas en la secuencia de DNA

Las evidencias basadas en la secuencia de DNA son aquellas evidencias derivadas de la naturaleza de la variante y su localización dentro de la secuencia génica. Por ejemplo, para una proteína donde la pérdida de función es el mecanismo de patogenicidad, los cambios que truncan un dominio funcional (por ejemplo cambios *nonsense* y *frameshift)* o los cambios que alteran las posiciones canónicas para el procesamiento del tránscrito de RNA o

splicing tendrán una alta probabilidad de ser patogénicos. Por contra, las consecuencias de los cambios *missense* o en posiciones intrónicas serán más difíciles de predecir.

4.1.2 Evidencias clínico-moleculares

Dentro de las evidencias clínico-moleculares quedan englobadas la cosegregación de la mutación con la enfermedad en la familia, la frecuencia poblacional de dicha mutación, las características moleculares del tumor y la co-ocurrencia con otras mutaciones.

Cosegregación: Una elevada cosegregación de la variante en los individuos afectos de cáncer indica una mayor probabilidad de ser la causa de la enfermedad; no obstante, este análisis no siempre es fácil debido a la penetrancia incompleta en el caso del SL, la posibilidad de fenocopias y el tamaño de las familias analizadas.

Frecuencia alélica: Si la variante se encuentra a una frecuencia elevada en la población general, se le atribuye una baja probabilidad de patogenicidad. Por ejemplo, se dan como neutras todas aquellas variantes con una frecuencia poblacional mayor al 1% (Goldgar et al., 2008). Sin embargo, hay que tener en cuenta que ciertas mutaciones patogénicas pueden darse a una elevada frecuencia debido a un efecto fundador o porque se localicen puntos calientes (*hotspots*) de mutación. Actualmente existen diversas bases de datos públicas con este tipo de información en series amplias de pacientes, como la *Genome Aggregation Database* (gnomAD, http://gnomad.broadinstitute.org/) o el *1000 Genomes Project* (http://www.internationalgenome.org/).

Características moleculares del tumor: Como se ha comentado anteriormente, la pérdida de expresión de las proteínas MMR y la presencia de MSI son indicadores de la deficiencia reparadora. La presencia de deficiencia MMR en los tumores de portadores de una determinada variante será sugestiva de patogenicidad.

Co-ocurrencia: Para aquellas variantes que se detectan junto a una variante patogénica en el mismo gen, el estudio de la fase alélica es sumamente importante porque este dato, en combinación con la información del fenotipo del individuo portador, puede ayudar en su clasificación. Por ejemplo, en un individuo con fenotipo de SL, la presencia de una mutación patogénica en *trans* (en el otro alelo) indicaría una baja probabilidad de patogenicidad de la variante en estudio, ya que de ser patogénica también, se asociaría a fenotipo de CMMRD.

4.1.3 Evidencias funcionales

Los estudios funcionales que incluyen las predicciones *in silico*, los estudios funcionales a nivel de RNA y los estudios a nivel de proteína, son clave para evaluar la patogenicidad de una variante ya que interrogan de forma directa el impacto de ésta a diferentes niveles moleculares (Heinen & Rasmussen, 2012; Peña-Diaz & Rasmussen, 2016). Sin embargo, la relevancia clínica de los resultados experimentales es a menudo limitada, ya que para poder respaldar firmemente la patogenicidad de una variante y aplicar estos resultados al diagnóstico, es necesario que dichos ensayos estén bien establecidos y sean robustos. Además, la falta de estandarización puede provocar resultados contradictorios entre diferentes laboratorios (Hoskinson et al., 2017; Richards et al., 2015). En 2014, el comité para la interpretación de variantes del InSiGHT propuso un diagrama de flujo para facilitar la interpretación de los resultados funcionales **(Figura 15)** (Thompson et al., 2014).

Predicciones *in silico*

Hoy en día existen numerosos algoritmos que predicen *in silico* el posible impacto funcional de una variante a nivel de RNA o de proteína (Richards et al., 2015). Debido a su moderada sensibilidad y especificidad, estas evidencias deben considerarse como evidencias adicionales a otras y es recomendable el uso de diferentes programas, considerando la coincidencia en el resultado de los mismos como un elemento importante. Por otro lado, los resultados de estos predictores también se encuentran actualmente integrados en los cálculos multifactoriales como una probabilidad de patogenicidad previa

a la observación de otras evidencias. Estos predictores también son útiles a la hora de priorizar qué ensayos funcionales se realizarán.

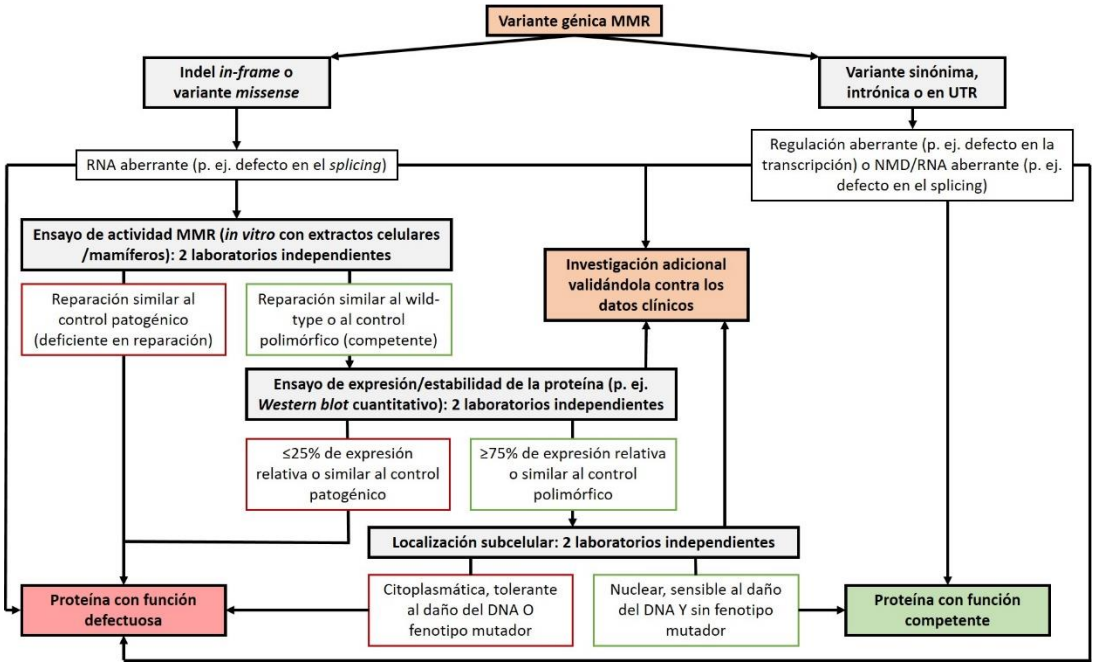


Figura 15. Diagrama de flujo para la interpretación de resultados funcionales. MMR, mismatch repair; NMD, nonsense-mediated decay. Adaptado de (Thompson et al., 2014).

Estudios funcionales a nivel de RNA

Los estudios funcionales a nivel de RNA evalúan, entre otras características, el impacto de una variante sobre el *splicing* o procesamiento del RNA mensajero y si existe una expresión alélica diferencial (ASE, del inglés "*Allele-Specific Expression*") de éste. Siempre que sea posible, es preferible realizar los ensayos a partir de RNA de linfocitos del paciente y habiendo inhibido antes el sistema *Nonsense-Mediated mRNA Decay* (NMD) mediante la suministración de puromicina o cicloheximida al cultivo celular. El sistema NMD se encarga de degradar los tránscritos de RNA mensajero aberrantes y su inhibición permite observar todos los tránscritos aberrantes generados por la mutación sin que estos queden enmascarados por la acción del NMD (Wimmer & Wernstedt, 2014). Sin embargo, dado que muchos laboratorios no disponen de este tipo de muestra, se han desarrollado ensayos de

RNA *in vitro* como los *minigenes*, que presentan una buena correlación con los ensayos en linfocitos (Gaildrat et al., 2010; Thompson et al., 2015; Tournier et al., 2008; van der Klift et al., 2015).

Estudios funcionales a nivel de proteína

Los principales estudios a nivel de proteína que se recomiendan realizar para las variantes MMR son los dirigidos a determinar (i) la expresión y estabilidad de la proteína, (ii) la localización subcelular de ésta y (iii) su capacidad reparadora, siendo esta última característica la que ha sido considerada como clave para evaluar la patogenicidad (Thompson et al., 2014). No obstante, también se han desarrollado ensayos que estudian otras funciones de la proteína MMR, como la interacción con otras proteínas reparadoras, la unión al DNA o la unión e hidrólisis del ATP (Heinen & Rasmussen, 2012).

i) Expresión y estabilidad de la proteína: Los estudios de expresión y estabilidad de la proteína se realizan, de manera mayoritaria, mediante la técnica de *Western blot*. Para generar la proteína con la variante, normalmente se transfecta de forma transitoria un plásmido de expresión con la variante a estudio en una línea celular deficiente en el gen en cuestión y, días después, se extraen las proteínas para su cuantificación (Borras et al., 2012; Hinrichsen et al., 2013; Kosinski et al., 2010; Takahashi et al., 2007). Sin embargo, la interpretación de los resultados obtenidos no siempre es fácil, ya que tanto la sobreexpresión como la subexpresión de la proteína debido al mismo sistema de transfección pueden influir en los resultados. Es más, teniendo en cuenta que la dosis de proteína también afecta a la capacidad reparadora, estos fenómenos podrían tanto enmascarar una pérdida parcial de la actividad como simularla (Hinrichsen et al., 2013).

ii) Localización subcelular: Para poder realizar su función, es necesario que las proteínas reparadoras se localicen en el núcleo. Para evaluar si la variante está afectando esta localización, se pueden realizar estudios de inmunotinción o fusionar la proteína con un

fluorocromo y analizarla mediante microscopía confocal (Borras et al., 2012; Borras et al., 2013).

iii) Actividad reparadora: Dentro de la gran diversidad de ensayos utilizados para evaluar a nivel funcional las variantes MMR, se proponen aquellos métodos dirigidos a testar la capacidad reparadora los de referencia para estudiar este tipo de variantes (Couch et al., 2008; Thompson et al., 2014). Existen multitud de ensayos para medir la actividad reparadora, pero en general se pueden agrupar en 3 categorías: los ensayos *in vivo* en levadura, los ensayos basados en células de mamíferos o *ex vivo* y los ensayos *in vitro* que utilizan extractos celulares y que, por lo tanto, son independientes de célula (también conocidos como *in vitro cell-free MMR assays* por su terminología en inglés). Actualmente, los métodos más utilizados son los *in vitro* con extractos celulares, ya que presentan ventajas significativas. Estos ensayos consisten en la reconstitución de extractos nucleares de células deficientes en reparación con proteínas reparadoras humanas purificadas o generadas *in vitro* que portan la variante a estudiar y un sustrato a reparar. En función de si el sustrato es reparado o no, se puede determinar si la variante está afectando la capacidad reparadora de la proteína. Esta condición evita que las condiciones fisiológicas de la célula, o que la acumulación de mutaciones en ellas tras múltiples rondas de replicación, afecten al proceso de reparación y falseen los resultados. También, el uso de proteínas humanas permite el estudio de todas las variantes detectadas en los pacientes, y no solo las ubicadas en los dominios conservados en levadura, y la posibilidad de extrapolar los resultados de las modificaciones post-traduccionales y otros aspectos del sistema de reparación (Peña-Diaz & Rasmussen, 2016). Además, recientemente se ha demostrado que este tipo de ensayos son útiles para el diagnóstico de CMMRD a partir de tejido no neoplásico, realizando el ensayo de reparación con proteínas purificadas de los linfocitos del paciente (Shuen et al., 2019). Sin embargo, la solidez de estos ensayos de reparación *in vitro*, punto crítico para su uso clínico, rara vez se ha evaluado.

4.1.4 Cálculos multifactoriales

Muchas de las evidencias previamente mencionadas se han incorporado en la actualidad a algoritmos bayesianos multifactoriales tras ser calibradas para ello. Estos algoritmos se basan en razones de verosimilitud (LR, del inglés "*Likelihood Ratios*") que comparan para cada componente la probabilidad de que se observe un dato asumiendo que la variante es patogénica frente a la probabilidad en caso de que sea neutra. Las LR para cada tipo de evidencia pueden ser combinadas para obtener una probabilidad posterior de patogenicidad. En el caso de los genes MMR, el modelo multifactorial combina el resultado de las predicciones *in silico*, las LR de cosegregación y las LR de las características moleculares de los tumores (MSI y BRAF) (Thompson, Greenblatt, et al., 2013). A pesar de su utilidad para clasificar las variantes desde un punto cuantitativo, estos modelos aún necesitan mejorar para ser más precisos. La inclusión de los resultados de la IHC sobre el tumor o de los resultados de los estudios funcionales deberían ser el siguiente paso para refinar su utilidad.

4.2 Categorías de clasificación

Actualmente existen varios esquemas para la clasificación de variantes genéticas, todos ellos destinados a categorizar de forma estandarizada las variantes según su potencial impacto clínico. El más comúnmente utilizado es el esquema en 5 categorías propuesto por la *International Agency for Research on Cancer* (IARC), que clasifica las variantes en patogénicas (clase 5), probablemente patogénicas (clase 4), de significado desconocido (clase 3), probablemente benignas (clase 2) y benignas (clase 1). Cada categoría puede vincularse a una probabilidad de patogenicidad concreta (datos cuantitativos) y/o a una interpretación de datos cualitativos validados. Estas categorías de clasificación están asociadas a unas recomendaciones clínicas de manejo y vigilancia específicas **(Tabla 10)** (Plon et al., 2011). Así, para las variantes de clase 5 y 4 se recomiendan estudios predictivos en los familiares y un seguimiento de alto riesgo, mientras que para las variantes de clase 1

y 2 se aconseja tratar a los individuos como no portadores de variantes responsables. El problema reside, sin embargo, en las recomendaciones para las variantes de clase 3. Para estas variantes no se recomiendan estudios predictivos y se ofrece un seguimiento a todos los miembros de la familia basado en la historia familiar y otros factores de riesgo, ya que no se puede realizar un seguimiento selectivo en función de ser portador o no de la variante.

Tabla 10. Sistema de clasificación de variantes en 5 clases y recomendaciones de seguimiento. Adaptado de (Plon et al., 2011).

Clase	Probabilidad de patogenicidad	Estudio predictivo en familiares en riesgo	Recomendaciones de seguimiento en familiares portadores
5. Patogénica	>0,99	Sí	Seguimiento de alto riesgo
4. Probablemente patogénica	0,95 - 0,99	Sí *	Seguimiento de alto riesgo
3. De significado incierto	0,05 - 0,949	No *	Basado en historia familiar y otro factores de riesgo
2. Probablemente benigna	0,001 - 0,049	No *	Tratado como individuo sin variantes responsables detectadas para el síndrome
1. Benigna	<0,001	No *	Tratado como individuo sin variantes responsables detectadas para el síndrome

* Se recomienda ampliar el estudio en el probando si existen técnicas adicionales disponibles (p. ej. estudio de grandes reordenamientos).

4.3 Guías de clasificación de variantes

4.4.1 Guías de clasificación gen-específicas

En 2014, el comité para la interpretación de variantes del InSiGHT desarrolló un esquema de clasificación estandarizado específico para variantes en genes MMR utilizando las 5 categorías propuestas por la IARC **(Figura 16)**. Este esquema se basa en múltiples líneas de evidencias, incluyendo datos clínico-moleculares y funcionales, y gracias a su aplicación se pudieron reclasificar dos tercios de las variantes informadas hasta el momento (Thompson et al., 2014). Los criterios InSiGHT para los genes MMR se encuentran disponibles en su página web (https://www.insight-group.org/criteria/) y están sujetos a revisiones periódicas en base a nuevos hallazgos que puedan contribuir al refinamiento de

las reglar de interpretación. Por este motivo, se recomienda su uso para la clasificación de variantes MMR.

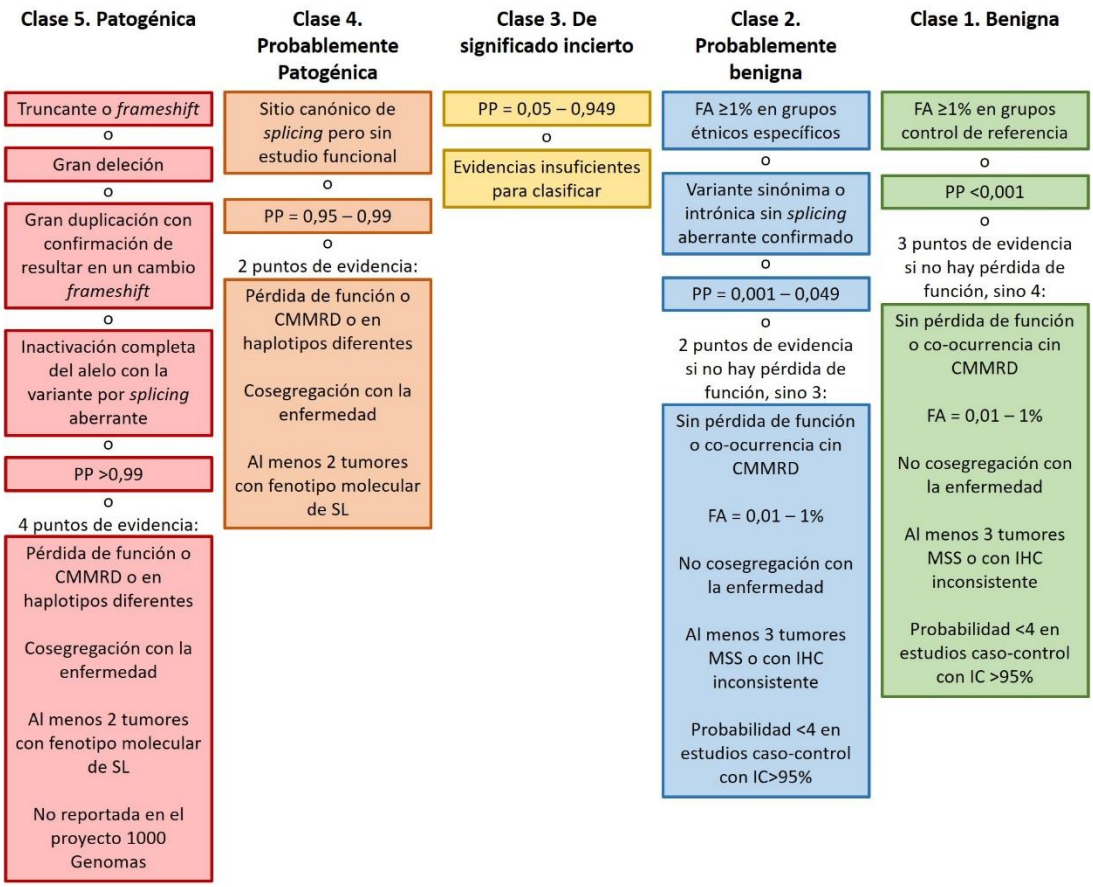


Figura 16. Descripción general de las reglas de clasificación InSiGHT en 5 niveles para los genes MMR. PP, probabilidad de patogenicidad; CMMRD, deficiencia constitucional de reparación de apareamientos erróneos; SL, síndrome de Lynch; FA, frecuencia alélica; MSS, estabilidad de microsatélites; IHC, inmunohistoquímica; IC, intervalo de confianza. Adaptado de (Thompson et al., 2014).

4.4.2 Guías generales de clasificación

En 2015, debido a la ausencia de criterios específicos para algunos genes, la *American College of Medical Genetics and Genomics* (ACMG) junto con la *Association for Molecular Pathology* (AMP) estandarizaron la interpretación clínica de variantes genéticas asociadas a enfermedades mendelianas (Richards et al., 2015). Para ello, establecieron

también 5 niveles de categorización con 28 criterios de clasificación basados principalmente en datos poblacionales, *in silico*, funcionales y de segregación. En la actualidad, estas reglas ACMG-AMP se han empezado a refinar para poderse aplicar de forma específica a algunos genes. Así, encontramos guías específicas basadas en el sistema ACMG-AMP para, por ejemplo, los genes *PTEN* o *CDH1*, causantes del síndrome tumoral hamartomatoso asociado a *PTEN* y cáncer gástrico difuso, respectivamente (K. Lee et al., 2018; Mester et al., 2018; Rivera-Munoz et al., 2018). Para facilitar la aplicación de estas reglas, existen diversas herramientas informáticas gratuitas *online*, como *InterVar* (http://wintervar.wglab.org/) (Q. Li & Wang, 2017) o *Franklin* (https://franklin.genoox.com/).

Por otro lado, la compañía Invitae también ha intentado refinar los criterios iniciales de clasificación de las guías ACMG-AMP a través del sistema Sherloc, que define un total de 33 reglas con 108 refinamientos adicionales, proporcionando un enfoque más coherente y transparente para la clasificación de variantes (Nykamp et al., 2017). Sin embargo, Sherloc es hoy en día poco utilizado y la mayoría de laboratorios siguen refiriéndose a las guías ACMG-AMP cuando no existen guías gen-específicas.

4.4 Bases de datos

Actualmente existen diversas bases de datos globales con acceso libre para el registro e interpretación de variantes. En el caso de las variantes en genes MMR, cabe destacar la base de datos de InSIGHT (http://www.insight-database.org/classifications/), que se focaliza en recopilar y clasificar exclusivamente variantes MMR. Además, la base de InSIGHT se sirve de la *Leiden Open Variation Database* (LOVD), un software de código abierto que puede ser descargado e instalado libremente por cualquier persona u organización, para almacenar y compartir la información clínica de probandos y variantes, así como para clasificarlas y notificárselo a las partes pertinentes. A día de hoy, la base de datos de InSIGHT se encuentra en la LOVDv3 (http://insight-database.org/).

Por otro lado, la base de datos ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), es una recopilación de clasificaciones de diferentes fuentes (contribuciones individuales y paneles de expertos), por lo que, a pesar de su gran utilidad, es necesario interpretar las clasificaciones críticamente. Recientemente, se han recopilado todos los datos agregados a ClinVar en una nueva aplicación web que ofrece diferentes estadísticas a nivel de gen, variante y enfermedad, para así facilitar la interpretación clínica de la variación genética (http://simple-clinvar.broadinstitute.org/) (Perez-Palma et al., 2019).

HIPÓTESIS

El síndrome de Lynch y el síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos son síndromes de predisposición hereditaria al cáncer cuyo diagnóstico molecular viene dado por la identificación de mutaciones patogénicas en los genes MMR en la línea germinal. Sin embargo, se estima que alrededor del 30% de las variantes en genes MMR que se detectan en la rutina diagnóstica son variantes de significado desconocido, lo que impide el diagnóstico temprano del paciente y sus familiares, esencial para establecer recomendaciones terapéuticas y de seguimiento óptimas.

Para establecer la patogenicidad de una variante, es necesario integrar múltiples líneas de evidencia, tanto cualitativas como cuantitativas. Los ensayos funcionales han resultados pueden muy útiles para elucidar el grado de patogenicidad de las variantes MMR. No obstante, la falta de estandarización y protocolos validados dificulta su implementación en el diagnóstico rutinario. Por otro lado, el cálculo multifactorial de probabilidad representa una manera alternativa de determinar de forma cuantitativa la patogenicidad de una variante MMR, aunque precisa de la recopilación de datos clínico-patológicos en un importante número de familiares.

La presencia de inestabilidad de microsatélites es característica de los tumores asociados a los síndromes asociados a deficiencia del sistema MMR. Asimismo, la inestabilidad de microsatélites también se ha detectado en bajas proporciones en tejido normal de estos individuos mediante técnicas muy laboriosas o poco sensibles.

Por estas razones, la presente tesis doctoral sostiene las siguientes hipótesis:

1) La validación del ensayo *in vitro* de actividad reparadora (*in vitro cell-free MMR assay*), así como el establecimiento de protocolos estandarizados para generar sus diferentes componentes, mejorará la caracterización funcional de las variantes MMR y, en consecuencia, el manejo clínico de los individuos con síndrome de Lynch o Deficiencia Constitucional de Reparación de Apareamientos Erróneos. Además, el análisis exhaustivo de

estas variantes mediante la combinación de ensayos funcionales y cálculo multifactorial de probabilidad incrementará el número total de variantes de significado desconocido que se reclasificarán.

2) La evaluación mediante técnicas de alta sensibilidad de la inestabilidad de microsatélites en muestras biológicas no tumorales de individuos con síndromes asociados a deficiencia del sistema MMR mejorará su diagnóstico, especialmente en individuos portadores de variantes de significado desconocido.

OBJETIVOS

Objetivo principal

Esta tesis doctoral tiene por objetivo principal la mejora del diagnóstico de los síndromes de predisposición hereditaria al cáncer asociados a una deficiencia del sistema de reparación de apareamientos erróneos mediante el desarrollo y puesta a punto de nuevas aproximaciones moleculares.

Objetivos específicos

1) Mejorar la evaluación de la patogenicidad de las variantes de significado desconocido en genes reparadores de apareamientos erróneos mediante la implementación de estudios funcionales exhaustivos, el uso de modelos multifactoriales, y la validación del ensayo *in vitro* de actividad reparadora.

2) Desarrollar una nueva metodología para la detección con alta sensibilidad de inestabilidad de microsatélites en tejido normal de portadores de mutaciones germinales en los genes del sistema de reparación de apareamientos erróneos.

RESULTADOS

La sección de **Resultados** de esta tesis doctoral consta de dos artículos publicados y uno pendiente de aceptación por parte de la revista a la que se ha enviado. Además, en la sección **Anexos** se adjuntan otras tres publicaciones en las que la estudiante de doctorado ha colaborado y es coautora (Anexo I. Otras publicaciones).

ARTÍCULO 1

Elucidating the clinical significance of two PMS2 missense variants coexisting in a family fulfilling hereditary cancer criteria. **Familial Cancer, 2017**.
DOI: 10.1007/s10689-017-9981-1.

ARTÍCULO 2

Validation of an in vitro mismatch repair assay used in the functional characterization of mismatch repair variants. **Manuscrito enviado.**

ARTÍCULO 3

High-sensitivity microsatellite instability assessment for the detection of mismatch repair defects in normal tissue of biallelic germline mismatch repair mutation carriers. **Journal of Medical Genetics, 2019.**
DOI: 10.1136/jmedgenet-2019-106272

ARTÍCULO 1

Elucidating the clinical significance of two *PMS2* missense variants coexisting in a family fulfilling hereditary cancer criteria

Maribel González-Acosta, Jesús del Valle, Matilde Navarro, Bryony A. Thompson, Sílvia Iglesias, Xavier Sanjuan, María José Paúles, Natàlia Padilla, Anna Fernández, Raquel Cuesta, Àlex Teulé, Guido Plotz, Juan Cadiñanos, Xavier de la Cruz, Francesc Balaguer, Conxi Lázaro, Marta Pineda*, Gabriel Capellá*.
* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Familial Cancer, 2017. DOI: 10.1007/s10689-017-9981-1

RESUMEN:

En este trabajo se identificaron *en trans* dos VUS en el gen *PMS2*, c.2149G>A (p.V717M) y c.2444C>T (p.S815L), en un mismo individuo diagnosticado de cáncer colorrectal a edad temprana y que pertenecía a una familia que cumplía clínica criterios clínicos de cáncer hereditario. Para determinar la relevancia clínica de las dos variantes, se utilizaron los datos clínico-patológicos, el cálculo multifactorial de probabilidad y los resultados de los estudios funcionales.

El cálculo multifactorial de probabilidad, basado en la cosegregación de la variante con la enfermedad y las características tumorales, clasificó la variante c.2444C>T como patogénica, lo que fue corroborado por los estudios funcionales que demostraron una actividad reparadora alterada de la variante, asociada a una disminución de la expresión de la proteína. Por el contrario, la variante c.2149G>A mostró competencia reparadora y estabilidad de la proteína. Estos resultados, sumados a la expresión conservada de PMS2 en tejido normal y la ausencia de inestabilidad de microsatélites basal en sangre del paciente portador de las dos variantes, descartaron un diagnóstico de CMMRD.

En conclusión, el uso de estrategias que integran la información funcional con los datos clínico-patológicos mejoró la interpretación clínica de las variantes germinales detectadas en los genes MMR, aspecto clave para el apropiado manejo clínico de los síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema MMR.

Familial Cancer
DOI 10.1007/s10689-017-9981-1

ORIGINAL ARTICLE

Elucidating the clinical significance of two *PMS2* missense variants coexisting in a family fulfilling hereditary cancer criteria

Maribel González-Acosta[1] · Jesús del Valle[1] · Matilde Navarro[1] · Bryony A. Thompson[2,3] · Sílvia Iglesias[1] · Xavier Sanjuan[4] · María José Paúles[4] · Natàlia Padilla[5] · Anna Fernández[1] · Raquel Cuesta[1] · Àlex Teulé[1] · Guido Plotz[6] · Juan Cadiñanos[7] · Xavier de la Cruz[5,8] · Francesc Balaguer[9] · Conxi Lázaro[1] · Marta Pineda[1] · Gabriel Capellá[1]

Abstract The clinical spectrum of germline mismatch repair (MMR) gene variants continues increasing, encompassing Lynch syndrome, Constitutional MMR Deficiency (CMMRD), and the recently reported *MSH3*-associated polyposis. Genetic diagnosis of these hereditary cancer syndromes is often hampered by the presence of variants of unknown significance (VUS) and overlapping phenotypes. Two *PMS2* VUS, c.2149G>A (p.V717M) and c.2444C>T (p.S815L), were identified in trans in one individual diagnosed with early-onset colorectal cancer (CRC) who belonged to a family fulfilling clinical criteria for hereditary cancer. Clinico-pathological data, multifactorial likelihood calculations and functional analyses were used to refine their clinical significance. Likelihood analysis based on cosegregation and tumor data classified the c.2444C>T variant as pathogenic, which was supported by impaired MMR activity associated with diminished protein expression in functional assays. Conversely, the c.2149G>A variant displayed MMR proficiency and protein stability. These results, in addition to the conserved PMS2 expression in normal tissues and the absence of germline microsatellite instability (gMSI) in the biallelic carrier ruled out a CMMRD diagnosis. The use of comprehensive strategies, including functional and clinico-pathological information, is mandatory to improve the clinical interpretation of naturally occurring MMR variants. This is critical for appropriate clinical management of cancer syndromes associated to MMR gene mutations.

Marta Pineda and Gabriel Capellá have contributed equally to this work and share senior authorship.

Electronic supplementary material The online version of this article (doi:10.1007/s10689-017-9981-1) contains supplementary material, which is available to authorized users.

✉ Gabriel Capellá
gcapella@iconcologia.net

1 Hereditary Cancer Program, Catalan Institute of Oncology (ICO), IDIBELL and CIBERONC, Av. Gran Via de l'Hospitalet, 199-203, 08908 Hospitalet de Llobregat (Barcelona), Spain

2 Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA

3 Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Melbourne, Australia

4 Pathology Department, Hospital Universitari de Bellvitge, IDIBELL, Hospitalet de Llobregat (Barcelona), Spain

5 Research Unit in Translational Bioinformatics, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

6 Medical Clinic 1, Johann Wolfgang Goethe-University, Frankfurt, Germany

7 Instituto de Medicina Oncológica y Molecular de Asturias (IMOMA), Oviedo, Spain

8 ICREA, Barcelona, Spain

9 Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

Published online: 01 April 2017

Keywords Constitutional mismatch repair deficiency · Lynch syndrome · Mismatch repair · *PMS2* gene · Variant of unknown significance

Abbreviations

CMMRD	Constitutional mismatch repair deficiency
CRC	Colorectal cancer
gMSI	Germline microsatellite instability
LR	Likelihood ratio
LS	Lynch syndrome
MMR	Mismatch repair
MSI	Microsatellite instability
VUS	Variant of unknown significance

Introduction

In humans, germline mutations in five of the mismatch repair (MMR) genes [1] can result in three currently identified hereditary cancer syndromes: Lynch syndrome (LS), Constitutional MMR deficiency (CMMRD) and the recently reported recessive polyposis syndrome associated with biallelic mutations in *MSH3* [2–4]. The identification of these inherited conditions has important consequences for clinical management, allowing targeted preventive measures in mutation carriers.

LS (OMIM #120435), caused by monoallelic pathogenic germline (epi)mutations in *MLH1, MSH2, MSH6* and *PMS2*, is characterized by early adult-onset CRC and an increased risk of other associated tumors [3, 5]. CMMRD (OMIM #276300), caused by biallelic mutations in the same MMR genes, is a more severe syndrome characterized by the development of café-au-lait skin lesions, hematological malignancies and brain and colorectal cancer, most often during childhood and adolescence [4, 5]. Recently, reported cases harboring biallelic *MSH3* mutations were diagnosed with colorectal and duodenal adenomas, CRC, gastric cancer and astrocytoma in the adulthood [2]. Interestingly, overlapping phenotypes have been described between the MMR gene-associated syndromes [2, 6–12]. At the somatic level, their associated tumors display MSI and/or loss of MMR protein expression. Of note, MMR deficiency is also evident in non-neoplastic tissues from CMMRD patients [413].

PMS2 monoallelic mutations account for a small portion of LS cases (6–15%) [14, 15], though its contribution might be underestimated due to its lower penetrance [16, 17]. In contrast, biallelic *PMS2* mutations account for over 50% of CMMRD cases [4]. Several strategies have been developed to refine the mutational analysis of *PMS2* [18, 19], hampered by the presence of highly homologous pseudogenes [20]. Advances in the mutational analysis of *PMS2* has lead to the increasing identification of variants of unknown significance (VUS), which preclude appropriate clinical management of carriers and their relatives [5]. In spite of the existence of a standardized scheme based on multiple lines of evidence, up to 30% of the MMR variants in the LOVD locus-specific database remain as VUS [21]. Specifically, *PMS2* VUS accounts for 22% of the *PMS2* reported variants. Their functional characterization is increasingly needed as current classification guidelines hamper their classification as class 4/5 mainly due to their low penetrance [21, 22].

Here we report the coexistence of two *PMS2* variants of unknown significance, c.2149G>A (p.V717M) and c.2444C>T (p.S815L), in a women affected by early-onset CRC. A comprehensive analysis has lead to the appropriate management of the entire family.

Materials and methods

Patients and germline mutational analysis

A Spanish family fulfilling Hereditary Breast and Ovarian Cancer syndrome and Amsterdam II criteria was identified. Clinico-pathological data from affected individuals were collected, including age at diagnosis, tumor location, MSI testing and immunohistochemistry of MMR proteins. Mutational analysis of the *BRCA1/2* and *PMS2* genes was performed as described in Supplementary Methods. Informed consent was obtained from all individuals and the ethics committee of Bellvitge University Hospital approved the study.

PMS2 variant frequency in controls

The variant frequencies were obtained from the NHLBI Exome Sequencing Project (ESP) (http://evs.gs.washington.edu/EVS), 1000Genomes (http://www.1000genomes.org/) and ExAC (http://exac.broadinstitute.org/) databases. Screening of the identified *PMS2* variants in a Spanish control population cohort was previously reported [14].

Multifactorial likelihood and bioinformatic analyses

Multifactorial likelihood analysis was conducted as described [23]. Briefly, multifactorial analysis was based on estimated prior probabilities of pathogenicity and likelihood ratios (LR) for segregation and tumor characteristics (MSI phenotype and recruitment location). The *PMS2* risk estimates from *ten Broeke* et al. [24] were used to calculate the segregation LRs. Risk associated with each identified *PMS2* variant (c.2149G>A and c.2444C>T) has been analyzed separately. Variants were classified according to the 5-class IARC quantitative scheme [25], based on their

posterior probability. In silico analyses were performed to evaluate evolutionary conservation and the impact of *PMS2* variants on splicing, protein function and protein structure. See Supplementary Methods for details.

Functional analyses of *PMS2* variants

The effect of variants on splicing was evaluated using RNA extracted from cultured lymphocytes, in the absence or presence of puromycin (Sigma), from individual IV:3. Total RNA was extracted from cultured lymphocytes and cDNA was synthesized as described [14]. *PMS2* cDNA fragments were amplified using LaTaq polymerase (Takara) in two overlapping fragments and sequenced [26] (see the used primers in Supplementary Table S1). Amplified transcripts from carriers were compared with transcripts from two control cultured lymphocyte samples.

pcDNA3.1_MLH1 and pN1_PMS2 expression plasmids, kindly provided by Dr. Kolodner and Dr. Nyström-Lahti, were used for in vitro MMR assays and expression analyses. The PMS2 variants p.V717M (c.2149G>A) and p.S815L (c.2444C>T) were constructed by site-directed mutagenesis using the QuikChange Site Directed Mutagenesis Kit (Agilent Technologies, La Jolla, CA), according to manufacturer's instructions (Supplementary Table S1). Sanger sequencing was used to verify the presence of the variants. *PMS2* cDNAs containing each variant were subcloned into pN1_PMS2-wild-type plasmid. In addition, three control plasmids, PMS2 p.D70N (c.208G>A), p.P470S (c.1408C>T) and p.S46I (c.137G>T) were used as controls, as previously described [14].

Transfection of HEK293T cells (deficient for endogenous MLH1 and PMS2) was carried out as described [27]. In brief, HEK293T cells were transfected at 30–40% confluence with MLH1 and PMS2 expression plasmids (0.5 µg/ml each) and 0.05 µg/ml of pGFP, as a transfection control, using 20 µl/ml of the cationic polymer polyethylenimine (Polysciences, Warrington, Pennsylvania, USA). After 48 h, cells were prepared for protein extraction and cytometer analysis. MLH1 and PMS2 expression levels were examined by Western blot as described [14]. Alpha-actin expression was assessed in parallel and used as loading control. All experiments were performed in triplicate. See Supplementary Methods for details.

MMR assays were performed as described [27]. 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 by the value of a wild-type protein that had been expressed, processed, and tested in parallel. Experiments were performed in triplicate. See Supplementary Methods for details.

Germline MSI (gMSI) analysis

PCR amplification of the dinucleotide microsatellites D17S791, D2S123, and D17S250 was performed as reported [13]. The gMSI ratios were determined by dividing the height of an allele's trailing "stutter" peak ($n+1$) by the height of the allele's major peak (n). See Supplementary Methods for details.

Statistical analysis

Significant differences between groups were analyzed using the non-parametric Mann–Whitney U test for quantitative data. All reported P values are two sided, and $P<0.05$ was considered significant. All calculations were performed using SPSS 19.0 software (IBM).

Results

Two *PMS2* germline missense variants, c.2149G>A [p.(V717M)] and c.2444C>T [p.(S815L)], were identified in a patient diagnosed with CRC at age 41. The tumor was MSI and PMS2 loss of expression was exclusively present in neoplastic cells (individual IV:3, Fig. 1 and Supplementary Fig. S1) [28]. Transcript analysis demonstrated the identified *PMS2* variants were located in trans

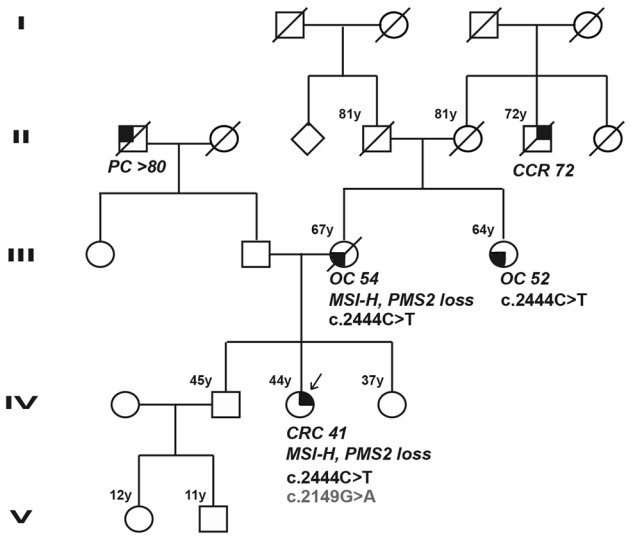


Fig. 1 Pedigree of the reported family. Current age (or age at death) and the carrier status of the variants *PMS2* c.2444C>T (in *black*) or c.2149G>A (in *grey*) are indicated above and below the individual's symbol, respectively. Tumor characteristics are indicated in *italics* below the individual's symbol. An *arrow* indicates the proband for PMS2 analysis. Tumor types are represented as *black sectors* inside an individual's symbol: *top right CRC* colorectal cancer, *top left PC* pancreatic cancer, *bottom left, OC* ovarian cancer, *MSI-H* microsatellite instability high

(Supplementary Fig. S2). Moreover, these variants were rare in control populations: c.2149G>A variant was identified at minor allele frequency <1% in public databases and in our Spanish control cohort (Table 1) [14], and the c.2444C>T variant was described in one individual in the control population and reported in a CRC patient [29]. According to the InSiGHT rules both variants were considered of unknown clinical significance (IARC class 3) [21].

Ad hoc dermatological evaluation identified a café-au-lait macula on proband's left leg with a size of 3.5 × 2.2 cm. Her mother and maternal aunt were previously diagnosed with ovarian cancer at ages 54 and 52 respectively (Fig. 1). Of note, the mother was previously tested negative for germline *BRCA1/2* mutations. A maternal granduncle was affected by CRC at age 72.

A comprehensive study was undertaken to elucidate the pathogenicity of the identified *PMS2* variants. Testing of both *PMS2* variants in affected individuals III:3 and III:4 only identified the c.2444C>T variant (Fig. 1). The ovarian tumor from III:3 also showed MSI and loss of PMS2 expression. The presence of *PMS2* c.2444C>T in three affected first-degree relatives resulted in a segregation likelihood ratio (LR) of 1.83:1, leading to a posterior probability of pathogenicity of 0.993 and therefore being classified as pathogenic (Fig. 1; Table 1). The probability of the c.2149G>A was 0.7965, remaining uncertain (Table 1).

No aberrant transcripts were detected in the RNA analysis of biallelic c.[2149G>A; 2444C>T] carrier lymphocytes, in line with RNA in silico predictions (Supplementary Fig. S3 and Supplementary Table S2) and the absence of splicing alterations reported in a c.2444C>T variant carrier [29]. At the protein level, transfection of MLH1 and PMS2 p.S815L variant in HEK293T cells resulted in diminished PMS2 and MLH1 protein expression (18.53 ± 9.25 and 15.53 ± 8.25% of the MLH1/PMS2 wildtype level, respectively). The p.S815L variant showed impaired MMR activity (10.11 ± 7.14% of the wild-type level) in in vitro complementation assays using the same protein extracts (Fig. 2a, b respectively). Accordingly, in silico predictions labeled p.S815L as deleterious (Supplementary Table S2). In contrast, p.V717M did not affect protein expression and was MMR proficient (Fig. 2a, b). While both variants mapped at relatively external locations in the MLH1 interaction domain, S815 has more residue–residue interactions than V717 (Fig. 2c) and affects a more conserved residue (Supplementary Fig. S4). Of note, the p.S815L variant is predicted not to affect the standard geometry of the Zn-binding site (Supplementary Fig. S5).

Finally, the gMSI ratios in blood DNA in the biallelic carrier for D2S123, D17S250 and D17S791 (0.030 ± 0.001, 0.041 ± 0.001 and 0.092 ± 0.005) were in the range of the two healthy controls analyzed [13], ruling out CMMRD (Supplementary Fig. S6 and Supplementary Table S3).

Table 1 Result of the multifactorial likelihood analysis of the identified *PMS2* variants

Gene	Variant	Protein	Reference SNP ID	Frecuency in controls (%) (our cohort*/ ESP-EA / ExAC /1000G)	Patient ID	Cancer	MSI	IHC	Prior probability of pathogenicity	Prior used	MSI LR	Tumor characteristics LR	Bayes	Segregation LR	Odds for causality	Posterior odds	Posterior probability of pathogenicity	Classification
PMS2	c.2149G>A	p.V717M	rs201671325	0.54/0.02/0.09/0.02	IV:3	CRC 41y**	MSI-H	PMS2 loss	0.594556783	0.59456	8.66	8.66	0.3083	0.3083	2.669878	3.91520689	0.796549764	Class 3 (uncertain)
PMS2	c.2444C>T	p.S815L	rs587779338	NR/NR/0.0008/NR	IV:3	CRC 41y**	MSI-H	PMS2 loss	0.935302404	0.9	8.66	8.66	1.8252	1.8252	15.806232	142.256088	0.993019494	Class 5 (pathogenic)
					III:3	OC 54y	MSI-H	PMS2 loss			–	–						
					III:4	OC 52y	–	–			–	–						

CRC colorectal cancer, *OC*, ovarian cancer, *MSI-H* microsatellite instability high, *LR* likelihood ratio, *NR* Not reported

*Reported in Borras et al. [14]

**Germline *PMS2* c.2149G>A and c.2444C>T variants were found in trans in patient IV:3, affected by CRC

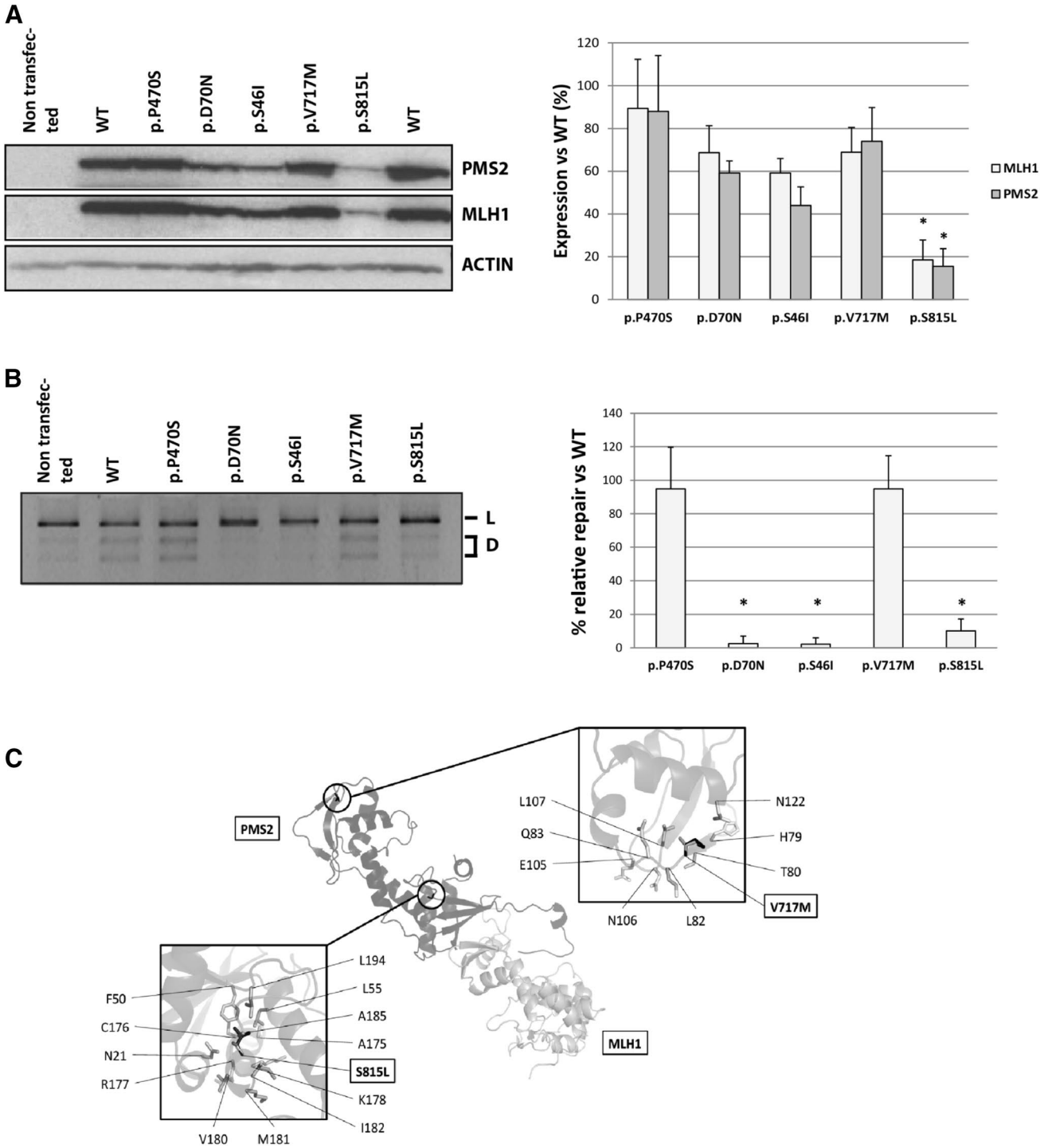


Fig. 2 Protein expression levels of PMS2 p.V717M and p.S815L variants, in vitro mismatch repair activity and location in the PMS2-MLH1 structure. **a** Western-blot analysis of MLH1 and PMS2. Quantification of MLH1 or PMS2 is shown in *light grey* and *grey columns*, respectively. **b** Agarose gel showing digestion products of MMR assay. *D* double-digested vector DNA, *L* linearized vector DNA. Quantification of repair levels of PMS2 variants in direct comparison to PMS2 wild-type is shown. Statistically significant differences with the wild-type group are indicated (*, $P < 0.05$). The PMS2 variants p.D70N and p.S46I (deficient in MMR activity and proficient in MLH1 and PMS2 expression levels) and the neutral variant p.P470S (proficient in MMR activity and expression) were used as controls [14] to confirm the reliability of the technique. **c** Location and three-dimensional neighborhood of the variants. The PMS2 (*dark grey*)-MLH1 (*light grey*) model is shown at the *center* of the figure. Note that for simplicity both variants are shown in the same structure, although they are in trans. The two *boxes* provide a close view of the neighboring residues (interatomic distance ≤ 5 Å) around V717 and S815. At the variant loci, the wild-type and mutant residues are shown in *grey* and *black*, respectively

Discussion

Our study presents a comprehensive assessment of two *PMS2* VUS identified in an individual diagnosed with an early-onset CRC. The *PMS2* c.2444C>T variant demonstrated decreased MLH1/PMS2 protein expression and impaired MMR activity in in vitro functional assays. Accordingly, the variant has been classified as pathogenic using multifactorial likelihood calculations, which enabled the diagnosis of LS in monoallelic c.2444C>T carriers from the reported family. Interestingly, the MMR deficiency of *PMS2* c.2444C>T variant has also been demonstrated in a recent report [30] using a different MMR assay approach based on the production of the variant protein in vitro previous to the MMR complementation assay. Of note, the confirmation in different laboratories of the MMR deficiency of a given variant is mandatory to demonstrate abrogated protein function according to the current classification guidelines [21].

Although the presence of a single café-au-lait macula has a high prevalence in the population [31], the presence of the additional *PMS2* c.2149G>A variant located in trans in the proband (IV:3) in addition to the early-onset CRC did not rule out the possibility of CMMRD. Six CMMRD cases with mild phenotype—with an age of onset of the first tumor ≥ 30 year-old—have been reported [6–8]. Most of them are carriers of likely hypomorphic MMR variants and, therefore, may represent an intermediate phenotype between CMMRD and LS [4]. The functional proficiency in in vitro testing of c.2149G>A variant, the preserved expression of PMS2 in normal tissue and the absence of gMSI detection in blood, most likely rules out CMMRD in our case. Recently, functional assays using lymphoblastoid cell lines have been also proposed to confirm CMMRD diagnoses [32]. Unfortunately, this testing was not possible due to the unavailability of samples. Consequently, the possibility of c.2149G>A being a hypomorphic allele, although unlikely, cannot be fully discarded.

Here we show the relevance of assays demonstrating MMR deficiency, either by impaired MMR activity or decreased protein expression, in providing strong evidence supporting the pathogenicity of a given MMR variant [22]. The information obtained from functional analysis has not been widely utilized in the classification of *PMS2* variants as only a limited number of *PMS2* VUS have been analyzed at the mRNA [14, 29, 33] and protein level [14, 34] in human cell models. At a time when next generation sequencing is being routinely implemented in diagnostic laboratories and the detection of multiple VUS in the same or distinct cancer genes in the same patient increases, the need for robust assays for functional VUS characterization becomes more relevant [35]. The identification of novel phenotypes associated with MMR compound heterozygotes further highlights the importance of appropriate variant classification, especially for *PMS2* gene.

The molecular diagnosis of LS allows for the appropriate management of patients and their families, particularly with regard to clinical follow-up of carriers, with surveillance colonoscopies starting at the age of 20–25 [3, 5]. The recent recommendation of less intense cancer surveillance in *PMS2* monoallelic carriers based on its lower penetrance is controversial [24, 36, 37]. In fact, an important proportion of *PMS2* mutation carriers develop cancer before the age of 45 [36], as the case reported here. Moreover, due to phenotype overlapping with CMMRD, the search for a second germline *PMS2* mutation should be considered in patients with early onset LS-associated PMS2-deficient tumors.

In conclusion, the molecular diagnosis of cancer syndromes associated to MMR gene mutations is hampered when overlapping phenotypes and the identification of VUS coincide. This report illustrates the utility of in-depth characterization of naturally occurring variants, including functional and clinico-pathological analyses, in order to improve the clinical interpretation of genetic data.

Acknowledgements This work was funded by the Spanish Ministry of Economy and Competitiveness and co-funded by FEDER funds- a way to build Europe-(SAF2012-33636 and SAF2015-68016); the Scientific Foundation Asociación Española Contra el Cáncer; the Government of Catalonia (2014SGR338), Fundación Mutua Madrileña (AP114252013) and Red Temática de Investigación Cooperativa en Cáncer (RTICC RD12/0036/0031 and RD12/0036/0008). MG-A was supported by grants AP114252013, RD12/0036/0031 and the Scientific Foundation Asociación Española Contra el Cáncer. BAT is an NHMRC Early Career Fellow. We are indebted to the patients and the members of the Hereditary Cancer Genetic Counseling Units. We also thank Benjamin Puliafito for the English revision of the manuscript and valuable remarks.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Kunkel TA, Erie DA (2015) Eukaryotic mismatch repair in relation to DNA replication. Annu Rev Genet 49:291–313
2. Adam R, Spier I, Zhao B et al (2016) Exome sequencing identifies biallelic MSH3 germline mutations as a recessive

subtype of colorectal adenomatous polyposis. Am J Hum Genet 99(2):337–351
3. Vasen HF, Tomlinson I, Castells A (2015) Clinical management of hereditary colorectal cancer syndromes. Nat Rev Gastroenterol Hepatol 12(2):88–97
4. Wimmer K, Kratz CP, Vasen HF et al (2014) Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'care for CMMRD' (C4CMMRD). J Med Genet 51(6):355–365
5. Sijmons RH, Hofstra RM (2016) Review: clinical aspects of hereditary DNA Mismatch repair gene mutations. DNA Repair 38:155–162
6. Kets CM, Hoogerbrugge N, van Krieken JH, Goossens M, Brunner HG, Ligtenberg MJ (2009) Compound heterozygosity for two MSH2 mutations suggests mild consequences of the initiation codon variant c.1A > G of MSH2. Eur J Hum Genet 17(2):159–164
7. Aronson M, Gallinger S, Cohen Z et al (2016) Gastrointestinal findings in the largest series of patients with hereditary biallelic mismatch repair deficiency syndrome: report from the International Consortium. Am J Gastroenterol 111(2):275–284
8. Li L, Hamel N, Baker K et al (2015) A homozygous PMS2 founder mutation with an attenuated constitutional mismatch repair deficiency phenotype. J Med Genet 52(5):348–352
9. Bodas A, Perez-Segura P, Maluenda C, Caldes T, Olivera E, Diaz-Rubio E (2008) Lynch syndrome in a 15-year-old boy. Eur J Pediatr 167(10):1213–1215
10. Durno CA, Sherman PM, Aronson M et al (2015) Phenotypic and genotypic characterisation of biallelic mismatch repair deficiency (BMMR-D) syndrome. Eur J Cancer 51(8):977–983
11. Herkert JC, Niessen RC, Olderode-Berends MJ et al (2011) Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines. Eur J Cancer 47(7):965–982
12. Will O, Carvajal-Carmona LG, Gorman P et al (2007) Homozygous PMS2 deletion causes a severe colorectal cancer and multiple adenoma phenotype without extraintestinal cancer. Gastroenterology 132(2):527–530
13. Ingham D, Diggle CP, Berry I et al (2013) Simple detection of germline microsatellite instability for diagnosis of constitutional mismatch repair cancer syndrome. Hum Mutat 34(6):847–852
14. Borras E, Pineda M, Cadinanos J et al (2013) Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. J Med Genet 50(8):552–563
15. Palomaki GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN (2009) EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. Genet Med 11(1):42–65
16. Moller P, Seppala T, Bernstein I et al. (2015) Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. Gut 66:464–472
17. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW (2015) ACG clinical guideline: genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 110(2):223–262 (**quiz 63**)
18. Clendenning M, Hampel H, LaJeunesse J et al (2006) Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat 27(5):490–495
19. Vaughn CP, Robles J, Swensen JJ et al (2010) Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Hum Mutat 31(5):588–593
20. Ganster C, Wernstedt A, Kehrer-Sawatzki H et al (2010) Functional PMS2 hybrid alleles containing a pseudogene-specific missense variant trace back to a single ancient intrachromosomal recombination event. Hum Mutat 31(5):552–560
21. Thompson BA, Spurdle AB, Plazzer JP et al (2014) Application of a 5-tiered scheme for standardized classification of 2360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 46(2):107–115
22. Heinen CD (2016) Mismatch repair defects and Lynch syndrome: the role of the basic scientist in the battle against cancer. DNA Repair 38:127–134
23. Thompson BA, Goldgar DE, Paterson C et al (2013) 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 34(1):200–209
24. ten Broeke SW, Brohet RM, Tops CM et al (2015) Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 33(4):319–325
25. Plon SE, Eccles DM, Easton D et al (2008) Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29(11):1282–1291
26. Etzler J, Peyrl A, Zatkova A et al (2008) RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. Hum Mutat 29(2):299–305
27. Plotz G, Welsch C, Giron-Monzon L et al (2006) Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. Nucleic Acids Res 34(22):6574–6586
28. Brea-Fernandez AJ, Cameselle-Teijeiro JM, Alenda C et al (2014) High incidence of large deletions in the PMS2 gene in Spanish Lynch syndrome families. Clin Genet 85(6):583–588
29. van der Klift HM, Tops CM, Bik EC et al (2010) Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. Hum Mutat 31(5):578–587
30. van der Klift HM, Mensenkamp AR, Drost M et al. (2016) Comprehensive mutation analysis of PMS2 in a large cohort of probands suspected of lynch syndrome or constitutional mismatch repair deficiency syndrome. Hum Mutat 37:1162–1179
31. Shah KN (2010) The diagnostic and clinical significance of cafe-au-lait macules. Pediatr Clin N Am 57(5): 1131–1153
32. Bodo S, Colas C, Buhard O et al (2015) Diagnosis of constitutional mismatch repair-deficiency syndrome based on microsatellite instability and lymphocyte tolerance to methylating agents. Gastroenterology 149(4):1017–1029
33. van der Klift HM, Jansen AM, van der Steenstraten N et al (2015) Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genom Med 3(4):327–345
34. Drost M, Koppejan H, de Wind N (2013) Inactivation of DNA mismatch repair by variants of uncertain significance in the PMS2 gene. Hum Mutat 34(11):1477–1480
35. Richards S, Aziz N, Bale S et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405–424
36. Goodenberger ML, Thomas BC, Riegert-Johnson D et al (2016) PMS2 monoallelic mutation carriers: the known unknown. Genet Med 18(1):13–19
37. ten Broeke SW, Nielsen M (2015) A PMS2-specific colorectal surveillance guideline. Genet Med 17(8):684

SUPPLEMENTARY METHODS

Germline mutational analysis in *BRCA1/2* and *PMS2* genes

Genomic DNA was extracted from whole blood using the FlexiGene DNA kit (Qiagen). The analysis of point mutations in *BRCA1* (NM_007294.2; NG_005905.2) and *BRCA2* (NM_000059.3; NG_012772.1) genes was performed by D-HPLC. Genomic rearrangements in these genes were analyzed using the multiplex ligation dependent probe amplification (MLPA) commercial kits SALSA P002B BRCA1 and SALSA P045 BRCA2/CHEK2, respectively (MRC-Holland).

Point mutations in *PMS2* (NM_000535.5, NG_008466.1) were analyzed using previously described LR-PCR procedures [1, 2]. In brief, amplicons spanning exons 1–6 (long range amplicon LR1), 6–10 (LR2) and 10–15 (LR3) were generated using LaTaq polymerase (TaKaRa Bio Inc). Fifteen microlitres of each PCR product were diluted in Tris-EDTA buffer up to a final volume of 180 µl. One microlitre of this dilution was used as the template for exon specific PCR using Megamix Double (Microzone limited) and specific primers. PCR products were sequenced using Big Dye Terminator v.3.1 Cycle Sequencing kit (Applied Biosystems) on an Applied Biosystems 3130XL Genetic Analyzer. Rearrangements in *PMS2* gene were analyzed by MLPA using the SALSA P008-B1 PMS2 commercial kit (MRC-Holland). DNA samples from relatives were screened for the two identified *PMS2* variants by LR-PCR and direct Sanger sequencing. The identified *PMS2* changes have been submitted to the Leiden Open Variation Database (LOVD) database (https://LOVD.nl). Variant nomenclature is according to HGVS recommendations (version 2.0) with nucleotide 1 corresponding to the A of the ATG-translation initiation codon.

Immunohistochemistry for DNA mismatch repair proteins

Immunohistochemistry was performed on 3-µm section slides from formalin-fixed, paraffin-embedded tissue, incubating with primary monoclonal antibodies against MLH1 (clone G168-15; BD Pharmingen), MSH2 (clone G219-1129; BD Pharmingen), MSH6 (clone 44; BD Pharmingen) and PMS2 (clone A16-4, BD Pharmingen). Normal positive DNA MMR protein expression was defined as nuclear staining within tumor cells, using adjacent normal non-neoplastic tissue on the same slide as positive internal control. Negative protein expression was defined as complete absence of nuclear staining within tumor cells. Results were confirmed on different sample blocks or slides from the same cancers.

Microsatellite instability analysis in tumor sample

Microsatellite instability (MSI) status was studied using MSI Analysis System v1.2 kit (Promega) following manufacturer recommendations. Briefly, the commercial kit uses five quasi monomorphic mononucleotide markers (BAT25, BAT26, NR21, NR24, and MONO27) for MSI determination and two polymorphic markers (PentaC and PentaD) for PBL / tumor sample matching. Fluorescent PCR products were resolved by capillary electrophoresis using an 3130XL Genetic Analyzer (Applied Biosystems). Tumor samples with two or more of the five microsatellite markers unstable were considered MSI positive.

Clinical data and *in silico* prediction analyses

Reports addressing the analyzed *PMS2* variants were identified using the LOVD, PubMed and Google. DNA sequences containing the identified *PMS2* variants were analyzed using several bioinformatic tools addressed to evaluate its impact at the RNA and protein level, as previously reported [3].

The impact of PMS2 variants was evaluated *in silico* using three standard pathogenicity predictors: SIFT (http://sift.jcvi.org), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2) and Condel (http://bg.upf.edu/fannsdb). Protein stability changes upon mutation were obtained from: PopMuSic (http://dezyme.com/en/Software), CUPSAT (http://cupsat.tu-bs.de), ERIS (http://troll.med.unc.edu/eris/login.php), I-Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) and FoldX 4 (http://foldxsuite.crg.eu).

The structure of the PMS2-MLH1 complex is a model obtained with MODELLER (https://salilab.org/modeller/) version 9.14, default parameters. As a template, we used the structure of yeast PMS1-MLH1 complex (PDB code: 4FMN), where PMS1 and MLH1 have sequence similarities of 48% and 34% with human PMS2 and MLH1, respectively.
PyMOL Molecular Graphics System v1.5.0.4 (Schrödinger, LLC) was used to visualize structures and to create Figure 2C.

Germline microsatellite instability analysis in DNA from peripheral blood lymphocytes

Germline microsatellite instability (gMSI) analysis was performed as described in Ingham *et al.* 2013 [4]. PCR amplification of the dinucleotide microsatellites D17S791, D2S123, and D17S250 was performed (primers detailed in Supplementary Table B.3). PCR products were analyzed on an Applied Biosystems 3130XL Genetic Analyzer using GeneMapper software (Applied Biosystems, Forster City, California, USA). The gMSI ratios were determined by dividing the height of an allele's trailing "stutter" peak (n+1) by the height of the allele's major

peak (*n*). DNA from two healthy control individuals and one CMMRD patient (homozygous carrier of *PMS2* c.24-2A>G, r.24_28del, p.S8Rfs*4; data not shown) were used as controls in the analysis. Experiments were performed in triplicate.

MMR activity assay

MMR assays were performed as described [5]. In short, the reaction was performed in 15 µl total volume with reaction buffer (25mM Tris–HCl pH 7.5, 110 mM KCl, 5 mM MgCl2, 50 µg/ml BSA, 1.5 mM ATP, 0.1 mM each dNTP), 50 ng DNA mismatched substrate pUC19CPDC, 50 µg nuclear extract of HEK293T cells, which are deficient in mismatch repair, and 5 µg protein extract from transfected HEK293T cells. Reactions were incubated at 37 ^{o}C for 15 min and terminated with 25 µl stop-buffer (100 mM EDTA, 10% SDS, 20 mg/ml proteinase K) by an additional incubation for 10 min at 37 ^{o}C. Plasmids were extracted from the reaction mixture by phenol-chloroform extraction and purified by ethanol co-precipitation with tRNA. Subsequent digestion with AseI, EcoRV and RNAse A produced two smaller fragments besides the linearized vector when repair was successful. Restriction digests were separated on 2% agarose gels. Band intensity was quantified using Image Lab Software v.2.0.1 (Bio-Rad). 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 by the value of a wild-type protein that had been expressed, processed and tested in parallel. Assays were performed in triplicate from 3 independent transfection experiments.

MLH1 and PMS2 protein expression analysis

MLH1 and PMS2 expression levels in transfected HEK293T cells were examined by SDS-PAGE, followed by Western blotting analysis with anti-MLH1 (clone G168-15, BD Pharmigen) and anti-PMS2 (clone 16-4, BD Biosciencies) antibodies. Band intensities were quantified using Quantity One v.4.4 (BioRad). Alpha-actin expression was assessed in parallel and used as loading control. Expression of MLH1 and PMS2 was normalized to alpha-actin expression. The relative protein expression was calculated by dividing the normalized protein expression in variant-transfected cells by the expression in wild-type MLH1/PMS2-transfected cells, processed and tested in parallel. Protein expression analyses were performed in triplicate from 3 independent transfection experiments.

REFERENCES

1. C.P. Vaughn, J. Robles, J.J. Swensen, C.E. Miller, E. Lyon, R. Mao, P. Bayrak-Toydemir, W.S. Samowitz, Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes, Human mutation, 31 (2010) 588-593.

2. H.M. van der Klift, C.M. Tops, E.C. Bik, M.W. Boogaard, A.M. Borgstein, K.B. Hansson, M.G. Ausems, E. Gomez Garcia, A. Green, F.J. Hes, L. Izatt, L.P. van Hest, A.M. Alonso, A.H. Vriends, A. Wagner, W.A. van Zelst-Stams, H.F. Vasen, H. Morreau, P. Devilee, J.T. Wijnen, Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients, Human mutation, 31 (2010) 578-587.

3. E. Borras, M. Pineda, J. Cadinanos, J. Del Valle, A. Brieger, I. Hinrichsen, R. Cabanillas, M. Navarro, J. Brunet, X. Sanjuan, E. Musulen, H. van der Klift, C. Lazaro, G. Plotz, I. Blanco, G. Capella, Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants, J Med Genet, 50 (2013) 552-563.

4. D. Ingham, C.P. Diggle, I. Berry, C.A. Bristow, B.E. Hayward, N. Rahman, A.F. Markham, E.G. Sheridan, D.T. Bonthron, I.M. Carr, Simple detection of germline microsatellite instability for diagnosis of constitutional mismatch repair cancer syndrome, Human mutation, 34 (2013) 847-852.

5. G. Plotz, C. Welsch, L. Giron-Monzon, P. Friedhoff, M. Albrecht, A. Piiper, R.M. Biondi, T. Lengauer, S. Zeuzem, J. Raedle, Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair, Nucleic acids research, 34 (2006) 6574-6586.

Supplementary Table S1. Primers used in this study.

Primer name	Sequence	Analysis
PMS2_c.2444_F	5' GAGCCTGCCGGAAGTTGGTGATGATTGGGAC 3'	Site-Directed Mutagenesis
PMS2_c.2444_R	5' GTCCCAATCATCACCAACTTCCGGCAGGCTC 3'	Site-Directed Mutagenesis
PMS2_c.2149_F	5' GCTGCAGCAGCACACCATGCTCCAGGGGCAGAG 3'	Site-Directed Mutagenesis
PMS2_c.2149_R	5' CTCTGCCCCTGGAGCATGGTGTGCTGCTGCAGC 3'	Site-Directed Mutagenesis
RT_PCR1_pms2_F	5' GGATCGGGTGTTGCATC 3'	Subcloning
RT_PCR1_pms2_R	5' CTTTCTCCTGAGAGTCCACATG 3'	Subcloning
RT_PCR2_pms2_F	5' GCAGCCACTGCTGGATGTTGAAG 3'	Subcloning
RT_PCR2_pms2_R	5' GGTTTGAAAAGGTTCTAAGATCAC 3'	Subcloning
pN1_pms2_A_dw	5' GATGCGTGGCAGGTAGAAAT 3'	Subcloning
pN1_pms2_A_up	5' TAGCGCTACCGGACTCAGAT 3'	Subcloning
pN1_pms2_B_dw	5' TATGCAGAGCATCGGAACAG 3'	Subcloning
pN1_pms2_B_up	5' CTTTGTGCACTGAGCGATGT 3'	Subcloning
pN1_pms2_C_dw	5' CAGTGGCTGCTGACTGACAT 3'	Subcloning
pN1_pms2_C_up	5' CCCCTAGTGACTCCGTGTGT 3'	Subcloning
pN1_pms2_D_dw	5' TCCGGTATCTTCCTGGTTTG 3'	Subcloning and splicing analysis
pN1_pms2_D_up	5' AAATGTCAGTCAGCAGCCACT 3'	Subcloning and splicing analysis
pN1_pms2_E_dw	5' GAGGTGCTATGAGCCTCTGC 3'	Subcloning and splicing analysis
pN1_pms2_E_up	5' AGAAAGCGCCTGAAACTGAC 3'	Subcloning and splicing analysis
pN1_pms2_E_2_up	5' GCAAAGTGAAGGGGAACAGA 3'	Subcloning and splicing analysis
pN1_pms2_E_2_dw	5' TGACTGGAGCATTTTCATCG 3'	Subcloning and splicing analysis
pN1_pms2_F_dw	5' AGAAATGACACCCAGGTTGG 3'	Subcloning and splicing analysis
pN1_pms2_Fup	5' TGCCACGGACGAGAAGTATAA 3'	Subcloning and splicing analysis
pN1_pms2_G_dw	5' GGGAGGTGTGGGAGGTTTT 3'	Subcloning and splicing analysis
pN1_pms2_G_up	5' GACCACCCCTGGAACTGTC 3'	Subcloning and splicing analysis
ct-c2149-wt_up	5' GCTGCAGCAGCACACCG 3'	Evaluation of allelic location of PMS2 variants
ct-c2149-A_up	5' GCTGCAGCAGCACACCA 3'	Evaluation of allelic location of PMS2 variants
ct-c2444-wt_dw	5' GCAGTCCCAATCATCACCG 3'	Evaluation of allelic location of PMS2 variants
ct-c2444-T_dw	5' GCAGTCCCAATCATCACCA 3'	Evaluation of allelic location of PMS2 variants
D2S123_up	5' (6-FAM)AAACAGGATGCCTGCCTTTA 3'	gMSI analysis
D2S123_dw	5' GGACTTTCCACCTATGGGAC 3'	gMSI analysis
D17S250_up	5' (ROX)GGAAGAATCAAATAGACAAT 3'	gMSI analysis
D17S250_dw	5' GCTGGCCATATATATATTTAA 3'	gMSI analysis
D17S791_up	5' (TAM)GTTTTCTCCAGTTATTCCCC 3'	gMSI analysis
D17S791_dw	5' GCTCGTCCTTTGGAAGAGTT 3'	gMSI analysis

Supplementary Table S2. Bioinformatic predictions of the *PMS2* variants. A: Result of the in silico predictions at the RNA level. **B:** Result of the in silico predictions at the protein level. Predictions are interpreted as inconclusive when discordant results are obtained from different programs. Abbreviations: SS, splice site; A, consensus acceptor splice site; D, consensus donor splice site; NR, consensus splice site not recognized.

A

			Splice Site Prediction									Enhancer site prediction		
			NNSplice		Spliceport		NetGene2		Softberry		Interpretation	Rescue ESE	ESE finder	Interpretation
Variant	Exon	SS	wild-type	variant	wild-type	variant	wild-type	variant	wild-type	variant				
c.2149G>A; p.V717M	12	A	0.54	0.54	_	_	0.00	0.00	9.72	9.72	**No effect**	No change	2 site destroyed	**Inconclusive**
		D	0.95	0.95	11.624	10.199	0.93	0.93	8.34	8.34				
c.2444C>T; p.S815L	14	A	0.90	0.90	0.5624	0.5624	0.23	0.23	4.88	4.88	**Inconclusive**	1 site created	No change	**Inconclusive**
		D	1.00	0.99	11.519	1.004	0.00	0.00	15.62	14.92				

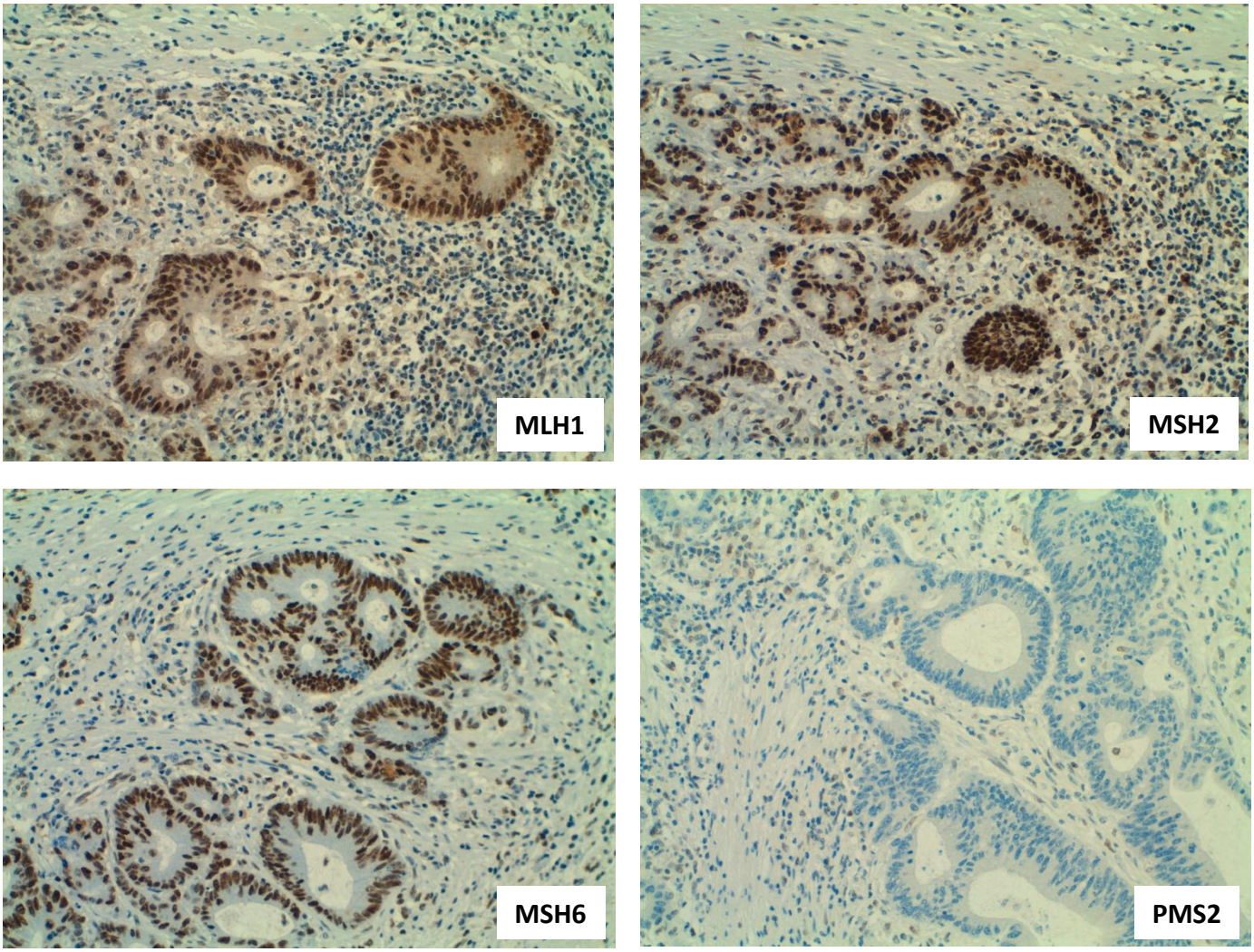
B

		Predicted impact on protein function					Structure prediction					
Variant	Exon	Functional domain	PolyPhen-2 (score)	SIFT (score)	Condel (score)	Interpretation	PoPMuSic	CUPSAT	ERIS	I-MUTANT 3.0	FoldX 4	Interpretation
c.2149G>A; p.V717M	12	Interaction with MLH1	Pr Damaging (0.982)	Tolerated (0.1)	Neutral (0.053)	**Inconclusive**	0.54 kcal/mol (Dstb)	-1.41 kcal/mol (Dstb)	-10.76 kcal/mol (Stb)	-0.66 kcal/mol (Dstb)	-2.89 kcal/mol (Stb)	**Inconclusive**
c.2444C>T; p.S815L	14	Interaction with MLH1	Pr Damaging (1.00)	Damaging (0.00)	Deleterious (1.00)	**Impaired**	-0.08 kcal/mol (Stb)	-0.72 kcal/mol (Dstb)	-4.20 kcal/mol (Stb)	0.00 kcal/mol (?)	0.67 kcal/mol (Dstb)	**Inconclusive**

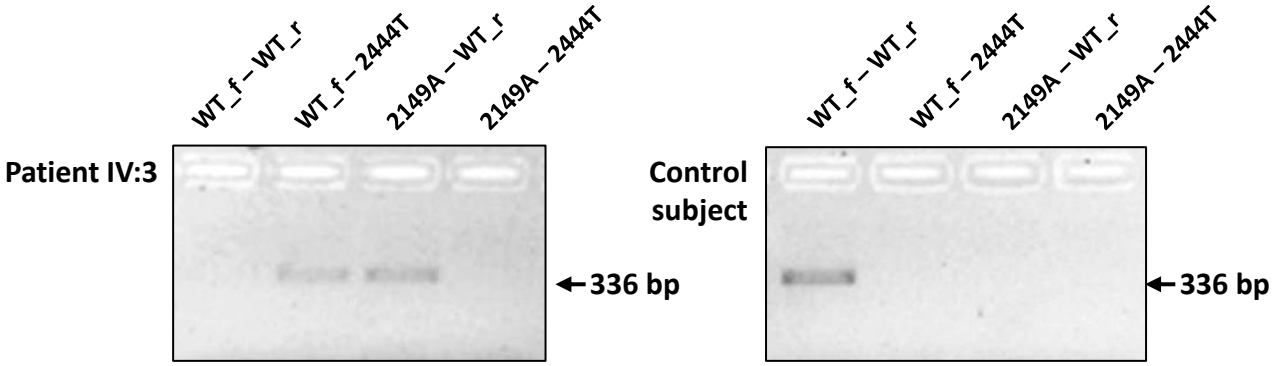
Supplementary Table S3. Germline microsatellite instability analysis of D2S123, D17S250 and D17S791 markers. Dinucleotide markers were amplified from genomic DNA from patient IV:3, a CMMRD control and two healthy controls. The gMSI ratios were determined by dividing the height of an allele's trailing "stutter" peak (n + 1) by the height of the allele's major peak (n). The threshold used for elevated gMSI ratio was previously determined as follows: >0.109 for D2S123, >0.074 for D17S250 and >0.095 for D17S791 (8).

Microsatellite	Sample	Exp. 1	Exp. 2	Exp. 3	Mean	SD
D2S123	Patient IV:3	0,0302	0,0293	0,0306	**0,0300**	**0,0007**
	Control CMMRD	0,1094	0,1113	0,1079	**0,1095**	**0,0017**
	Normal Control 1	0,0339	0,0326	0,0307	**0,0324**	**0,0016**
	Normal Control 2	0,0333	0,0345	0,0464	**0,0380**	**0,0073**
D17S250	Patient IV:3	0,0429	0,0417	0,0402	**0,0416**	**0,0013**
	Control CMMRD	0,1139	0,1117	0,1027	**0,1094**	**0,0059**
	Normal Control 1	0,0424	0,0404	0,0426	**0,0418**	**0,0012**
	Normal Control 2	0,0512	0,0590	0,0572	**0,0558**	**0,0041**
D17S791	Patient IV:3	0,0887	0,0983	0,0889	**0,0920**	**0,0055**
	Control CMMRD	0,1931	0,1912	0,1975	**0,1939**	**0,0033**
	Normal Control 1	0,0847	0,0815	0,0813	**0,0825**	**0,0019**
	Normal Control 2	0,0810	0,0748	0,0765	**0,0774**	**0,0032**

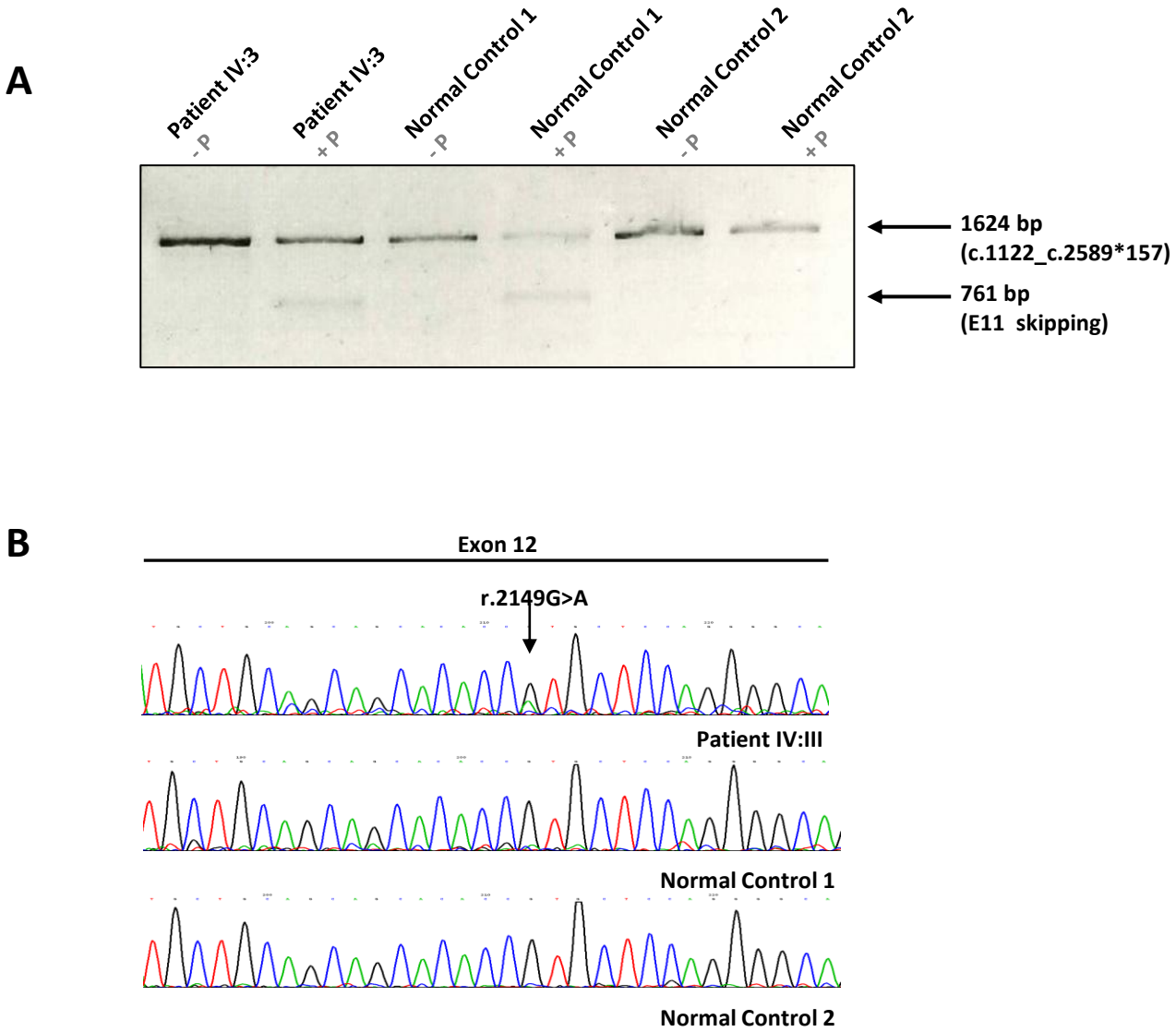
Supplementary Fig. S1. Immunohistochemistry analysis for MMR proteins in the colorectal adenocarcinoma from individual IV:3. Tumor sections immunostained with MLH1, MSH2, MSH6 and PMS2 antibodies (x200). The nuclear expression of MLH1, MSH2 and MSH6 is preserved in tumor and non-neoplastic cells. In contrast, PMS2 expression is lost in neoplastic tissue but conserved in nonmalignant cells (e.g., peritumoral and infiltrating lymphocytes as shown in the image).



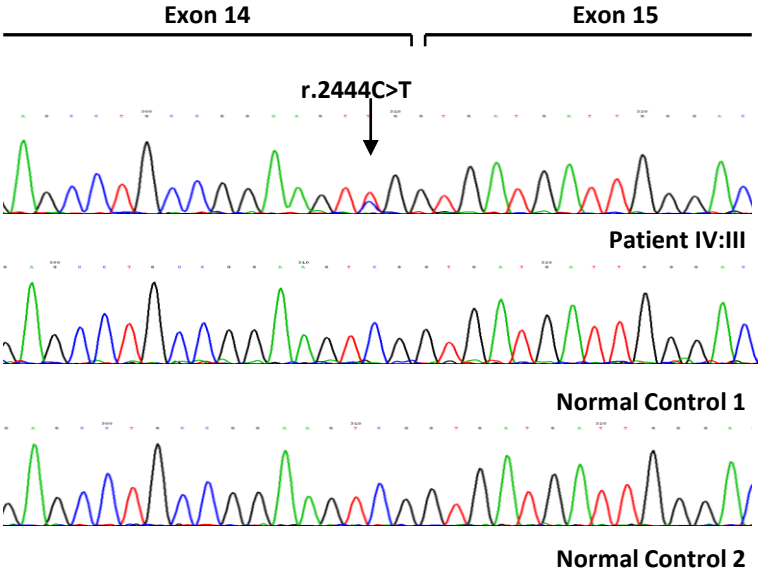
Supplementary Fig. S2. Evaluation of the allelic location of *PMS2* variants identified in individual IV:3. In order to determine the allelic location of the two genetic *PMS2* variants, RNA was extracted from cultured lymphocytes (in the absence of puromycin) of individual IV:3 and a wild-type *PMS2* control, and reverse-transcribed as detailed in Materials and Methods. cDNA amplification was performed using four different combinations of primers which carried specific sequences for *PMS2* wild-type or variants at the 3'-end [forward: either wild-type (WT_f) or variant c.2149A (2149A); reverse: either wild-type (WT_r) or variant c.2444T (2444T)]. A PCR band was amplified when the indicated combination of sequences was present on the template cDNA. The amplified bands indicated that the variants identified in *PMS2* are located *in trans* in individual IV:3.



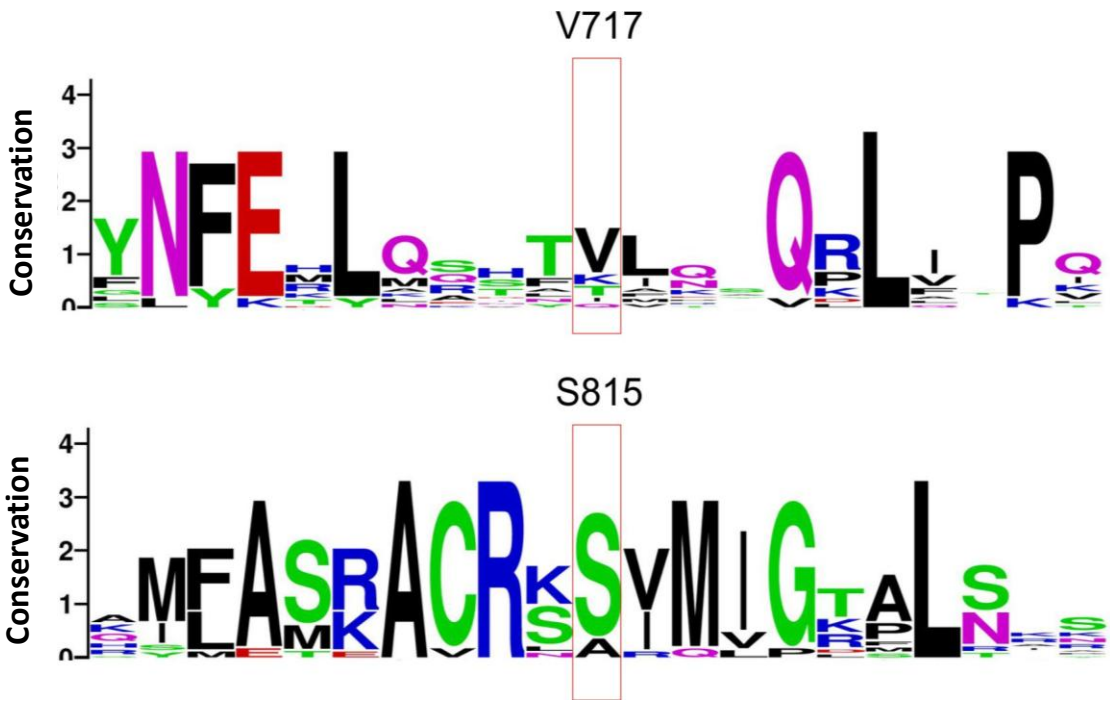
Supplementary Fig. S3. Splicing analysis of *PMS2* c.2149G>A andc.2444C>T variants. A: Agarose gel showing the RT-PCR products amplified using primers located at *PMS2* exons 10 and 15 in individual IV:3 (c.[2149G>A];[2444C>T] carrier), and in two control individuals. The upper band corresponds to the whole cDNA fragment, and the bottom band corresponds to an alternative transcript with exon 11 skipping. **B:** Direct sequencing of variant c.2149G>A. **C:** Direct sequencing of variant c.2444C>T. In B and C, RT-PCR products were obtained from lymphocytes cultured in the absence of puromycin from the *PMS2* c.[2149G>A];[2444C>T] carrier and two control individuals. In the presence of puromycin sequencing analysis showed the same pattern (data not shown)**.** Abbreviations: -P; RNA obtained from lymphocytes cultured in the absence of puromycin; +P; RNA obtained from lymphocytes cultured in the presence of puromycin.



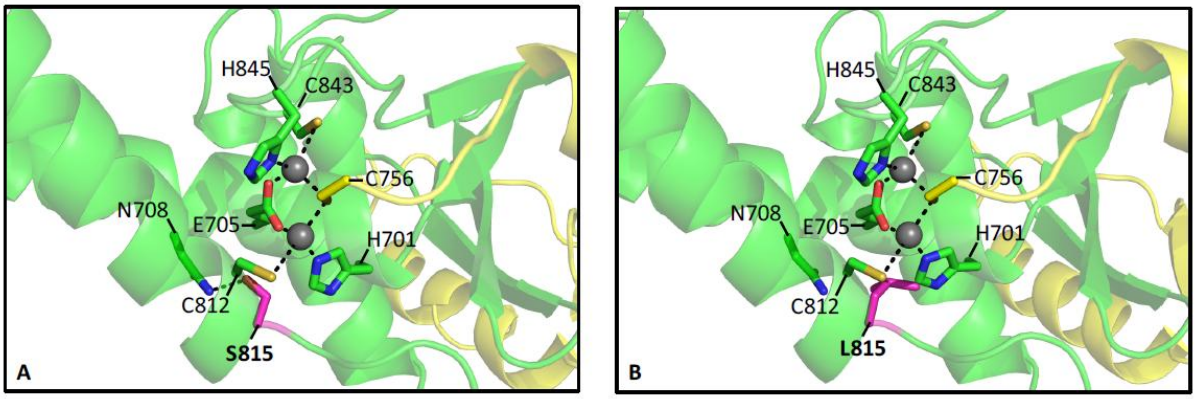
C


Exon 14
Exon 15
r.2444C>T
Patient IV:III
Normal Control 1
Normal Control 2

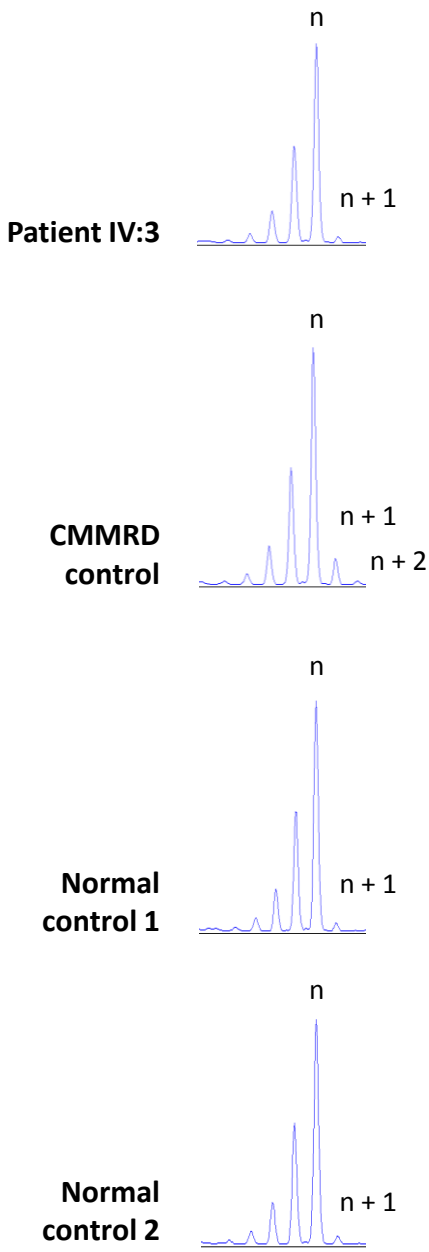
Supplementary Fig. S4. Evolutionary conservation of the positions affected in the *PMS2* variants. Conservation graphic for the amino acid residues affected by the variants analyzed. Affected residues are indicated by red boxes. Overall/individual font sizes indicate overall/individual conservation, respectively.



Supplementary Fig. S5. Structure model of the environment of the wildtype PMS2 S815 (A) and mutant residue L815 (B). Zinc molecules and Zinc binding residues are represented with dark grey spheres and sticks, respectively. Amino acid 815 is highlighted in pink. Dashed lines indicate their interactions.



Supplementary Fig. S6. Germline microsatellite instability analysis of D2S123 marker. Electropherograms showing fluorescence PCR products of D2S123 marker amplified from genomic DNA from patient IV:3, a CMMRD patient (homozygous carrier of the *PMS2* c.24-2A>G, r.24_28del, p.S8Rfs*4; data not shown) and two healthy controls. In patient IV:3 and healthy controls a stable microsatellite is detected, with very little PCR "stutter" of a size larger (*n* + 1) than the principal allele peaks (*n*). In contrast, in the CMMRD patient the presence of higher proportion of larger alleles (n+1, n+2) is evidenced.



ARTÍCULO 2

Validation of an *in vitro* mismatch repair assay used in the functional characterization of mismatch repair variants

Maribel González-Acosta, Inga Hinrichsen, Anna Fernández, Conxi Lázaro, Marta Pineda*, Guido Plotz*, Gabriel Capellá*.
* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Manuscrito enviado al *Journal of Molecular Diagnostics*, en revisión.

RESUMEN:

El ensayo *in vitro* de actividad reparadora se utiliza para evaluar la capacidad de reparación de las variantes en genes MMR, la función más importante de una proteína MMR. Sin embargo, la solidez del ensayo, fundamental para su uso en el entorno clínico, rara vez se ha evaluado. El objetivo de este trabajo fue validar uno de los ensayos *in vitro* de actividad reparadora para la caracterización funcional de las variantes MMR.

El ensayo *in vitro* de actividad reparadora se optimizó testando diferentes reactivos y condiciones experimentales. También, se establecieron materiales de referencia y protocolos estándar. Para determinar la variabilidad intra- e inter-experimental del ensayo y su reproducibilidad entre centros, se estudiaron funcionalmente seis variantes en *MLH1*, previamente caracterizadas, en dos laboratorios independientes.

Como resultado, se establecieron los reactivos y condiciones óptimas para realizar el ensayo *in vitro* de actividad reparadora. Además, el ensayo no demostró una variabilidad intra- o inter-experimental significativa y presentó buena reproducibilidad entre laboratorios.

En conclusión, se ha establecido un ensayo sólido que puede proporcionar evidencias funcionales relevantes para la evaluación de la patogenicidad de las variantes MMR, mejorando eventualmente el diagnóstico molecular de los síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema MMR.

TITLE: Validation of an *in vitro* mismatch repair assay used in the functional characterization of mismatch repair variants.

AUTHORS:
Maribel González-Acosta[1], Inga Hinrichsen[2], Anna Fernández[1], Conxi Lázaro[1], Marta Pineda*[1], Guido Plotz*[2], Gabriel Capellá[1*]
(*) Equally contributed to this work.

[1]Hereditary Cancer Program, Catalan Institute of Oncology - ICO, Hereditary Cancer Group, Molecular Mechanisms and Experimental Therapy in Oncology Program, Institut d'Investigació Biomèdica de Bellvitge – IDIBELL, Ciber Oncología (CIBERONC) Instituto Salud Carlos III, L'Hospitalet de Llobregat, Barcelona, Spain.
[2]Biomedizinisches Forschungslabor, Medizinische Klinik 1, Universitätsklinikum Frankfurt, Frankfurt, Germany.

NUMBER OF TEXT PAGES, TABLES AND FIGURES: 14 text pages and 4 figures.

RUNNING HEAD: Mismatch repair assay validation

GRANT NUMBERS AND SOURCES OF SUPPORT: This work was funded by the Spanish Ministry of Economy and Competitiveness and cofunded by FEDER funds -a way to build Europe- (grant SAF2015-68016-R), CIBERONC, the Spanish Association Against Cancer (grant 080253), the Government of Catalonia (grants 2014SGR338, 2017SGR1282 and PERIS SLT002/16/0037) and Fundación Mutua Madrileña. MG was supported by a grant from AECC, AF by the Catalonian Health Department (SLT002/16/00409) and GP by the Deutsche Forschungsgemeinschaft. We thank CERCA Programme for institutional support.

CORRESPONDING AUTHOR:
Gabriel Capellá, MD, PhD
Hereditary Cancer Program
Catalan Institute of Oncology, IDIBELL
Av. Gran Via de l'Hospitalet, 199-203
08908 Hospitalet de Llobregat, Spain
Tel: (+34) 93 260 7319
Fax: (+34) 93 260 7466
E-mail: gcapella@iconcologia.net

Footnote: Portions of this work were presented at the 7th biennial Scientific Conference of InSiGHT, held on 5-8th July 2017 in Florence, Italy.

ABSTRACT

A significant proportion of DNA mismatch repair (MMR) variants are classified as of unknown significance (VUS), precluding diagnosis. The *in vitro* MMR assay is used to assess their MMR capability, likely the most important function of a MMR protein. However, robustness of the assay, critical for its use in the clinical setting, has been rarely evaluated. The aim of the present work was to validate an *in vitro* MMR assay approach for the functional characterization of MMR variants, as a first step to meeting quality standards of diagnostic laboratories.

The MMR assay was optimized by testing a variety of reagents and experimental conditions. Reference materials and standard operating procedures were established. To determine the intra- and inter-experimental variability of the assay and its reproducibility among centers, independent transfections of six previously characterized *MLH1* variants were performed in two independent laboratories. Optimal reagents and conditions to perform the *in vitro* MMR assay were determined. The validated assay demonstrated no significant intraexperimental or interexperimental variability and good reproducibility between centers.

We have set up a robust *in vitro* MMR assay that can provide relevant *in vitro functional* evidence for MMR variant pathogenicity assessment, eventually improving the molecular diagnosis of hereditary cancer syndromes associated with MMR deficiency.

INTRODUCTION

Mismatch repair (MMR) system corrects base-base mismatches errors and small insertions and deletions mainly introduced by DNA polymerases during replication, but also mispairs formed during recombination or chemically modified bases.[1, 2] In humans, base-base MMR is initiated when the heterodimer formed by MSH2/MSH6 recognizes a mismatch. Then, the MLH1/PMS2 protein complex is recruited and binds other proteins, such as PCNA and RFC. PMS2 introduces a nick in the daughter strand, ExoI degrades the sequence containing the error and the strand is finally resynthesized.

Germline monoallelic mutations and epimutations in the major MMR genes (*MLH1, MSH2, MSH6,* and *PMS2*) cause Lynch syndrome[3] (LS; MIM 120435), whereas its biallelic inactivation underlies Constitutional MMR deficiency syndrome[4, 5] (CMMRD; MIM 276300). Hence, detection of germline MMR gene pathogenic mutations allows the diagnosis of MMR-associated cancer syndromes and the appropriate management of patients and their families.[6, 7] However, in routine diagnosis MMR variants of unknown significance (VUS) are often identified, representing up to 30% of all the identified variants,[8] precluding diagnosis for carriers and their relatives.[9] Moreover, the number of MMR VUS identified is increasing by the implementation of LS population screening of colorectal cancer and the use of multigene panel testing.[10-13] To facilitate classification of MMR variants in terms of pathogenicity, several guidelines have been developed during the last years, including generic and gene-specific rules.[8, 14, 15] All of them integrate multiple lines of evidence to classify variants, including that obtained by functional assays. These assays evaluate the impact of a variant at different molecular levels (e.g. specific function, expression and stability, subcellular localization), contributing to sort out uncertainty on variant pathogenicity assessment.[16] While well-established functional tests are mandatory to support the pathogenicity of a VUS, lack of standardization may entail discordant results between distinct laboratories.[17]

Within the large diversity of assays used in functional assessment of MMR variants, methods addressed to evaluate the MMR capability, likely the most important function of a MMR protein, are proposed as the gold standard.[16] *In vitro* cell-free assays, using protein extracts or human purified proteins together with nuclei extracts and a substrate to repair, are currently the most commonly used, providing advantage over yeast-based *in vivo* or cell-based *ex vivo* assays.[18] For years, they have been used in functional assessment of MMR variants of unknown significance identified in LS suspected individuals. Moreover, a repair assay has recently proven to be useful for the diagnosis of CMMRD analyzing non-neoplastic tissue.[19] However, the robustness of these approaches, critical for their routine use in the clinical setting, has been rarely evaluated.[20]

The aim of the present work is to validate an *in vitro* MMR assay approach that uses cell-free protein extracts[21] by providing optimized protocols as a first step for meeting quality standards

118 of diagnostic laboratories. The validation includes the evaluation of intraexperimental and
119 interexperimental variability of the assay, as well as reproducibility among centers.
120
121 **MATERIALS AND METHODS**
122
123 **CELL LINES AND PLASMIDS**
124 HEK293T cells, defective for MLH1 expression,[22] were grown in Dulbecco's Modified Eagle
125 Medium (Gibco, Thermo Fisher Scientific, Waltham, Massachusetts, US) with 10% fetal bovine
126 serum (Gibco) and 1% penicillin-streptomycin (Gibco).
127
128 pcDNA3.1_MLH1 and pSG5_PMS2 plasmids, kindly provided by Dr. Kolodner and Dr. Nyström-
129 Lahti. Six selected *MLH1* missense variants were constructed by site directed mutagenesis using
130 QuikChange Site Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, California, US)
131 according to the manufacturer's instructions (Supplemental Table S1). Sanger sequencing was
132 used to verify the presence of the *MLH1* variants in pcDNA3.1 expression plasmid. Of note, the
133 original pcDNA3.1_MLH1 plasmid harbored a silent variant at the 2221 position (c.2221T>C;
134 p.L741=) although it did not affect the MMR activity of the wild-type MLH1 (Supplemental Figure
135 S1). The 6 selected *MLH1* variants were previously characterized at functional and expression
136 level[23-30] (Supplemental Table S1 and S2), of which *MLH1* p.I219V and p.G67R variants were used
137 as proficient and deficient control variants, respectively. pUC19CPDC plasmid, kindly provided
138 by Dr. John B. Hays, was used in the construction of mismatched plasmids.
139
140 **MMR ASSAY OPTIMIZATION**
141 The *in vitro* MMR assay previously described by our group[21] was used as the basis for assay
142 optimization (Figure 1).
143
144 To evaluate the optimal mismatch-nick distance and orientation for specific DNA MMR activity,
145 twelve pUC19CPDC mismatched plasmids were constructed at different distances and
146 orientations between the localization of the mismatch and the nick by using the enzymes
147 Nt.BbvCI or Nb.BbvCI and adding their restriction sequences at different positions of the
148 pUC19CPDC plasmid (Figure 2A and Supplemental Table S3). Resulting constructions were used
149 in the described MMR assay to determine the optimal combination.
150
151 Different amounts of whole HEK293T cell protein extract (in the range of 0 to 10 µg) containing
152 transfected wild-type MLH1/PMS2 proteins were tested to determine the optimal amount for
153 the assay. Salt concentration was evaluated by using different KCl concentrations from 45 to 180
154 mM. The incubation time of the MMR reaction was evaluated for 1, 2, 5, 7.5, 10 and 15 minutes.
155 To evaluate whether nick-ligation or protein-DNA-complex formations during pre-incubation on
156 ice influence repair efficiency, the reaction was pre-incubated on ice for 1, 2, 5, 10 and 15
157 minutes.

158

159 ***IN VITRO* MMR ASSAY VALIDATION**

160 Standard Operating Procedures (SOP) for reagent preparation (SOP 1), nuclear proteins
161 extraction (SOP 2), HEK293T cells transfection and whole cell protein extraction (SOP 3),
162 mismatched plasmid substrate generation (SOP 4) and MMR assay (SOP 5) are detailed in
163 http://dx.doi.org/10.17632/z8z3yvkv9g.1.

164

165 *Nuclear protein extraction*

166 Nuclear protein extraction protocol is detailed in SOP 2
167 (http://dx.doi.org/10.17632/z8z3yvkv9g.1). Briefly, HEK293T cells were resuspended in 3-times
168 their packed cell volume in ice-cold hypotonic buffer and lysed with Dounce pestle. After
169 centrifugation, the cytoplasmic supernatant was removed and centrifuged again to remove the
170 residual supernatant. The pellet was resuspended in resuspension buffer and then high-salt
171 buffer was added under soft agitation. Extraction was performed for 30 min at 4ºC and extracted
172 nuclei were removed by centrifugation. The supernatant was dialyzed, centrifuged and stored
173 in aliquots at -80ºC.

174

175 *Transfection and whole cell protein extract preparation*

176 The protocol for HEK293T cell transfection and preparation of whole cell protein extracts is
177 detailed in SOP 3 (http://dx.doi.org/10.17632/z8z3yvkv9g.1). In brief, HEK293T cells were
178 transfected at 30–40% confluence with MLH1 and PMS2 expression plasmids (0.5 µg/ml of each)
179 and 0.05 µg/ml of pGFP, as a transfection control, using 2 µl/ml of the cationic polymer
180 polyethylenimine (stock solution 1 mg/ml) (Polysciences, Warrington, Pennsylvania, US). After
181 48h, cells were harvested and prepared for protein extraction and cytometer analysis.

182

183 After 48h of transfection, cells were washed in PBS and resuspended in 2-times the packed cell
184 volume of hypotonic buffer. The suspension was frozen at -80ºC and thawed on ice for lysis. This
185 suspension was supplemented with an identical volume of hypertonic buffer. The suspension
186 was rocked on ice for 30 min and then centrifuged. The supernatant (the whole cell extract) was
187 finally stored in aliquots at -80ºC.

188

189 *Mismatched plasmid construction*

190 Construction of the 3' nicked G-T mismatched plasmid substrate is detailed in SOP 4
191 (http://dx.doi.org/10.17632/z8z3yvkv9g.1). In brief, 50 ng of pUC19CPDC Bbv 83 plasmid was
192 digested with N.BstNBI restriction enzyme for 3 h at 55ºC, which generated two single-strand
193 breaks. Complete digestion was assured by running aliquots of the reaction on an agarose gel.
194 If the enzymatic digestion was not complete, reaction was continued by adding additional
195 enzyme and incubating for one additional hour. The digested single stranded 32-bp oligomer
196 was captured by denaturing at 85ºC for 5 min in the presence of 50-fold excess of WHCPDPuriAS
197 antisense oligomer and subsequent slow cooling to room temperature. Oligomers were

198 removed by centrifugation through *Amicon 50K 500 µl* spin columns (Millipore, Merck KGaA,
199 Darmstadt, Germany) and extensive washing with TE buffer. Gapped plasmid was then ligated
200 with a 10-fold molar excess of WHpCPD7 oligomer that contains the mismatched residue.
201 Ligation was carried out overnight with T4 DNA ligase at 16ºC. Ligated product was ethanol
202 precipitated and subsequently treated with EcoRV and Exonuclease V to eliminate the residual
203 original plasmid. After another round of precipitation, the mismatched plasmid was digested
204 with Nt.Bbv CI restriction enzyme in order to introduce a single strand break in a specific location
205 separated from the mismatched. Finally, the nicked mismatched plasmid was purified by
206 centrifugation through *Amicon 50K 500 µl* spin columns. The resulting plasmid contains one AseI
207 restriction site, a GT mismatch within an overlapping AseI/EcoRV restriction site as well as a
208 single strand nick in the 3' position of the mismatch, which serves to direct MMR to convert the
209 GT mismatch to GC, thereby restoring the EcoRV restriction site. Digestion of this preparation
210 with AseI and EcoRV must yield only linearized vector, but no detectable fragments.
211
212 *Mismatch repair assay*
213 MMR assay conditions are detailed in SOP 5 (http://dx.doi.org/10.17632/z8z3yvkv9g.1). In
214 short, the reaction was performed in 15 µl total volume with reaction buffer (25 mM Tris–HCl
215 pH 7.5, 110 mM KCl, 5 mM MgCl2, 50 µg/ml BSA, 1.5 mM ATP, 0.1 mM each dNTP), 50 ng DNA
216 mismatched plasmid substrate, 50 µg HEK293T cells nuclear extract and 5 µg of whole protein
217 extract from transfected-HEK293T cells. Reaction was incubated at 37ºC for 15 min and
218 terminated with 25 µl stop-buffer (100 mM EDTA, 10% SDS, 20 mg/ml proteinase K) by an
219 additional incubation for 10 min at 37ºC. Plasmids were extracted from the reaction mixture by
220 phenol-chloroform extraction and purified by ethanol co-precipitation with tRNA. Subsequent
221 digestion with AseI, EcoRV and RNAse A produced two smaller fragments besides the linearized
222 vector when repair was successful. Restriction digestions were separated in 2% agarose gels.
223 Bands intensity was quantified using ImageLab™ Software v.2.0.1 (Bio-Rad, Hercules, California,
224 US).
225
226 Repair efficiency was measured as follows: i) ***Absolute Repair*** value was calculated as the
227 percentage of the intensities of those bands indicating repair in relation to the sum of all band
228 intensities; ii) ***Relative repair:*** The repair efficiency of the *MLH1* variants was analyzed in direct
229 comparison with the activity of the wild-type protein that had been expressed, processed and
230 tested in parallel. The relative repair value was calculated by subtracting the absolute repair
231 value of the non-transfected group from the absolute repair value in wild-type
232 and *MLH1* variants and, then, dividing the result for the tested variants by that of the wild-type
233 protein, multiplied by 100. The minimum accepted value of wild-type absolute repair was 25%.
234
235 **EVALUATION OF THE VARIABILITY AND REPRODUCIBILITY OF THE MMR ASSAY**
236 To analyze the intra- and inter-experimental variability of the MMR assay, three independent
237 transfections were performed including the following plasmids: eGFP (negative control, 1 plate

per transfection), wild-type MLH1/PMS2 (3 independent plates per transfection), I219V-MLH1/PMS2 (proficient control, 1 plate per transfection) and G67R-MLH1/PMS2 (deficient control, 1 plate per transfection) (Supplemental Figure 2A). Different nuclear extracts were used in every transfection experiment. The MMR assay reaction of each plate was performed in triplicate. Intra-experimental variability was estimated from the observed values of the transfection of wild-type MLH1/PMS2 plasmids in 3 independent plates in 3 independent transfection experiments. Inter-experimental variability was determined from the observed values of the transfection of wild-type MLH1/PMS2 as well as the control variants (1 plate per experiment in three independent transfection experiments). One MMR assay was performed for each condition.

To analyze the reproducibility of the technique between different centers, six *MLH1* variants (p.I219V, p.G67R, p.V716M, p.T82A, p.A618T and p.L622H) were studied in two independent laboratories. Each laboratory carried out three independent transfections of each variant and subsequently performed one MMR assay for each transfection.

MLH1 AND PMS2 PROTEIN EXPRESSION ANALYSIS

MLH1 and PMS2 protein expression levels in transfected HEK293T cells were examined by SDS-PAGE, followed by Western blotting analysis with anti-MLH1 (clone G168-15) (BD Pharmingen, BD, Franklin Lakes, New Jersey, US) and anti-PMS2 (clone 16-4, BD Pharmingen) antibodies. Band intensities were quantified using QuantityOne v.4.4 (Bio-Rad). Alfa-actin expression was assessed in parallel and used as loading control. Expression of MLH1 and PMS2 was normalized to alpha-actin expression. The relative protein expression value was calculated by dividing the normalized protein expression in variant-transfected cells by the expression in wild-type MLH1/PMS2-transfected cells, processed and tested in parallel. Protein expression analyses were performed in triplicate from 3 independent transfection experiments.

STATISTICAL ANALYSIS

To analyze intra-experimental and inter-experimental variability of the MMR assay, absolute repair and relative repair values were calculated, respectively. A two-way repeated measures ANOVA was used to compare the mean differences between and within experiments of the % of repair. All reported p-values were 2 sided, and $p<0.05$ was considered statistically significant. All calculations were performed using R version 3.1.2.

To assess reproducibility of the assay, significant differences between centers were analyzed using the non-parametric Mann–Whitney U test for quantitative data. All reported p-values were 2 sided, and $p<0 .05$ was considered significant. All calculations were performed using SPSS 19.0 (IBM, Armonk, New York, US).

278 **RESULTS**

279

280 **MMR ASSAY OPTIMIZATION**

281 A variety of reagents and parameters were tested in order to optimize the *in vitro* MMR assay
282 previously reported by our group,[21] that used HEK293T nuclei extracts reconstituted with
283 purified transfected proteins in the presence of a mismatched plasmid substrate (Figure 1).

284

285 Twelve DNA mismatched pUC19CPDC plasmid substrates were designed, differing in the
286 mismatch-nick distance and orientation (Figure 2A and Supplemental Table S3). The pUC19CPDC
287 Bbv 83 plasmid, with a nick-mismatch distance of 82 bp in the 3' orientation, gave the highest
288 absolute repair yield and specificity for MLH1/PMS2-dependent repair. In contrast,
289 MLH1/PMS2-independent 5'-repair occurred when the nick is in the 5' orientation (Figure 2B
290 and 2C) as previously reported.[31, 32] Accordingly plasmid pUC19CPDC Bbv 83-3' was chosen for
291 the successive experiments.

292

293 Different amounts of whole cell protein extracts from HEK293T MLH1/PMS2-transfected cells
294 were used to test the dependency of the quantity of the protein extract on the efficiency of the
295 MMR assay. The obtained results showed a drop of repair activity below 2 µg of extract whereas
296 higher amounts (from 2.5 to 10 µg) offered similar repair levels (Figure 3A). KCl concentration
297 was not critical: 45 to 180 mM levels demonstrated to be optimal for the MMR reaction (Figure
298 3B). Finally, the role of the incubation time was studied: a proportional increase of repair with
299 time was observed from 0 to 10 minutes with negligible increases from 10 to 15 minutes. In this
300 time frame a good discrimination between unrepaired and repaired substrate was depicted
301 (Figure 3C). Of note, the preincubation of the MMR reaction on ice before the addition of the
302 whole cell protein extracts caused a reduction in the final repair activity proportional to the time
303 of preincubation (Figure 3D).

304

305 In conclusion, the plasmid substrate pUC19CPDC Bbv 83, 5 µg of whole transfected-cell protein
306 extracts, 110mM KCl concentration and a 15-minute reaction time without preincubation on ice
307 were chosen as the optimal conditions to perform the *in vitro* MMR assay. Subsequently,
308 reference materials and SOP for nuclear protein extraction, HEK293T cell transfection and whole
309 cell protein extraction, mismatched plasmid substrate generation and MMR assay reaction were
310 defined (http://dx.doi.org/10.17632/z8z3yvkv9g.1). Quality control measures included in the
311 SOPs are detailed in Supplemental Table S4.

312

313 **VALIDATION OF THE *IN VITRO* MMR ASSAY**

314 The MMR activity of 6 *MLH1* variants was assessed using our optimized MMR assay
315 (Supplemental Table S1). Minimal dispersion was seen in non-transfected group as well as MLH1
316 p.I219V and p.G67R control variants (Supplemental Figure 2B). No significant intra-experimental
317 variability was observed in the absolute repair values of wild-type MLH1 protein in three

independent experiments using distinct preparations of nuclear extracts (Figure 4A). However, inter-experimental differences in absolute repair were observed (Figure 4A). Of note, differences disappeared when the same nuclear extract preparation was used (Supplemental Figure 3), suggesting that absolute repair depends on the intrinsic efficiency of each nuclear extract preparation. Thus, the relative repair values of *MLH1* variants (depicting their repair efficiency in relation to absolute repair of the wild-type protein analyzed in parallel) were used to evaluate the inter-experimental variability. No significant differences were detected in the relative repair values of MLH1 p.I219V and p.G67R variants (Figure 4B), allowing a meaningful comparison of MMR activities from different experiments even using different preparations of nuclear extracts.

Furthermore, no significant differences were observed among the data obtained in two independent laboratories on the evaluation of the relative repair of six *MLH1* variants (Figure 4C), demonstrating reproducibility between two centers.

For the interpretation of MMR activity results, conservative cut-offs previously suggested by InSiGHT[8] -set as <35% (for MMR-deficiency) and >65% (for MMR-proficiency)- were used. Functional results were in the range of previous analyses for five of the six analyzed variants.[23-25, 29, 30] In contrast, our results showed MMR deficiency for the pathogenic p.L622H variant, previously catalogued as MMR proficient, also associated with decreased expression levels (Supplemental Table S1 and Supplemental Figure 4).[24, 30] In all, the sensitivity and specificity of our optimized assay was similar to that offered by other approaches (Supplemental Table S5).

DISCUSSION

Here we report a validated and optimized protocol to perform an *in vitro* cell-free MMR assay. The dependency of the repair activity of MLH1 and the mismatched substrate on the amount of whole protein cell extracts, the KCl concentration and the incubation time has been assessed. Although cell-free MMR assay approaches require significant technical expertise in molecular and cellular biology methods, the 5 SOPs reported are critical to limit intra- and inter-experimental variability making the assay robust, as a first step towards meeting laboratory quality standards (Supplementary Table 6).

The optimized reconstitution assay is sufficiently insensitive to small deviations of the experimental conditions demonstrating robustness and reproducibility in the relative repair values observed in the six *MLH1* variants analyzed. Of note, our validated methodology has already been used in the characterization of several variants including *MLH1* c.121G>C, *PMS2* c.2149G>A and *PMS2* c.2444C>T.[33, 34] The impaired MMR activity of *MLH1* c.121G>C and *PMS2* c.2444C>T variants was further confirmed by other groups.[35, 36]

356
357 The in-depth characterization of our validated MMR assay approach pointed to the MMR
358 efficiency of nuclear extracts as the main cause of differences in the inter-experimental absolute
359 repair. Nevertheless, in the analysis of MMR variants, these differences can be easily overcome
360 by using the relative repair, obtained as the percentage of the absolute repair of a variant
361 divided by the same result of the wild-type MLH1. In the *in vitro* MMR assay the absolute repair
362 reached by wild-type MLH1 is not 100%, in line with previous reports.[23-25, 35, 37] The ligation of
363 the 3'-nick of the mismatched plasmid substrate during the repair reaction, that would make
364 the substrate refractory to repair, may underlie this observation. Accordingly, the results
365 obtained after preincubation on ice also highlighted that a fraction of the mismatched DNA
366 plasmid substrate is modified in a way that subsequently does not allow repair in a time
367 dependent manner.
368

369 Noteworthy, our methodology allows the analysis of the MMR activity and expression of MMR
370 variants by using the same whole cell protein extract of transfected cells, in contrast to other *in*
371 *vitro* cell-free MMR assay approaches (Supplemental Table S6). For example, while the
372 production of recombinant proteins in S*f*9 insect cells is a good strategy to generate high
373 amounts of protein, the levels of expression in this heterologous system do not always correlate
374 with those obtained from human cells.[23, 38] Similarly, the expression of human mutant proteins
375 by *in vitro* translation used by Drost and collaborators, called CIMRA,[20, 25, 37] which facilitates the
376 protein generation process, is a good approach to analyze the intrinsic MMR activity of a variant,
377 although it precludes the simultaneous analysis of the levels of protein expression.
378

379 The protein expression of p.L622H variant was initially analyzed because of apparently non-
380 concordant MMR activity results with previously reported functional assays.[24, 30] This variant
381 represented an analytical challenge because of its milder expressivity.[26] The reduced expression
382 observed in our analysis was also reported in other studies after transfection in HEK293T and
383 HCT116 cells.[24, 26, 27, 30] The variable level of reduction observed in those studies is probably a
384 consequence of its reduced stability, as reported in HCT116 cells after cyclohexamide
385 treatment.[26] In essence, the data suggested a better sensitivity of our MMR assay when testing
386 variants associated with decreased protein stability.
387

388 Significant progress has been made in providing clinically calibrated cut-offs for high-throughput
389 functional results to differentiate deficient from proficient *BRCA1* variants.[39, 40] This is in contrast
390 with MMR variants since none of the previous studies using cell-free MMR assay approaches
391 provided thresholds for absolute or relative repair data (**Supplemental Table S6**). In the MMR
392 setting deficiency and proficiency of a given variant has been established in comparison with
393 control variants. A non-arbitrary threshold can only be defined after analyzing a larger set of
394 MMR variants and calibrating the values to other validated pieces of evidence. So far this has

been partially approached by Drost and collaborators,[20] where the CIMRA assay was calibrated with *in silico* variant predictions. The bayesian integration of the results obtained by the CIMRA assay highlighted the potential and limitations of the added value of validated MMR assay approaches.[20]

The HEK293T cell-based methodology presented here is a useful tool for functional assessment of *MLH1* and *PMS2* variants. Interestingly, our validated approach might be eventually adapted to the functional analysis of *MSH2* and *MSH6* variants. For this purpose, a MSH2/MSH6-deficient cell line (such as LoVo cells[41]) would be used for nuclear protein extraction and transient transfection of the variants of interest, followed by the optimization and validation of the MMR assay. In this regard, LoVo cells have been previously used for MSH2 and MSH6 variant assessment in functional[38, 42] and expression studies.[43]

Functional analyses based on *in vitro* MMR assay approaches are currently making important contributions to LS and CMMRD diagnosis.[8, 19] Similar to the MMR assay approach used for the identification of CMMRD,[19] our validated assay is also a 3'-nicked G-T mismatch-based repair assay, suggesting its potential usefulness for CMMRD diagnosis. Irrespective of the MMR assay approach used, standardization of the assay and establishment of quality control standards are mandatory in order to avoid discordant results between experimental replicates and/or laboratories. Nevertheless, validation of the obtained functional results in an independent assay is recommended, following the InSiGHT classification rules.[8]

In summary, we have validated a robust *in vitro* MMR assay that can provide meaningful *in vitro* evidence for the classification of VUS detected in MMR genes. The need for reproducible assays for functional VUS characterization becomes even more relevant as next generation sequencing is routinely implemented in diagnostic laboratories and the number of VUS identified keeps increasing.

ACKNOWLEDGEMENTS

We are indebted to the patients and their families. We thank all members of the Hereditary Cancer Program at the Catalan Institute of Oncology.

435 **REFERENCES**

436

437 1. Reyes GX, Schmidt TT, Kolodner RD, Hombauer H: New insights into the mechanism of
438 DNA mismatch repair. Chromosoma 2015, 124:443-462.
439 2. Jiricny J: Postreplicative mismatch repair. Cold Spring Harbor perspectives in biology
440 2013, 5:a012633.
441 3. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP: Milestones of Lynch syndrome:
442 1895-2015. Nature reviews 2015, 15:181-194.
443 4. Bakry D, Aronson M, Durno C, Rimawi H, Farah R, Alharbi QK, Alharbi M, Shamvil A, Ben-
444 Shachar S, Mistry M, Constantini S, Dvir R, Qaddoumi I, Gallinger S, Lerner-Ellis J, Pollett
445 A, Stephens D, Kelies S, Chao E, Malkin D, Bouffet E, Hawkins C, Tabori U: Genetic and
446 clinical determinants of constitutional mismatch repair deficiency syndrome: report
447 from the constitutional mismatch repair deficiency consortium. Eur J Cancer 2014,
448 50:987-996.
449 5. Wimmer K, Kratz CP, Vasen HF, Caron O, Colas C, Entz-Werle N, Gerdes AM, Goldberg Y,
450 Ilencikova D, Muleris M, Duval A, Lavoine N, Ruiz-Ponte C, Slavc I, Burkhardt B, Brugieres
451 L: Diagnostic criteria for constitutional mismatch repair deficiency syndrome:
452 suggestions of the European consortium 'care for CMMRD' (C4CMMRD). Journal of
453 medical genetics 2014, 51:355-365.
454 6. Vasen HF, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, Bernstein I, Bertario L,
455 Burn J, Capella G, Colas C, Engel C, Frayling IM, Genuardi M, Heinimann K, Hes FJ,
456 Hodgson SV, Karagiannis JA, Lalloo F, Lindblom A, Mecklin JP, Moller P, Myrhoj T,
457 Nagengast FM, Parc Y, Ponz de Leon M, Renkonen-Sinisalo L, Sampson JR, Stormorken
458 A, Sijmons RH, Tejpar S, Thomas HJ, Rahner N, Wijnen JT, Jarvinen HJ, Moslein G: Revised
459 guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations
460 by a group of European experts. Gut 2013, 62:812-823.
461 7. Vasen HF, Ghorbanoghli Z, Bourdeaut F, Cabaret O, Caron O, Duval A, Entz-Werle N,
462 Goldberg Y, Ilencikova D, Kratz CP, Lavoine N, Loeffen J, Menko FH, Muleris M, Sebille
463 G, Colas C, Burkhardt B, Brugieres L, Wimmer K: Guidelines for surveillance of individuals
464 with constitutional mismatch repair-deficiency proposed by the European Consortium
465 "Care for CMMR-D" (C4CMMR-D). Journal of medical genetics 2014, 51:283-293.
466 8. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B,
467 Bernstein I, Capella G, den Dunnen JT, du Sart D, Fabre A, Farrell MP, Farrington SM,
468 Frayling IM, Frebourg T, Goldgar DE, Heinen CD, Holinski-Feder E, Kohonen-Corish M,
469 Robinson KL, Leung SY, Martins A, Moller P, Morak M, Nystrom M, Peltomaki P, Pineda
470 M, Qi M, Ramesar R, Rasmussen LJ, Royer-Pokora B, Scott RJ, Sijmons R, Tavtigian SV,
471 Tops CM, Weber T, Wijnen J, Woods MO, Macrae F, Genuardi M: Application of a 5-
472 tiered scheme for standardized classification of 2,360 unique mismatch repair gene
473 variants in the InSiGHT locus-specific database. Nature genetics 2014, 46:107-115.

474 9. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB,
475 Hoogerbrugge N, Spurdle AB, Tavtigian SV: Sequence variant classification and
476 reporting: recommendations for improving the interpretation of cancer susceptibility
477 genetic test results. Human mutation 2008, 29:1282-1291.
478 10. Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kaushik P, Roa BB, Wenstrup RJ,
479 Hartman AR, Syngal S: Identification of a Variety of Mutations in Cancer Predisposition
480 Genes in Patients With Suspected Lynch Syndrome. Gastroenterology 2015, 149:604-
481 613 e620.
482 11. Howarth DR, Lum SS, Esquivel P, Garberoglio CA, Senthil M, Solomon NL: Initial Results
483 of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer and Lynch
484 Syndrome. The American surgeon 2015, 81:941-944.
485 12. Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L,
486 Vaccari EM, Bissonnette J, Booker JK, Cremona ML, Gibellini F, Murphy PD, Pineda-
487 Alvarez DE, Pollevick GD, Xu Z, Richard G, Bale S, Klein RT, Hruska KS, Chung WK:
488 Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients
489 referred for next-generation cancer panel testing. Genet Med 2016, 18:823-832.
490 13. Rohlin A, Rambech E, Kvist A, Torngren T, Eiengard F, Lundstam U, Zagoras T, Gebre-
491 Medhin S, Borg A, Bjork J, Nilbert M, Nordling M: Expanding the genotype-phenotype
492 spectrum in hereditary colorectal cancer by gene panel testing. Familial cancer 2017,
493 16:195-203.
494 14. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E,
495 Spector E, Voelkerding K, Rehm HL: Standards and guidelines for the interpretation of
496 sequence variants: a joint consensus recommendation of the American College of
497 Medical Genetics and Genomics and the Association for Molecular Pathology. Genet
498 Med 2015, 17:405-424.
499 15. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho YY, Kobayashi Y, Patil N,
500 Thusberg J, Westbrook M, Topper S: Sherloc: a comprehensive refinement of the ACMG-
501 AMP variant classification criteria. Genet Med 2017, 19:1105-1117.
502 16. Heinen CD, Juel Rasmussen L: Determining the functional significance of mismatch
503 repair gene missense variants using biochemical and cellular assays. Hereditary cancer
504 in clinical practice 2012, 10:9.
505 17. Hoskinson DC, Dubuc AM, Mason-Suares H: The current state of clinical interpretation
506 of sequence variants. Current opinion in genetics & development 2017, 42:33-39.
507 18. Pena-Diaz J, Rasmussen LJ: Approaches to diagnose DNA mismatch repair gene defects
508 in cancer. DNA repair 2015, 38:147-154.
509 19. Shuen AY, Lanni S, Panigrahi GB, Edwards M, Yu L, Campbell BB, Mandel A, Zhang C,
510 Zhukova N, Alharbi M, Bernstein M, Bowers DC, Carroll S, Cole KA, Constantini S, Crooks
511 B, Dvir R, Farah R, Hijiya N, George B, Laetsch TW, Larouche V, Lindhorst S, Luiten RC,
512 Magimairajan V, Mason G, Mason W, Mordechai O, Mushtaq N, Nicholas G, Oren M,
513 Palma L, Pedroza LA, Ramdas J, Samuel D, Wolfe Schneider K, Seeley A, Semotiuk K,

514 Shamvil A, Sumerauer D, Toledano H, Tomboc P, Wierman M, Van Damme A, Lee YY,
515 Zapotocky M, Bouffet E, Durno C, Aronson M, Gallinger S, Foulkes WD, Malkin D, Tabori
516 U, Pearson CE: Functional Repair Assay for the Diagnosis of Constitutional Mismatch
517 Repair Deficiency From Non-Neoplastic Tissue. J Clin Oncol 2019, 37:461-470.
518 20. Drost M, Tiersma Y, Thompson BA, Frederiksen JH, Keijzers G, Glubb D, Kathe S, Osinga
519 J, Westers H, Pappas L, Boucher KM, Molenkamp S, Zonneveld JB, van Asperen CJ,
520 Goldgar DE, Wallace SS, Sijmons RH, Spurdle AB, Rasmussen LJ, Greenblatt MS, de Wind
521 N, Tavtigian SV: A functional assay-based procedure to classify mismatch repair gene
522 variants in Lynch syndrome. Genet Med 2018.
523 21. Plotz G, Welsch C, Giron-Monzon L, Friedhoff P, Albrecht M, Piiper A, Biondi RM,
524 Lengauer T, Zeuzem S, Raedle J: Mutations in the MutSalpha interaction interface of
525 MLH1 can abolish DNA mismatch repair. Nucleic acids research 2006, 34:6574-6586.
526 22. Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, Plotz G, Jiricny J, Marra
527 G: Functional analysis of hMLH1 variants and HNPCC-related mutations using a human
528 expression system. Gastroenterology 2002, 122:211-219.
529 23. Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lonnqvist KE, Holinski-Feder E,
530 Sutter C, McKinnon W, Duraisamy S, Gerdes AM, Peltomaki P, Kohonen-Ccorish M,
531 Mangold E, Macrae F, Greenblatt M, de la Chapelle A, Nystrom M: Functional
532 significance and clinical phenotype of nontruncating mismatch repair variants of MLH1.
533 Gastroenterology 2005, 129:537-549.
534 24. Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C:
535 Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair
536 assays. Cancer research 2007, 67:4595-4604.
537 25. Drost M, Zonneveld J, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind N:
538 A cell-free assay for the functional analysis of variants of the mismatch repair protein
539 MLH1. Human mutation 2010, 31:247-253.
540 26. Borras E, Pineda M, Blanco I, Jewett EM, Wang F, Teule A, Caldes T, Urioste M, Martinez-
541 Bouzas C, Brunet J, Balmana J, Torres A, Ramon y Cajal T, Sanz J, Perez-Cabornero L,
542 Castellvi-Bel S, Alonso A, Lanas A, Gonzalez S, Moreno V, Gruber SB, Rosenberg NA,
543 Mukherjee B, Lazaro C, Capella G: MLH1 founder mutations with moderate penetrance
544 in Spanish Lynch syndrome families. Cancer research 2010, 70:7379-7391.
545 27. Kosinski J, Hinrichsen I, Bujnicki JM, Friedhoff P, Plotz G: Identification of Lynch
546 syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and
547 mismatch repair. Human mutation 2010, 31:975-982.
548 28. Hardt K, Heick SB, Betz B, Goecke T, Yazdanparast H, Kuppers R, Servan K, Steinke V,
549 Rahner N, Morak M, Holinski-Feder E, Engel C, Moslein G, Schackert HK, von Knebel
550 Doeberitz M, Pox C, Hegemann JH, Royer-Pokora B: Missense variants in hMLH1
551 identified in patients from the German HNPCC consortium and functional studies.
552 Familial cancer 2011, 10:273-284.

553 29. Borras E, Pineda M, Brieger A, Hinrichsen I, Gomez C, Navarro M, Balmana J, Ramon y
554 Cajal T, Torres A, Brunet J, Blanco I, Plotz G, Lazaro C, Capella G: Comprehensive
555 functional assessment of MLH1 variants of unknown significance. Human mutation
556 2012, 33:1576-1588.
557 30. Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G: Expression defect size
558 among unclassified MLH1 variants determines pathogenicity in Lynch syndrome
559 diagnosis. Clin Cancer Res 2013, 19:2432-2441.
560 31. Constantin N, Dzantiev L, Kadyrov FA, Modrich P: Human mismatch repair:
561 reconstitution of a nick-directed bidirectional reaction. The Journal of biological
562 chemistry 2005, 280:39752-39761.
563 32. Zhang Y, Yuan F, Presnell SR, Tian K, Gao Y, Tomkinson AE, Gu L, Li GM: Reconstitution
564 of 5'-directed human mismatch repair in a purified system. Cell 2005, 122:693-705.
565 33. Pineda M, Gonzalez-Acosta M, Thompson BA, Sanchez R, Gomez C, Martinez-Lopez J,
566 Perea J, Caldes T, Rodriguez Y, Landolfi S, Balmana J, Lazaro C, Robles L, Capella G, Rueda
567 D: Detailed characterization of MLH1 p.D41H and p.N710D variants coexisting in a Lynch
568 syndrome family with conserved MLH1 expression tumors. Clinical genetics 2015,
569 87:543-548.
570 34. Gonzalez-Acosta M, Del Valle J, Navarro M, Thompson BA, Iglesias S, Sanjuan X, Paules
571 MJ, Padilla N, Fernandez A, Cuesta R, Teule A, Plotz G, Cadinanos J, de la Cruz X, Balaguer
572 F, Lazaro C, Pineda M, Capella G: Elucidating the clinical significance of two PMS2
573 missense variants coexisting in a family fulfilling hereditary cancer criteria. Familial
574 cancer 2017.
575 35. van der Klift HM, Mensenkamp AR, Drost M, Bik EC, Vos YJ, Gille HJ, Redeker BE, Tiersma
576 Y, Zonneveld JB, Garcia EG, Letteboer TG, Olderode-Berends MJ, van Hest LP, van Os TA,
577 Verhoef S, Wagner A, van Asperen CJ, Ten Broeke SW, Hes FJ, de Wind N, Nielsen M,
578 Devilee P, Ligtenberg MJ, Wijnen JT, Tops CM: Comprehensive Mutation Analysis of
579 PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional
580 Mismatch Repair Deficiency Syndrome. Human mutation 2016, 37:1162-1179.
581 36. Hinrichsen I, Schafer D, Langer D, Koger N, Wittmann M, Aretz S, Steinke V, Holzapfel S,
582 Trojan J, Konig R, Zeuzem S, Brieger A, Plotz G: Functional testing strategy for coding
583 genetic variants of unclear significance in MLH1 in Lynch syndrome diagnosis.
584 Carcinogenesis 2015, 36:202-211.
585 37. Drost M, Zonneveld JB, van Hees S, Rasmussen LJ, Hofstra RM, de Wind N: A rapid and
586 cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6
587 missense variants. Human mutation 2012, 33:488-494.
588 38. Ollila S, Sarantaus L, Kariola R, Chan P, Hampel H, Holinski-Feder E, Macrae F, Kohonen-
589 Corish M, Gerdes AM, Peltomaki P, Mangold E, de la Chapelle A, Greenblatt M, Nystrom
590 M: Pathogenicity of MSH2 missense mutations is typically associated with impaired
591 repair capability of the mutated protein. Gastroenterology 2006, 131:1408-1417.

592 39. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X,
593 Starita LM, Shendure J: Accurate classification of BRCA1 variants with saturation
594 genome editing. Nature 2018, 562:217-222.
595 40. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S, Shendure J,
596 Parvin JD: A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More
597 Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. American journal
598 of human genetics 2018, 103:498-508.
599 41. Umar A, Boyer JC, Thomas DC, Nguyen DC, Risinger JI, Boyd J, Ionov Y, Perucho M, Kunkel
600 TA: Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines
601 exhibiting microsatellite instability. The Journal of biological chemistry 1994, 269:14367-
602 14370.
603 42. Ollila S, Dermadi Bebek D, Jiricny J, Nystrom M: Mechanisms of pathogenicity in human
604 MSH2 missense mutants. Human mutation 2008, 29:1355-1363.
605 43. Brieger A, Trojan J, Raedle J, Plotz G, Zeuzem S: Transient mismatch repair gene
606 transfection for functional analysis of genetic hMLH1 and hMSH2 variants. Gut 2002,
607 51:677-684.
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631

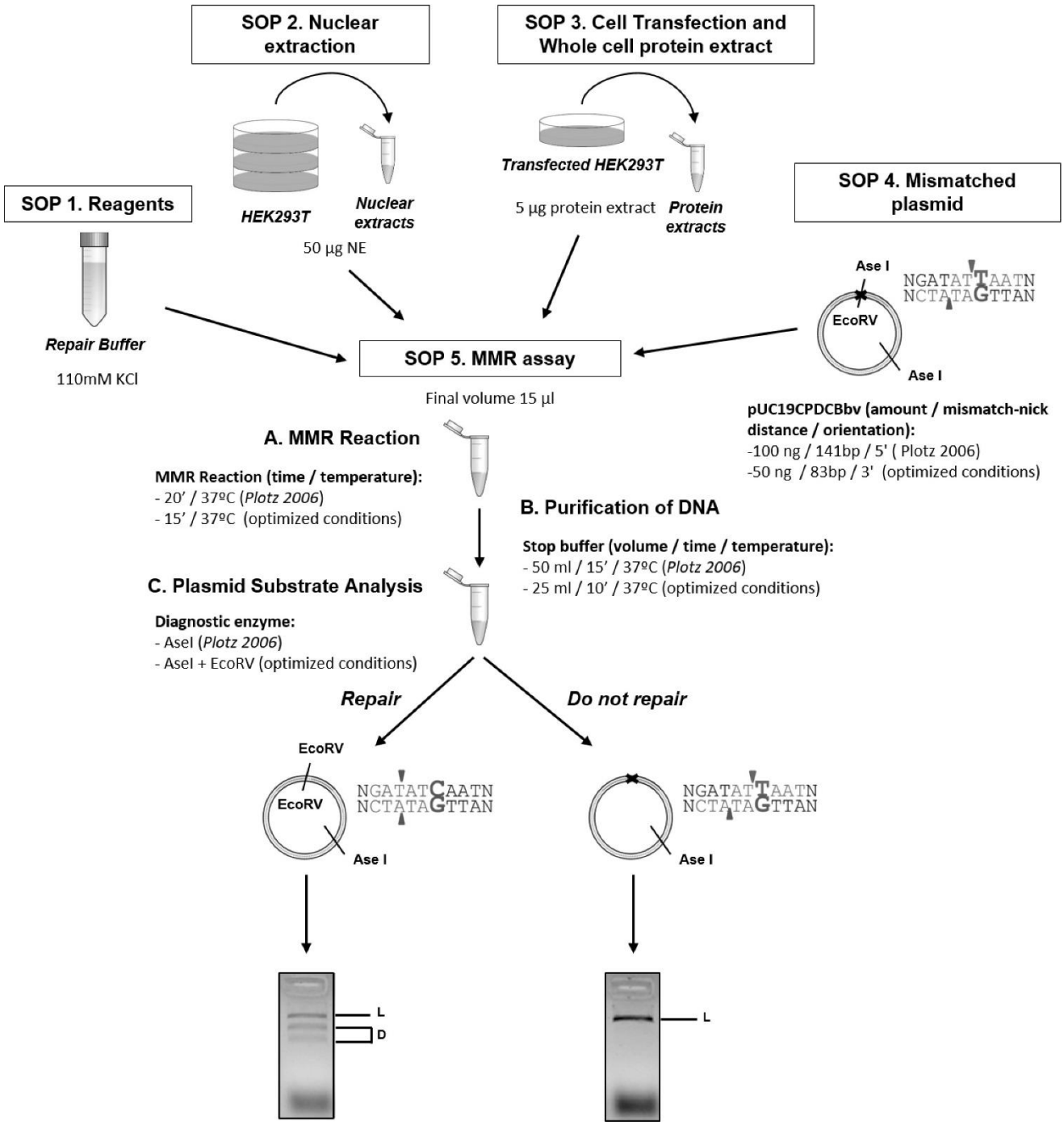
632 **FIGURE LEGENDS**

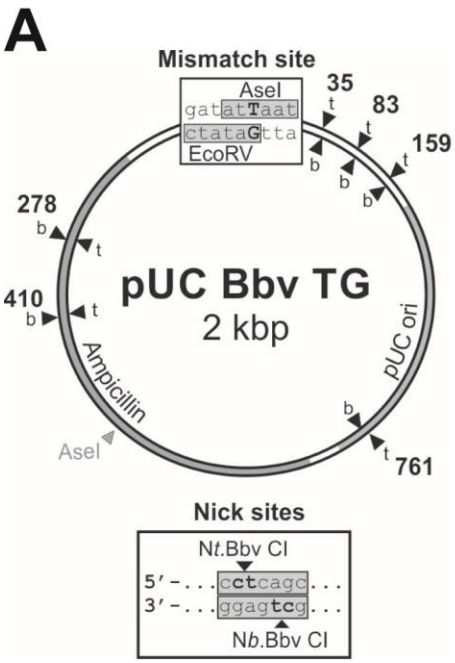
633 **Figure 1. *In vitro* MMR assay scheme.**
634 The standardized operating procedures (SOP) used in each step are indicated. The reagents and
635 parameters that have been optimized in this work are detailed together with those previously
636 reported by Plotz and collaborators (18). L, linearized vector DNA; D, double-digested vector
637 DNA.

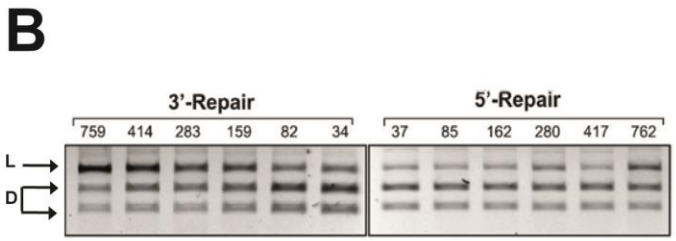
638 **Figure 2. Optimization of the mismatched plasmid.**
639 (A), Mismatch-nick distances and orientation for specific mismatched plasmids evaluated in this
640 work. (B), Agarose gel showing digestion products obtained in the *in vitro* MMR assay of nuclear
641 extracts complemented with 10% wild-type MLH1/PMS2-transfected cell extracts using the
642 designed mismatched plasmids. L, linearized vector DNA; D, double-digested vector DNA. (C),
643 Absolute repair obtained for each mismatched plasmid in HEK293T nuclear extracts (white
644 circles) and nuclear extracts complemented with 10% wild-type MLH1/PMS2-transfected cells
645 extracts (black circles). HEK293T nuclear extracts were tested in parallel to assess the
646 MLH1/PMS2-independent repair capability of the nuclear extracts.

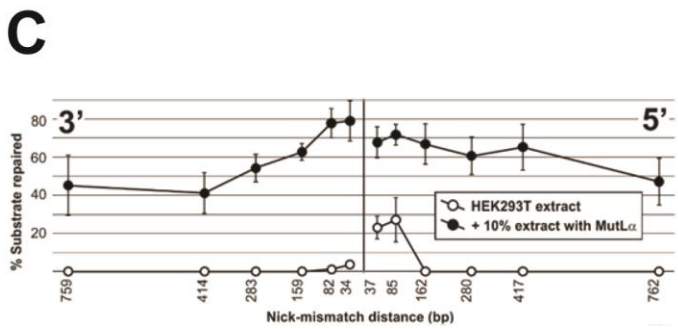
647 **Figure 3. Evaluation of the amount of whole cell protein extract, salt concentration and**
648 **reaction time.**
649 (A), Percentage of substrate repaired (absolute repair) obtained from testing different amounts
650 of whole protein extracts from wild-type MLH1/PMS2 transfected HEK293T cells. (B), Evaluation
651 of the salt concentration dependency by using different KCl concentrations from 40 to 180 mM.
652 For each condition repair levels of the wild-type MLH1/PMS2 is shown in direct comparison with
653 wild-type MLH1/PMS2 at 110 mM KCl, which is the currently used concentration. (C), Effect of
654 the incubation time of the MMR reaction on the repair efficiency. (D), Effect of the preincubation
655 of the MMR reaction on ice before the addition of MLH1/PMS2 proteins (MutLα complex) on
656 repair efficiency.

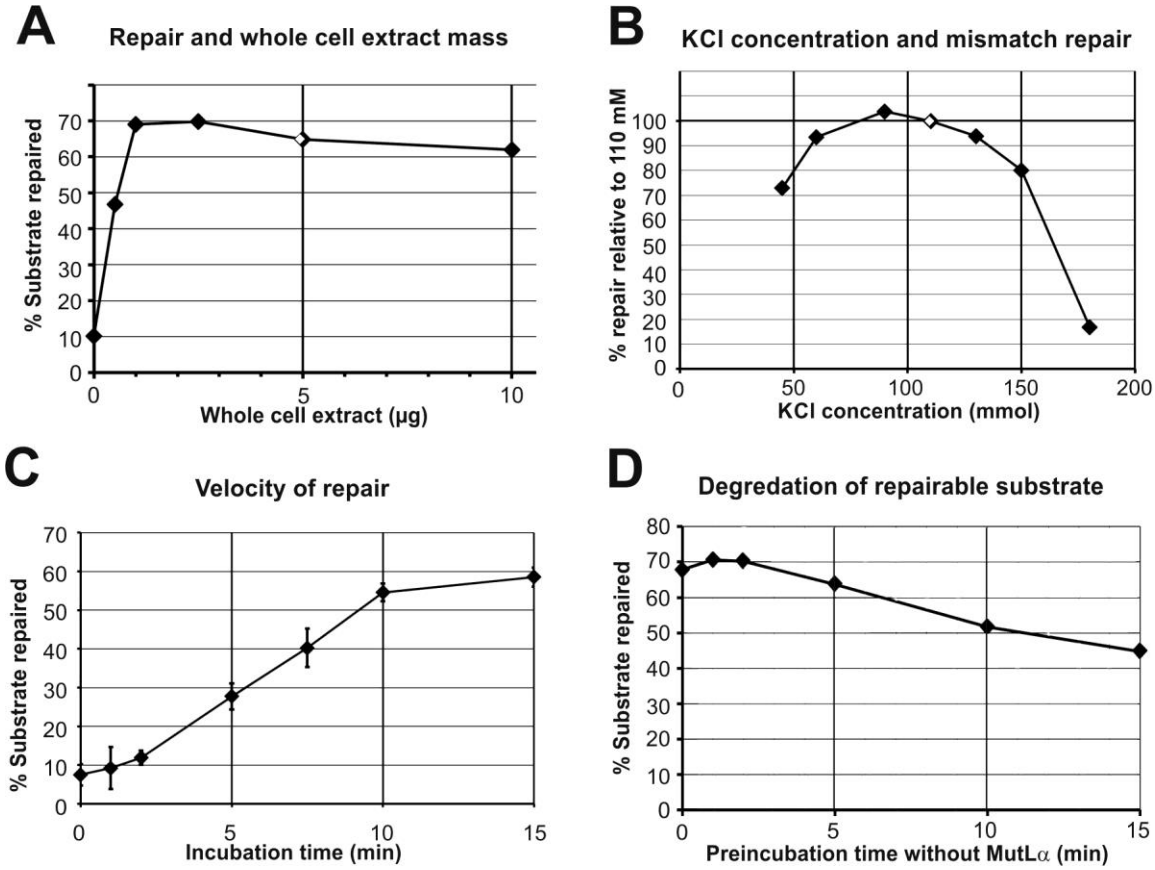
657 **Figure 4. Evaluation of the variability and reproducibility of the MMR assay.**
658 (A), Result of the intraexperimental variability analysis of the *in vitro* MMR assay by assessing
659 the absolute repair of wild-type MLH1/PMS2 proteins. Statistically significant differences
660 between experiments (*, $P<0.05$) and no significances (ns, $P > 0.05$) are indicated. (B), Result of
661 the interexperimental variability of the *in vitro* MMR assay by assessing the relative repair of
662 MLH1 p.I219V and p.G67R variants. No statistically significant differences were found between
663 the groups ($P > 0.05$). (C), Result of the reproducibility analysis of the *in vitro* MMR assay
664 between two independent laboratories. Left panel: Relative repair levels of the analyzed *MLH1*
665 variants. No statistically significant differences were found between the groups ($P > 0.05$). Right
666 panel: representative agarose gel showing digestion products of the MMR assay. Absolute repair
667 obtained in non-transfected cells in HCP and BF laboratories was 9.17±0.65% and 8.87±2.39%
668 respectively. D, double-digested DNA; L, linear DNA; HCP, Hereditary Cancer Program's lab; BF,
669 Biomedizinisches Forschungslabor's lab.

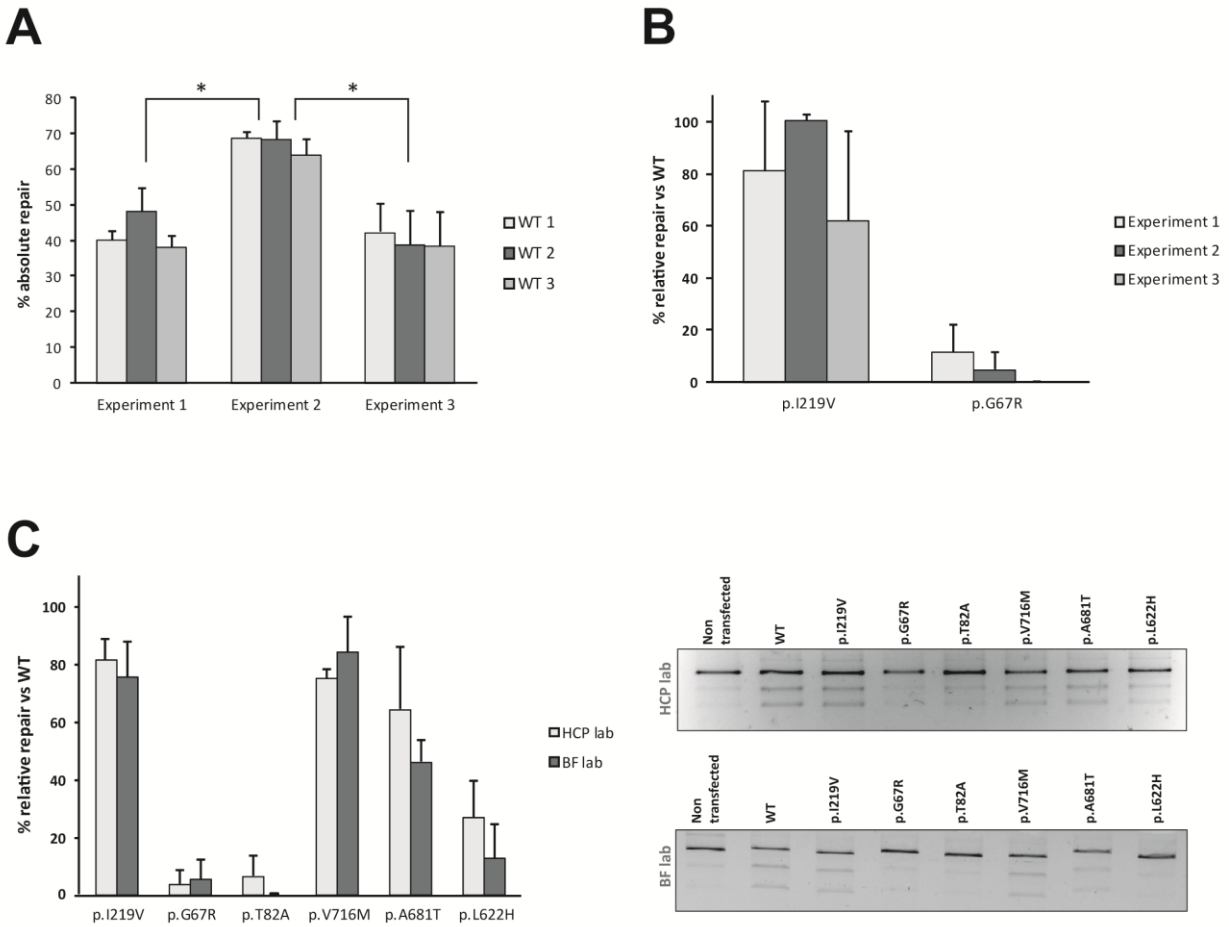

SOP 2. Nuclear extraction
SOP 3. Cell Transfection and Whole cell protein extract
SOP 1. Reagents
HEK293T
Nuclear extracts
50 µg NE
Transfected HEK293T
5 µg protein extract
Protein extracts
SOP 4. Mismatched plasmid
Ase I
NGATATTAATN
NCTATAGTTAN
EcoRV
Ase I
Repair Buffer
110mM KCl
SOP 5. MMR assay
Final volume 15 µl
pUC19CPDCBbv (amount / mismatch-nick distance / orientation):
-100 ng / 141bp / 5' (Plotz 2006)
-50 ng / 83bp / 3' (optimized conditions)
A. MMR Reaction
MMR Reaction (time / temperature):
- 20' / 37ºC (Plotz 2006)
- 15' / 37ºC (optimized conditions)
B. Purification of DNA
Stop buffer (volume / time / temperature):
- 50 ml / 15' / 37ºC (Plotz 2006)
- 25 ml / 10' / 37ºC (optimized conditions)
C. Plasmid Substrate Analysis
Diagnostic enzyme:
- AseI (Plotz 2006)
- AseI + EcoRV (optimized conditions)
Repair
Do not repair
EcoRV
EcoRV
NGATATCAATN
NCTATAGTTAN
Ase I
NGATATTAATN
NCTATAGTTAN
Ase I
L
D
L


A
Mismatch site
AseI
gatatTaat
ctataGtta
EcoRV
35
83
159
278
410
761
pUC Bbv TG
2 kbp
Ampicillin
pUC ori
AseI
Nick sites
Nt.Bbv CI
5'-...cctcagc...
3'-...ggagtcg...
Nb.Bbv CI


B
3'-Repair
5'-Repair
759 414 283 159 82 34
37 85 162 280 417 762
L
D


C
3'
5'
% Substrate repaired
80
60
40
20
HEK293T extract
+ 10% extract with MutLα
759 414 283 159 82 34 37 85 162 280 417 762
Nick-mismatch distance (bp)


A
Repair and whole cell extract mass
% Substrate repaired
Whole cell extract (µg)
B
KCl concentration and mismatch repair
% repair relative to 110 mM
KCl concentration (mmol)
C
Velocity of repair
% Substrate repaired
Incubation time (min)
D
Degredation of repairable substrate
% Substrate repaired
Preincubation time without MutLα (min)


A
% absolute repair
80
70
60
50
40
30
20
10
0
*
*
Experiment 1
Experiment 2
Experiment 3
WT 1
WT 2
WT 3
B
% relative repair vs WT
100
80
60
40
20
0
p.I219V
p.G67R
Experiment 1
Experiment 2
Experiment 3
C
% relative repair vs WT
100
80
60
40
20
0
p.I219V
p.G67R
p.T82A
p.V716M
p.A681T
p.L622H
HCP lab
BF lab
Non transfected
WT
p.I219V
p.G67R
p.T82A
p.V716M
p.A681T
p.L622H
HCP lab
BF lab

Supplemental Table S1. Selected *MLH1* MMR variants analyzed in this study and summary of results. Colored cells indicate interpretation of functional analyses following the recommendations of the Variant Interpretation Commitee of InSIGHT (Thompson et al., 2014): (*) abrogated MMR activity <35% (red highlighted cells), proficient MMR activity >64% (green cells), intermediate activity 35-65% (grey cells); (†) decreased expression <25% (red highlighted cells), normal expression >75% (green cells), intermediate expression 25-75% (grey cells). (‡) Different total protein extracts were used in the functional assays: baculovirus infected insect cells extracts were used in MMR assay whereas cell extracts from transfected MMR-deficient cells were used in protein expression assessment.

MLH1 variants	Population MAF (1000genomes / ESP EA / ExAc)	Functional assessment			Multifactorial likelihood analysis	InSiGHT Classification (date and version)	Reasoning for variant classification
		Reference	MMR activity (% normalized to WT)*	MLH1 expression (% normalized to WT)†			
c.655A>G (p.I219V)	0.13/0.3202/0.3176 (rs1799977)	Takahashi 2007 [24]	60.7%	25-75%	NP	**Class 1: Not pathogenic (2013/09/05 v1.9)**	Minor allele frequency in general population >1%. Proficient MMR activity and protein expression.
		Raevaara 2005 [23]‡	88%	Normal			
		Borràs 2010 [26], 2012 [29]	95%	≈120%			
		Hinrichsen 2013 [30]	96%	94%			
		Drost 2010 [25]	80% (cloned), 70% (PCR)				
		Current study	***78%±9.6***	***77,5% ±13,9***			
c.199G>A (p.G67R)	NR/NR/NR (rs63750206)	Takahashi 2007 [24]	5.9%	<25%	Posterior probability of pathogenicity >0.99	**Class 5: Pathogenic (2013/09/05 v1.9)**	PPP >0.99. Deficient MMR activity and protein expression
		Raevaara 2005 [23]‡	6%	Reduced			
		Borràs 2010 [26], 2012 [29]	5%	20%			
		Drost 2010 [25]	5% (cloned), 5% (PCR)				
		Current study	***4.4% ±5.1***	***32,5% ±8,1***			
c.244A>G (p.T82A)	NR/1.499e-05/NR (rs587778998)	Borràs 2012 [29]	12.2%	140%	Posterior probability of pathogenicity >0.95	**Class 4: likely pathogenic 2018/06/13 v2.4)**	PPP >0.95. Deficient MMR activity.
		Current study	***3.2% ±5.5***	***66% ±10,7***			
c.2146G>A (p.V716M)	0.00/0.0021/0.0018 (rs35831931)	Takahashi 2007 [24]	75.1%	25-75%	Posterior probability of pathogenicity <0.001	**Class 1: Not pathogenic (2013/09/05 v1.9)**	PPP <0.001. Proficient MMR activity. Intermediate expression.
		Raevaara 2005 [23]‡	110%	Normal			
		Hinrichsen 2013 [30]	86%	65%			
		Drost 2010 [25]	110% (cloned), 65% (PCR)				
		Current study	***79.3% ±9.4***	***60,8% ±9,2***			
c.2041G>A (p.A681T)	NR/NR/NR (rs63750217)	Takahashi 2007 [24]	69.8%	75%	Posterior probability of pathogenicity >0.99	**Class 5: Pathogenic (2013/09/05 v1.9)**	PPP >0.99. Inconclusive MMR activity. Intermediate protein expression.
		Raevaara 2005 [23]‡	115%	Normal			
		Hinrichsen 2013 [30]	99%	51%			
		Hardt 2011 [28]		41%			
		Current study	***54.9% ±17.6***	***33,3% ±17,5***			
c.1865T>A (p.L622H)	NR/NR/NR	Takahashi 2007 [24]	69.2%	25-75%	Posterior probability of pathogenicity >0.99	**Class 5: Pathogenic (2013/09/05 v1.9)**	PPP >0.99. Inconclusive MMR activity. Intermediate protein expression.
		Borràs 2010 [26]		25-48% (48h), 0-19% (72h)			
		Hinrichsen 2013 [30]	80%	42%			
		Kosinski 2010 [27]		25-75%			
		Current study	***19.5% ±15.5***	***15,7% ±12,6***			

Abbreviations: NR, not reported; NP, not performed.

Supplemental Table S2. Description of the methologies reported in the assessment of MMR activity and expression of the *MLH1* variants analyzed in the current study.

Reported study	*In vitro* MMR Assay											Analysis of MMR variant expression					
	Nuclear extracts	Generation of MMR variants					Mismatched subtrate	MMR assay									
	Cell line	Cell line	Protein production method	Cloning vector	Reagent	Post-transfection incubation time		Protein extracts	Mismatched plasmid amount	MMR reaction time	Quantification of the MMR efficiency	Cell line	Protein production method	Cloning vector	Reagent	Post-transfection incubation time	Evaluation of protein expession
Raevaara 2005 [23]	HCT116	Sf9 cells (insect)	Baculovirus expression system	pFastBac1		48h	DNA heteroduplex nicked at 369bp 3' to the mismatch	75 µg (NE) / 2.3 µg (TPE)	100 ng	30'	Analysis of bands intensity in agarose gel	HEK293T	Transient transfection	pMLH1-N1	FuGENE6	48h	Western blot
Takahashi 2007 [24]	HCT116	HCT116	Transient transfection	pCMV-Neo-Bam	FuGENE6	24h	M13mp2 substrate (described in *Thomas et al., 1991*)	75 µg (TPE)	5 ng	60'	Analysis of bands intensity in agarose gel	HCT116	Transient transfection	pCMV-Neo-Bam	FuGENE6	24h	Western blot
Drost 2010 [25]	HCT116	–	*In vitro* trancription and translation	pCITE4a	–	–	pJHGT30'InFAM substrate nicked at 138bp 3' to the mismatch	75 µg (NE) / 12 µl (ivP)	100 ng	40'	Analysis of fluorescent fragments	NP	NP	NP	NP	NP	NP
Borràs 2010 [26]	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	HCT116	Transient transfection	pcDNA3	Lipofectamine	48h	Western blot
Kosinski 2010 [27]	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	HEK293T	Transient transfection	pcDNA3	PEI	48h	Western blot
Hardt 2011 [28]	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	HEK293T	Transient transfection	pMCC-574	FuGENE6	48h	Western blot
Borràs 2012 [29]	HEK293T	HEK293T	Transient transfection	pcDNA3.1	PEI	48h	pUC19CPDCBbv substrate nicked at 83bp 3' to the mismatch	50 µg (NE) / 5 µg (TPE)	100 ng	20'	Analysis of bands intensity in agarose gel	HEK293T	Transient transfection	pcDNA3.1	PEI	48h	Western blot
Hinrichsen 2013 [30]	HEK293T	HEK293T	Transient transfection	pcDNA3	PEI	24h/48h	pUC19CPDCBbv substrate nicked at 83bp 3' to the mismatch	50 µg (NE) / 5 µg (TPE)	35 ng	20'	Analysis of bands intensity in agarose gel	HEK293T	Transient transfection	pcDNA3	PEI	24h/48h	Western blot
Current study	**HEK293T**	**HEK293T**	**Transient transfection**	**pcDNA3.1**	**PEI**	**48h**	**pUC19CPDCBbv substrate nicked at 83bp 3' to the mismatch**	**50 µg (NE) / 5 µg (TPE)**	**50 ng**	**15'**	**Analysis of bands intensity in agarose gel**	**HEK293T**	**Transient transfection**	**pcDNA3.1**	**PEI**	**48h**	**Western blot**

Abbreviations: PEI, polyethylenimine; NP, not performed; NE, nuclear extracts; TPE, total protein extracts; ivP, *in vitro* produced MMR protein.

Supplemental Table S3. pUC19CPDC mismatched plasmids tested in the optimization of *in vitro* MMR assay.

Type of pUC19CPDC plasmid	Plasmid nicked at 3'			Plasmid nicked at 5'		
	Enzyme used to perform the nick	Nick-mismatch distance (bp)	Diagostic enzyme in MMR assay	Enzyme used to perform the nick	Nick-mismatch distance (bp)	Diagostic enzyme in MMR assay
pUC Bbv GT 35	Nt.BbvCI	34	EcoRV	Nb.BbvCI	37	AseI
pUC Bbv GT 83	Nt.BbvCI	82	EcoRV	Nb.BbvCI	85	AseI
pUC Bbv GT 159	Nt.BbvCI	159	EcoRV	Nb.BbvCI	162	AseI
pUC Bbv GT 278	Nt.BbvCI	283	EcoRV	Nb.BbvCI	280	AseI
pUC Bbv GT 410	Nt.BbvCI	414	AseI	Nb.BbvCI	417	EcoRV
pUC Bbv GT 761	Nt.BbvCI	759	EcoRV	Nb.BbvCI	762	AseI

Supplemental Table S4. Quality control measures included in Standard Operating Procedures (SOP).

SOP #	Quality control mesure	Description	In-Process Control number (*IPC*):
1- Reagents	**Use of reference reagents**		
2- Nuclear Extraction	**Cell lysis monitoring**	Trypan blue staining allows to verify that cells are intact (not blue-stained) before the lysis step (*IPC1*) and lysed (blue-stained) after the lysis step (*IPC2*)	*IPC1, IPC2*
	Nuclear protein extract enrichment	Western blotting against a nuclear marker allows to verify the enrichment in nuclear extract proteins in the nuclear extract fraction (*IPC5*) compared to cytoplasmatic (*IPC3*) and nuclear membrane (*IPC4*) fractions	*IPC3, IPC4, IPC5*
	Nuclear extract protein concentration	Minimum concentration: 3 µg/µl	
	Verification of repair capability	Verify by in vitro MMR assay the repair capability of the nuclear extracts by complementation with cytoplasmatic MMR proficient cytoplasmic extracts. Minimum absolute repair activity: 25%	
3- Cell Transfection and Whole Cell Protein Extract	**Transfection efficiency**	Minimum level of transfection: 40%	
	Protein concentration	Minimum concentration: 3 µg/µl	
4- Mismatched Plasmid Substrate Generation	**Double nick of the strand**	Check that supercoiled plasmid band completely disappears by running in agarose gel (*IPC1*)	*IPC 1*
	Original oligomer removal	Verify that the flowthrough from sample washings (*IPC2, IPC3, IPC4*) only contains the oligomer by running in agarose gel	*IPC2, IPC3, IPC4*
	Mismatch oligomer annealing and ligation	Verify by running in agarose gel that the plasmid has the circular conformation before the annealing and ligation of the mismatched oligomer (*IPC5*)and the ligated conformation after this process (*IPC6*)	*IPC5, IPC6*
	Elimination of the original plasmid (without mismatch)	Digest the sample with EcoRV and ExonucleaseV to eliminate the original plasmid (which is EcoRV-sensitive) and verify this process in agarose gel (*IPC7, IPC8, IPC9*)	*IPC7, IPC8, IPC9*
	Second verification of elimination of the original plasmid	Verify the complete elimination of the original plamid by digestion with EcoRV and AseI and running in agarose gel (*IPC10*). Only the linearized band should appear.	*IPC10*
	Verification of the introduction of 3'-single strand break	Check by running in agarose gel that the plasmid has the circular conformation after digestion with Nt.Bbv CI (*IPC11*).	*IPC11*
	Mismatched plasmid production	Minimum DNA concentration: 3 ng/µl	
5- MMR Assay	**Absolute repair activity in non-transfected group**	Absolute repair in the non-transfected group should be <10%	
	Absolute repair activity in wild-type MLH1/PMS2 group	Absolute repair in the wild-type MLH1/PMS2 group should be >25%	
	Relative repair of p.I219V proficient control variant	No significant differences with wild-type variant	
	Relative repair of p.G67R deficient control variant	No significant differences with non-transfected group	

Table S5: Sensitivity and specificity of different MMR assay approaches to correctly classify *MLH1* variants. Only variants not affecting splicing and currently classified by InSIGHT as class 1-2 (true neutral) and 4-5 (true pathogenic) have been taken into account for calculations. Following the recommendations of the Variant Interpretation Commitee of InSight for interpretation of MMR assay results (Thompson et al., 2014), cut-offs were set at <35% (abrogated function) and >64% (normal function).

A. Sensitivity and specificity to detect pathogenic variants

	Raevaara 2005 (n=22) [23]		**Takahashi 2007 (n=43) [24]**		**Drost 2010 (n=14) [25]**		**Hinrichsen 2013 (n=20) [30]**		**Current study (n=6)**	
Relative MMR activity:	True pathogenic	True neutral	True pathogenic	True neutral	True pathogenic	True neutral	True pathogenic	True neutral	True pathogenic	True neutral
<35%	9	0	20	1	8	0	9	0	3	0
>35%	8	5	6	16	1	5	6	5	1	2
Sensitivity (95% CI)	*52.9% (27.81-77.02%)*		*76.9% (56.35-91.03%)*		*88.9% (51.75-99.72%)*		*60.0% (32.29-83.66%)*		*75.0% (19.41-99.37%)*	
Specificity (95% CI)	*100% (63.06-100%)*		*94.1% (71.31-99.85%)*		*100% (47.82-100%)*		*100% (47.82-100%)*		*100% (15.81-100%)*	

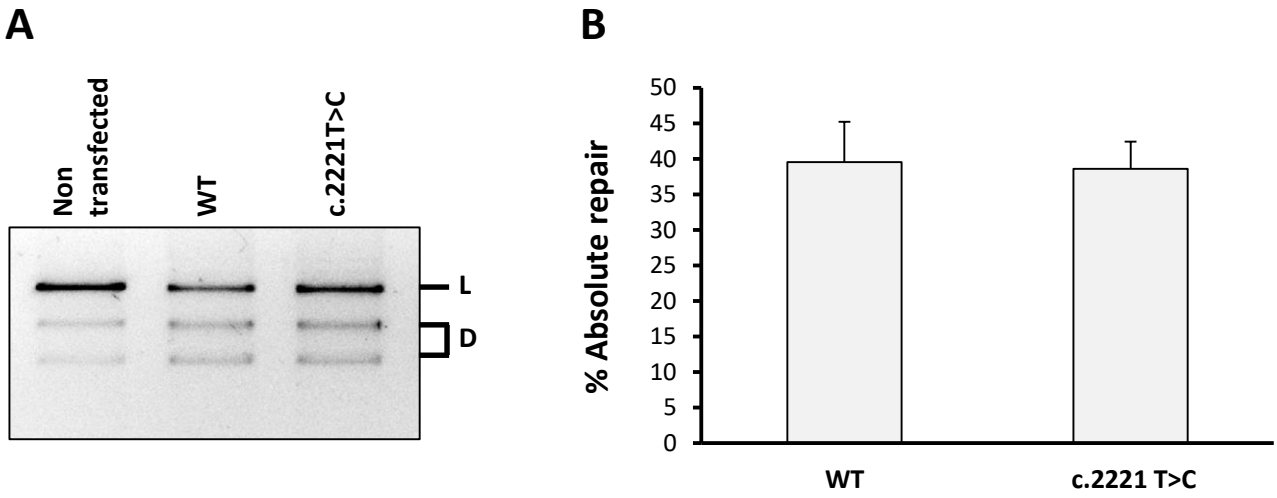
B. Sensitivity and specificity to detect neutral variants

	Raevaara 2005 (n=22) [23]		**Takahashi 2007 (n=43) [24]**		**Drost 2010 (n=14) [25]**		**Hinrichsen 2013 (n=20) [30]**		**Current study (n=6)**	
Relative MMR activity:	True neutral	True pathogenic	True neutral	True pathogenic	True neutral	True pathogenic	True neutral	True pathogenic	True neutral	True pathogenic
>64%	5	6	13	2	5	0	5	5	2	0
<64%	0	11	4	24	0	9	0	10	0	4
Sensitivity (95% CI)	*100% (47.82-100%)*		*76.5% (50.10-93.19%)*		*100% (47.82-100%)*		*100% (47.82-100%)*		*100% (15.81-100%)*	
Specificity (95% CI)	*64.7% (38.33-85.79%)*		*92.3% (74.87-99.05%)*		*100% (66.37-100%)*		*66.7% (38.38-88.18%)*		*100% (39.76-100%)*	

Supplemental Table S6. Advantages and limitations of *in vitro* MMR assay approaches used in the assessment of *MLH1* variants.

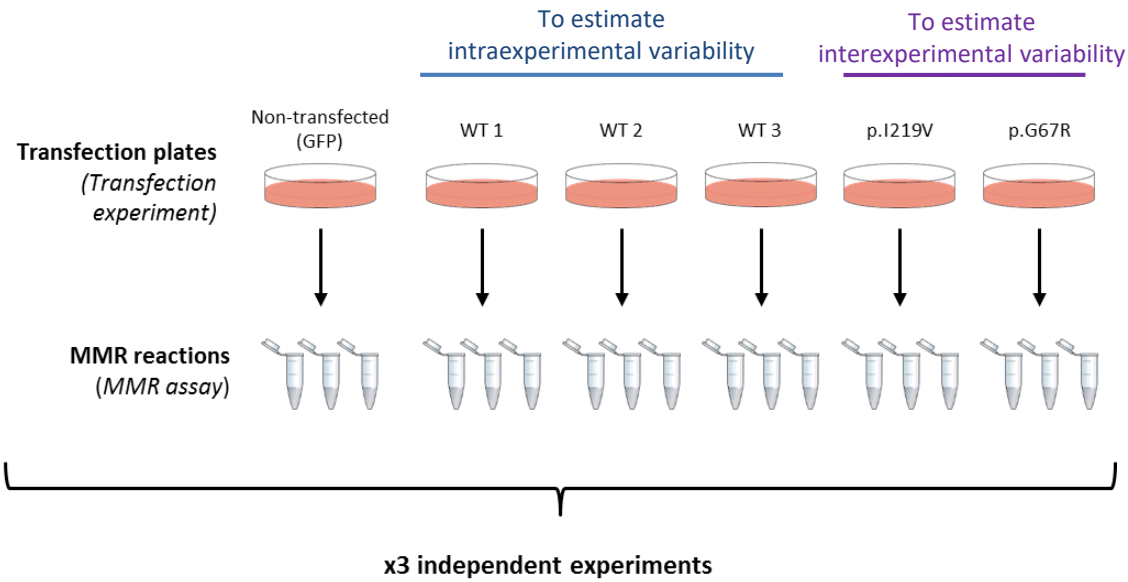
	***Cell-free in vitro* MMR assay** (current study)	***Cell-free in vitro* MMR assay** (Takahashi 2007 [24]; Borràs 2012 [29]; Hinrichsen 2013 [30])	***Cell-free in vitro* MMR assay** (Raevaara 2005 [23])	**Cell-free *in vitro* MMR activity (CIMRA) assay** (Drost 2010 [25], Drost 2018 [20])
Advantages	Available standardized protocols. Tested analytical variability and reproducibility. The protein extract (obtained after transcient transfection) can also be used to assess protein expresion.	The protein extract (obtained after transcient transfection) can also be used to assess protein expresion.	Baculovirus expression system allows the production of large amounts of protein.	Quantification of the MMR efficiency by a highly sensitive method. CIMRA results in combination with *in silico* predictions have recently been integrated in multifactorial likelihood models.
Limitations	Results have not been calibrated for integration in multifactorial models. Cut-offs for pathogenicity have not been established	Results have not been calibrated for integration in multifactorial models. Cut-offs for pathogenicity have not been established.	The same protein extract can not be used to assess protein expresion. Results have not been calibrated for integration in multifactorial models. Cut-offs for pathogenicity have not b.een established	Theprotein extract obtained by in vitro transcription and translation can not be used to assess variant expresion.

Supplemental Figure S1. *In vitro* MMR activity of the *MLH1* silent variant c. 2221T>C. A: Agarose gel showing digestion products obtained in the MMR assay. Absolute repair obtained in non-transfected group was 9.62±1.99%. D, double-digested DNA vector; L, linearized DNA vector. **B:** Quantification of absolute repair of wild-type *MLH1* and c.2221T>C variant. No statistically significant differences were found (P > 0.05). Data was obtained from three independent experiments.

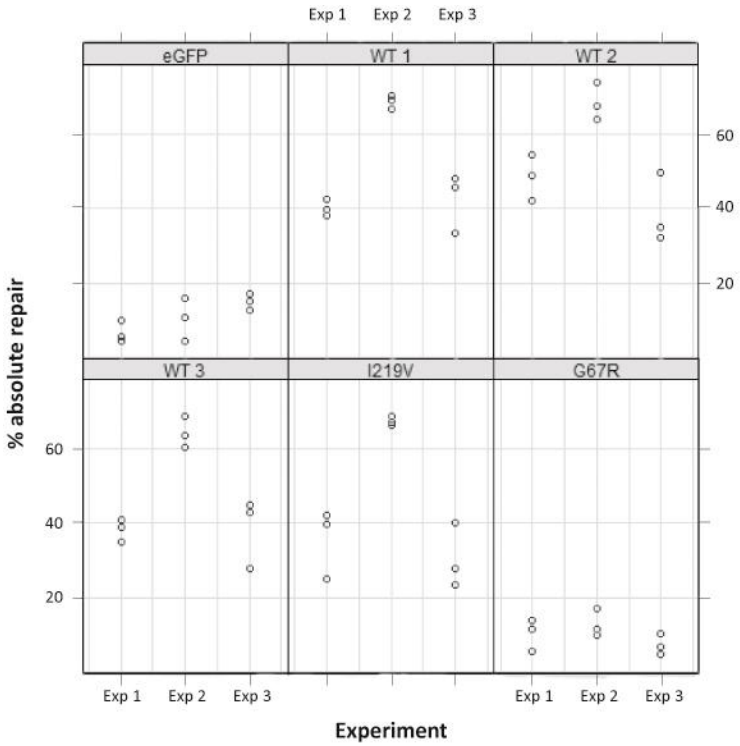


Supplemental Figure S2. Experimental design to evaluate the intra- and inter-experimental variability of the *in vitro* MMR assay and obtained results. A: Schematic representation of the experimental design to evaluate the intra- and inter-experimental variability of the MMR assay. Each transfection experiment includes 1 plate transfected with eGFP plasmid (negative control), 3 independent plates transfected with wild-type MLH1/PMS2 plasmid, 1 plate transfected with I219V-MLH1/PMS2 plasmid (proficient control) and 1 plate transfected with G67R-MLH1/PMS2 plasmid (deficient control). The MMR assay reaction performed by using the protein extracts from each plate was performed in triplicate. Three independent transfections were performed. **B:** Absolute repair obtained for each sample. WT, wild-type; exp, experiment.

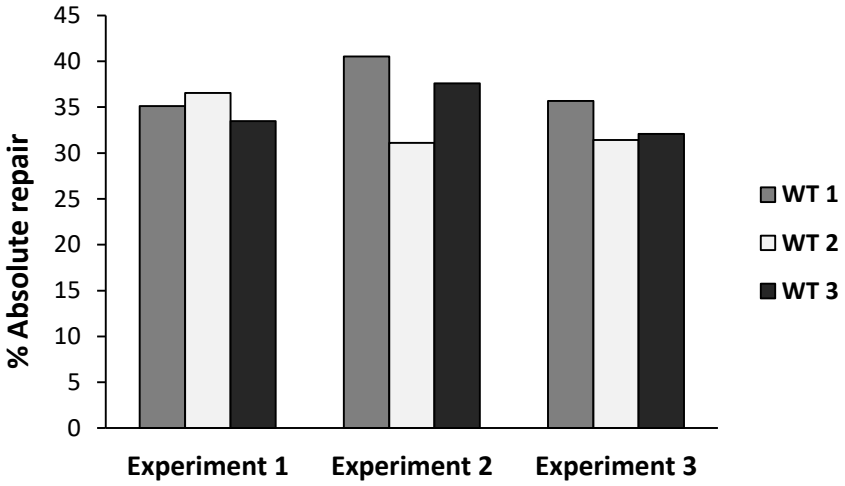
A



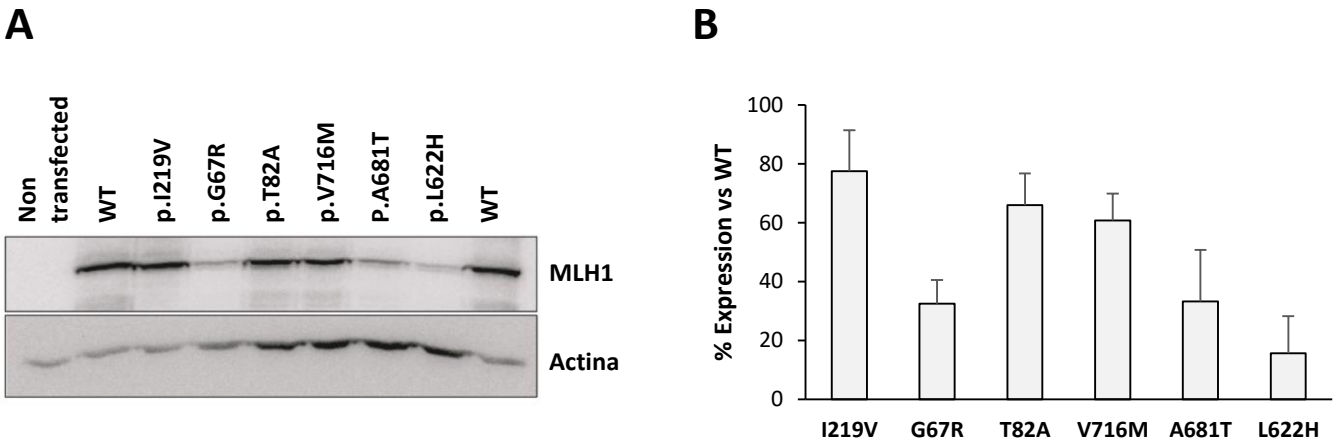
B



Supplemental Figure S3. Intraexperimental variability of the *in vi*tro MMR assay using the same nuclear extract preparation. Quantification of repair levels of wild-type (WT) proteins in terms of absolute repair using the same nuclear extract preparation. No statistically significant differences were found between the WT groups (P > 0.05).



Supplemental Figure S4. Protein expression levels of MLH1 variants. A: Western-blot analysis of the 6 MLH1 variants analyzed in this study. **B:** Quantification of expression levels of MLH1 variants in direct comparison to MLH1 wild-type. Data was obtained from three independent experiments.



Protocol 1 Reagents

Reagents for Nuclear Extraction (Protocol 2)

Protocol 1.1 Hypotonic Buffer

HEPES pH 7.6	20 mM
KCl	5 mM
$MgCl_2$	0,5 mM
DTT	0,5 mM
PMSF	0,1%

Store at +4°C.

PREPARE as follows/LABEL with:

Hypotoic Buffer (Protocol 1.1)

Date prepared: _____

HEPES KOH 250 mM pH 7.6 (***P. 1.7***)	4 ml
KCl 4 M (***Protocol 1.5***)	62.5 µl
$MgCl_2$ 250 mM (***Protocol 1.8***)	0.10 ml
Fill up with water to **50 ml**.	

Store at +4°C

Add before use:

PMSF 1000X (***Protocol 1.9***)	1 µl/ml
DTT 200 mM (***Protocol 1.10***)	2.5 µl/ml

Protocol 1.2 Resuspension Buffer

HEPES KOH pH 7.6	50 mM
Sucrose	10% (w/v)
DTT	1 mM
PMSF	0,2%

Store at +4°C.

PREPARE as follows/LABEL with:

Resuspension Buffer (Protocol 1.2)

Date prepared: _____

HEPES KOH 250 mM pH 7.6 (***P. 1.7***)	10 ml
Sucrose	5 g
Fill up with water to **50 ml**.	

Store at +4°C

Add before use:

PMSF 1000X (***Protocol 1.9***)	2 µl/ml
DTT 200 mM (***Protocol 1.10***)	5 µl/ml

Protocol 1.3 High Salt Buffer

HEPES KOH pH 7.6	50 mM
Sucrose	10%(w/v)
KCl	840 mM

Store at +4°C.

PREPARE as follows/LABEL with:

High Salt Buffer (Protocol 1.3)

Date prepared: _____

HEPES KOH 250 mM pH 7.6 (***P. 1.7***)	2 ml
Sucrose	1 g
KCl 4 M (***Protocol 1.5***)	2.1 ml
Fill up with water to **10 ml**.	

Store at +4°C

Protocol 1.4 Dialysis Buffer for Nuclear Extracts

HEPES KOH pH 7.6	25 mM
KCl	100 mM
EDTA	0.01 mM
DTT	1 mM
PMSF	0.1%

Store at +4°C.

PREPARE as follows/LABEL with:

Dialysis Buffer for Nuclear Extracts (Protocol 1.4)

Date prepared: _____

HEPES KOH pH 7.6 250 mM (***P. 1.7***)	100 ml
KCl (Mw 74.55)	7.46 g
EDTA 100 mM (***Protocol 1.6***)	100 µl
Fill up with water to **1L**.	

Store at +4°C

Add before use:

PMSF 1000X (***Protocol 1.9***)	1 µl/ml
DTT 200 mM (***Protocol 1.10***)	2.5 µl/ml

Protocol 1.5 KCl 4 M

PREPARE as follows/LABEL with:

KCl 4 M (Protocol 1.5)
Date prepared:_____.
14.92 g KCl (Mw 74.55) on 50 ml water.
Store at room temperature.

Protocol 1.6 EDTA 100 mM

PREPARE as follows/LABEL with:

EDTA 100 mM (Protocol 1.6)
Date prepared:_____.
EDTA (Di-Sodiumsalt, dihydrate, e.g.
AppliChem #A2937, Mw 372.24 g/mol) 372.24 mg
Dissolve in 10 ml water.
Store at room temperature.

Protocol 1.7 HEPES-KOH 250 mM pH 7.6

PREPARE as follows/LABEL with:

HEPES-KOH 250 m*M* pH 7.6 (Protocol 1.7)
Date prepared:_____.
HEPES (Mw 238 g/mol) 59.52 g
Dissolve in 900 ml Water
Adjust pH with KOH to 7.6
Adjust volume with water to **1000 ml**.
Store at room temperature.

Protocol 1.8 $MgCl_2$ solution 250 mM

PREPARE as follows/LABEL with:

Magnesium chloride solution 250 m*M* (Protocol 1.8)
Date prepared:_____.
$MgCl_2$ Hexahydrate (e.g. Roth 2189.2) Mw 203.3 g/mol 12.71 g
Dissolve in **250 ml** water.
Store at room temperature.

Reagents for whole cell protein extracts (Protocol 3)

Protocol 1.9 PMSF solution 1000x (200 mM)

PREPARE as follows/LABEL with:

PMSF solution 1000X (200 mM) (Protocol 1.9)
Date prepared:_____.
PMSF (Mw 174.2 g/mol) **348 mg**
Dissolve in 10 ml 2-Propanol water-free (Sigma 278475)
Store at room temperature (precipitates in the cold)

Protocol 1.10 DTT solution 500x (200 mM)

PREPARE as follows/LABEL with:

DTT solution 500x (200 mM) (Protocol 1.10)
Date prepared:_____.
DTT (Mw 154.3 g/mol) **308 mg**
Dissolve in 10 ml water.
Store at -20°C.

Protocol 1.11 Buffer A (293 Whole Cell Extract)

HEPES KOH pH 7.6	20	mM
$MgCl_2$	5	mM
EDTA	0.1	mM
NaCl	10	mM

PREPARE as follows/LABEL with:

Buffer A 293 Whole Cell Extract (Protocol 1.11)
Date prepared: _____

HEPES KOH 250 mM pH 7.6 (***P. 1.7***)	8 ml	20 mM
NaCl 1 *M* (***Protocol 1.13***)	1 ml	10 mM
$MgCl_2$ 250 mM (***Protocol 1.8***)	2 ml	5 mM
EDTA 100 mM (***Protocol 1.6***)	0.1 ml	0.1 mM
Water	89.5 ml	

Store at +4°C

Add before use:

PMSF 1000X (***Protocol 1.9***)	1 µl/ml
DTT 200 mM (***Protocol 1.10***)	2.5 µl/ml

Protocol 1.12 Buffer C (293 Whole Cell Extract)

HEPES pH 7.6	20	mM
$MgCl_2$	5	mM
EDTA	0.1	mM
NaCl	840	mM
Glycerol	40	% (v/v)

PREPARE as follows/LABEL with:

Buffer C 293 Whole Cell Extract (Protocol 1.12)
Date prepared: _____

HEPES KOH 250 mM pH 7.6 (***P. 1.7***)	8 ml	20 mM
NaCl (substance)	4.91 g	840 mM
$MgCl_2$ 250 mM (***Protocol 1.8***)	2 ml	5 mM
EDTA 100 mM (***Protocol 1.6***)	0.1 ml	0.1 mM
Glycerol (substance, δ=1,26)	50.4 g	40 %(v/v)
Water ad	100 ml	

Store at +4°C

Add before use:

PMSF 1000X (***Protocol 1.9***)	1 µl/ml
DTT 200 mM (***Protocol 1.10***)	2.5 µl/ml

Protocol 1.13	**NaCl solution 1 *M***

PREPARE as follows/LABEL with:

NaCl 1 *M* (Protocol 1.13)

Date prepared:_____.

NaCl (Mw 58.44 g/mol) **2.922 g**

Dissolve in 50 ml water.

Store at room temperature.

Reagents for MMR Assay (Protocol 5)

Protocol 1.14 tRNA solution 1 mg/ml

PREPARE as follows/LABEL with:

tRNA solution 1 mg/ml (Protocol 1.14)
Date prepared:_____.
tRNA (from yeast, e.g. Roche #109517) 10 mg
Add 10 ml water. Aliquot to 1 ml.
Store at -20°C.

Protocol 1.15 KCl solution 500 m*M*

PREPARE as follows/LABEL with:

KCl 500 m*M* for MMR assay (Protocol 1.15)
Date prepared:_____.
KCl (Mw 74.55 g/mol) 373 mg
Dissolve in 10 ml water.
Store at room temperature.

Protocol 1.16 RNAse A solution 1 mg/ml

PREPARE as follows/LABEL with:

RNAse A solution 1 mg/ml (Protocol 1.16)
Date prepared:_____.
RNAse A (bovine, lyophilized: e.g. Carl Roth GmbH, #7156) 10 mg
Dissolve in 10 ml water
Store in 1 ml aliquots at -20°C.

Protocol 1.17 SDS solution 10%

PREPARE as follows/LABEL with:

SDS solution 10% (Protocol 1.17)
Date prepared:_____.
Sodium dodecyle sulfate 1 g
Dissolve in 10 ml water
Store at room temperature.

Protocol 1.18 MMR Repair Buffer 10x

Reagent	Source	Concentration in 10x Buffer	For 1 ml	Added?
T4 RNA ligation buffer	New England Biolabs #B0216S	250 mM Tris-HCl pH 7.5 50 mM $MgCl_2$ 5 mM DTT	500	
dCTP 100 mM	Promega #U123A	1 mM	10	
dGTP 100 mM	Promega #U123A	1 mM	10	
dTTP 100 mM	Promega #U123A	1 mM	10	
dATP 100 mM	Fermentas #R0441	15 mM	150	
BSA solution 10 mg/ml	New England Biolabs #B9001S	500 µg/ml	50	
Water		-	270	

Vortex, aliquot to 50 µl.
Store at -20°C.

Protocol 2 Nuclear Extraction

Material required: Dounce Homogenizer 1 ml or 7 ml, Pestle „Tight“, on ice
Spectra/Por CE dialysis tubing 3.5-5kDa MWCO, 10 mm flat width, 0.32 ml/cm
Cell centrifuge (15 ml tubes) at 4°C
Microcentrifuge (2 ml cups) at 4°C
Cold PBS
Cold Extraction Buffers (***Protocols 1.1 – 1.4***) and Supplements DTT and PMSF (***Protocols 1.9*** and ***1.10***)

DATE: _____
CELL LINE: _____ (typical values are given for HEK293/HEK293T cells)
NUMBER AND SIZE DISHES: _____ (typical: 8-15 145 cm dishes)
CONFLUENCY (%): _____ (typical: 80-100%)

CELL HARVEST

approximate time required for 10 dishes: 25 min

Have ready: Open for waste, papertowel for dripping, 15 ml tube on ice to approximately measure 10 ml of PBS, 50 ml Tubes (2-4) on ice.

1. Pour (approximately) 10 ml cold PBS into tube (8-11 ml) and put tube on ice.
2. Put stack of dishes on desk. Beginning at the top, decant medium into waste container, shortly dry edge of dish overhead on paper towel, put dish on desk and add PBS from the tube. Move dish slightly to distribute PBS. Replace lid and continue with 1. for the next dish.
3. After all dishes are in PBS: begin with lowest dish and release cells softly with scraper. Use 10 ml pipette to aspirate suspension. Flush dish once softly with suspension to collect residual cells. Aspirate and transfer to 50 ml tube on ice.
4. Continue with 3. until cells from all dishes are collected in the 50 ml tubes on ice. **Take control sample *IPC1* (50µl) to analyze later by trypan blue staining and microscopy (See Figure 1).** In this step, cells should be seen intact. After cell lysis (*IPC2*, step 22), cells should be broken, showing blue membranes and nucleus.
5. Centrifuge tubes: 500 g room temperature 3 min.
6. Decant supernatants into waste container. Softly collect all pellets in 10 ml PBS using a 10 ml pipette. Transfer in 15 ml tube.
7. Centrifuge: 1.850g 4°C 5 min.

HYPOTONIC INCUBATION

approximate time required: 20-30 min

8. During centrifugation, prepare 10-15 ml (this is appropriate for 1-2 ml pCV) of Hypotonic Buffer on ice by supplementing DTT (2.5 µl/ml) and PMSF (1 µl/ml).
9. Measure packed cell volume (pCV) of the pellet and mark it on the tube with a pen.
   DOCUMENT **pCV:** _____ (typical: 1-2 ml)
10. Decant supernatant and shortly dry edge of tube overhead on papertowel.
11. Resuspend cells ***softly*** with 10 ml pipette in a volume of supplemented **Hypotonic Buffer** corresponding to **5 pCVs**.
    DOCUMENT **Volume for resuspending:** _____
12. Incubate on ice for 2 min.
13. Centrifuge: 1.850g 4°C 5 min. Put back on ice.
14. Estimate the volume increase of the packed cells.
    DOCUMENT **Increase:** _____ (typical: 0.1-1 ml)
15. Calculate required volume of Hypotonic Buffer for lysis: **2 x pCV – Increase**.
    DOCUMENT **Volume Hypotonic Buffer** (for use in step 17)**:** _____
    This will result in a final suspension with a volume of three times the pCV.
16. Remove supernatant with 10 ml pipette.
17. Resuspend pellet in the calculated **volume** of **Hypotonic Buffer** softly but thoroughly with 1 ml pipette.
18. Incubate on ice for 5-10 min.

CELL LYSIS

approximate time required: 15-30 min

19. Transfer suspension to Dounce homogenizer on ice.
20. Lyse cells with pestle „Tight“ by slowly moving pestly up and down while keeping the homogenizer on ice. During upward stroke, remove pestle completely out of suspension and re-insert cautiosuly to avoid excessive foaming. One complete stroke (up and down) should take approximately 1-3 seconds.
21. Perform 10-15 strokes.
    DOCUMENT **# strokes:** _____ (typical: 10-15)
22. Transfer lysed suspension to (1 or 2) 2 ml Cups. **Take control sample *IPC2* (50µl) to analyze together with *IPC1* (Figure 1).**
23. Centrifuge 10.000 g 4°C 2 min. Remove supernatant including fatty substances and foam at the top of the cup.

24. Centrifuge 12.000 g 4°C 3 min.
25. Remove residual supernatant accurately and put cup on ice. **Keep the residual supernatant (*IPC3*) to analyze nuclear protein extract enrichment by Western-blot, it is the cytoplasmic extract (Figure 2).**
26. During centrifugation, clean homogenizer, rinse with demineralized water, and dry with paper towel. Put it back on ice.
27. During centrifugation, prepare 1 ml supplemented Resuspension Buffer by adding DTT (5 µl/ml) and PMSF (2 µl/ml) (double concentration).

NUCLEAR EXTRACTION approximate time required: 1.5 hours

28. Estimate the pellet volume which contains the nuclei (pNV).
    DOCUMENT **pNV:** _____ (typically 75%-100% of the pCV)
29. Calculate required volume of Resuspension/High Salt Buffer for nuclear extraction: **0.4 times the pNV**.
    DOCUMENT **Volume Resuspension/High Salt Buffer:** _____
30. Completely remove supernatant.
31. Add calculated volume of **Resuspension Buffer** to pellets.
    Resuspend thoroughly with a 1 ml pipette whose tip has been shortened by cutting so as to provide a larger (2-4 mm) aspiration hole.
    Transfer suspension to cleaned Dounce homogenizer on ice.
32. Add calculated volume of **High Salt Buffer** (**0.4 times the pNV**) to suspension within Dounce homogenizer. Immediately continue.
33. Mix and homogenize on ice by 3-8 soft strokes with Dounce pestle „Tight“.
34. Transfer suspension to 1.5 ml or 2 ml cup(s) on ice.
35. Incubate at 4°C for 30 min under rolling soft agitation.
36. Centrifuge at 20.000 g 4°C for 30 min.
37. The supernatant is the nuclear extract[1]. It can be shock-frosted in liquid nitrogen and stored at -80°C. **Keep the pellet to analyze nuclear protein extract enrichment by Western-blot (*IPC4*), it is the nuclear membrane (Figure 2).**

DIALYSIS approximate time required: 3 hours

38. If the nuclear extract was frozen, thaw it slowly on ice or fingerthaw it. Keep it on ice.
39. Prepare dialysis tubing: cut appropriate length of tubing and water it in 200-400 ml on a magnetic stirrer at room temperature.
40. Put approximately 100x the volume of the nuclear extract of dialysis buffer in a glass beaker on ice.
41. Add PMSF and DTT to the dialysis buffer; mix on magnetic stirrer.
42. Rinse dialysis tubing from the inside and from the outside with dialysis buffer. Use 1 ml pipette for rinsing the inside.
43. Clamp one end of tubing. Clear liquids out of the other end by squeezing along with two fingers.
44. Open dialysis tubing, form an open tube. Hold tube vertical, the clamp at the bottom side.
45. Enter nuclear extract (hold tubing at a slight angle to allow extract to run properly inside to the bottom).
46. Close tubing at the top with your fingers, fold top of tubing once or twice to lock the air and use the air bubble to create a pressure to force the extract into the completely expanded bottom of the tubing above the clamp.
47. While keeping the pressure, apply second clamp just above the surface of the extract
    (a small bubble may be included or a bit of extract may be lost, both is acceptable).
48. Put tubing in beaker with dialysis buffer and dialyse on a magnetic stirrer for **three hours at +4°C**[2]. Nuclear extract turns more turbid during dialysis.
49. Remove one clip and retrieve extract with a 200 µl pipette. Transfer to 1.5 ml cup on ice. Use scissors to obstructive length of tubing.
50. Centrifuge at 20.000 g 4°C 30 min. Transfer supernatant (dialysed nuclear extract) to new, pre-chilled cup.
51. Determine volume and protein concentration of extract. **Protein concentration is acceptable above 3 µg/µl.**
    DOCUMENT **Volume extract:** _____
    DOCUMENT **Protein concentration:** _____ (usually 3-6 µg/µl)
52. Extract aliquots (100 µl) should be snap-frozen in liquid nitrogen and stored at -80°C. **Keep 20 µg of protein (*IPC5*) to analyze nuclear protein extract enrichment by Western-blot together with *IPC3* and *IPC4*.** Enrichment for nuclear extract proteins must be seen by Western blot to ensure nuclear extract quality (Figure 2).

[1] The nuclear extract may show a small cloud of insoluable floating material. This does not affect its quality. You may measure the protein concentration, which should be above 4-5 µg/µl.
[2] Conductivity may be checked in comparison to dialysis buffer to confirm that dialysis is complete. Normally, 2.5 hours are sufficient.

Figure 1. Cell lysis monitoring. Trypan blue staining and microscopy allow to determine the cell lysis state before and after the hypotonic incubation and the corresponding cell lysis using the mechanical process of the pestle homogenizer. Figure show the aspect of the cells before cell lysis (***IPC1***) and after (***IPC2***) this process.

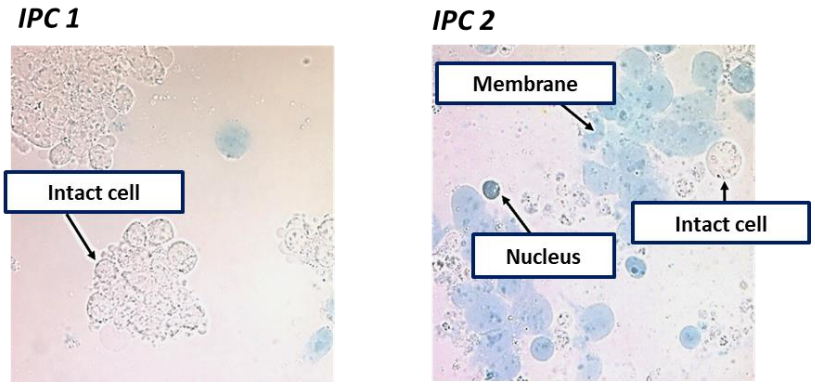
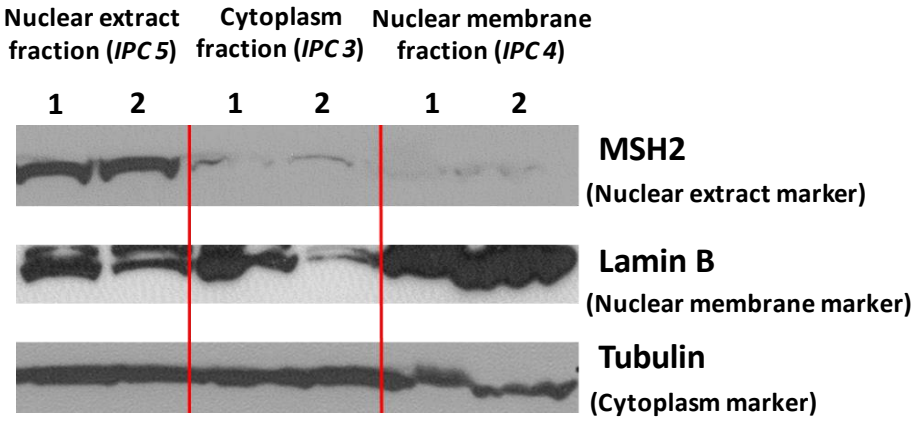


Figure 2. Nuclear protein extract enrichment. Western blotting allows verifying the enrichment in nuclear proteins. The figure is an example of the comparison between the different cellular fractions in two different experiments (1 and 2). MSH2 protein was used as nuclear extract marker because its localization is the nucleus. Lamin B, localized in nuclear membrane, was used as marker of the nuclear membrane fraction and Tubulin, component of the cellular cytoskeleton, was used as marker of cytoplasm fraction.



Protocol 3 Cell Transfection and Whole Cell Protein Extract

A. Cell Transfection using Polythylenimine (PEI)

Material required:
- Empty 1.5ml tubes
- Polyethylenimine (PEI) (1 mg/ml solution in water)
- Dulbecco's Modified Eagle Medium (containing no additives)
- Fetal bovine serum
- Penicillin-streptomycin
- Expression plasmids of the variants of interest

DATE: _____
CELL LINE: _____
NUMBER AND SIZE DISHES: _____
CONFLUENCY (%): _____ (typical: 30-40%)

TRANSFECTION approximate time required for 5 dishes: 45 min

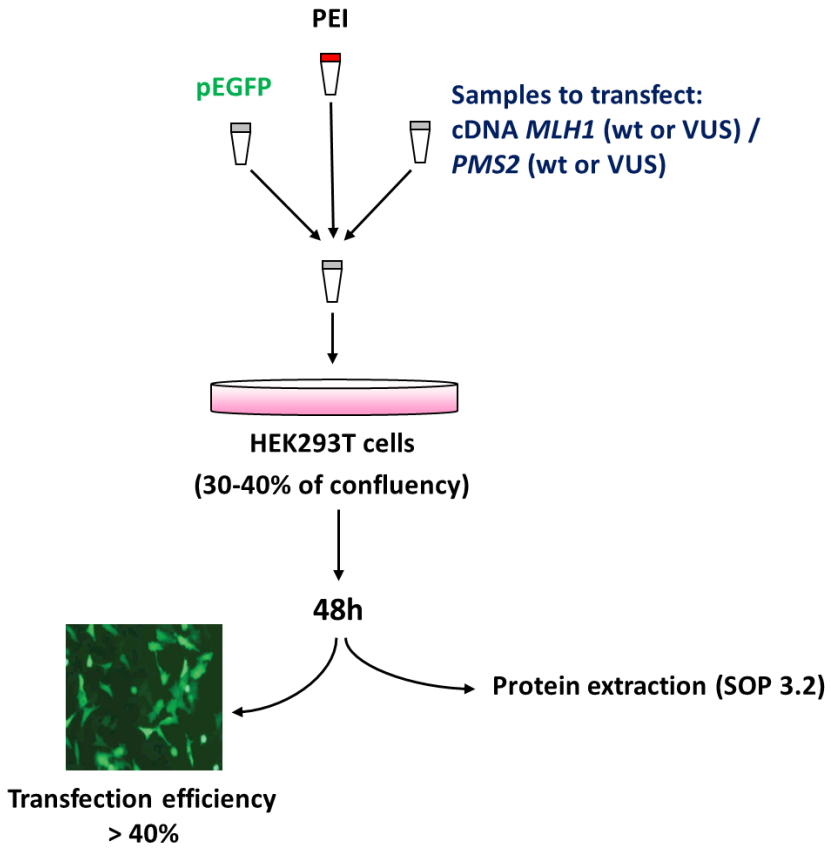
HEK293T cells are grown in Dulbecco's Modified Eagle Medium (Gibco) with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco). Cells should be seeded at least 6 hours prior transfection in 10 cm plates. The cell confluence at the time of transfection should be 30-40%.

1. Mix basal medium (containing no additives ie. serum, antibiotics or other proteins) with the plasmids of interest at the following conditions, process are exemplifying in Figure 1:

Reagent	Sample		
	NO DNA	**GFP**	**MLH1/PMS2**
Medium	980 µl	X µl	X µl
MLH1 plasmid	-	-	X µl (5 µg)
PMS2 plasmid	-	-	X µl (5 µg)
GFP plasmid	-	X (5 µg)	X µl (0,5 µg)
PEI 1mg/ml	20 µl	20 µl	20 µl
Final volum	**1000 µl**	**1000 µl**	**1000 µl**

2. Vortex for 10 s and spin down.
3. Add 20 µl PEI (from 1 mg/ml solution); mix by vortexing and spin down.
4. Incubate for 15 min at room temperature.
5. Apply the mixture (1 ml volume) to cells, drop by drop. Move the dish carefully to distribute the mixture.
6. Incubate 48 hours at 37 ºC, in a cell culture incubator.
7. After 48 hours, harvest the cells to determine the transfection efficiency in the cytometer and continue the protein extraction following *SOP 3.2 Whole Cell Protein Extract*.

Figure 1. HEK293T transfection. HEK293T cells were transfected at 30–40% confluence with MLH1 and PMS2 expression plasmids (3 µg/ml, respectively) and 0.5 µg/ml of pGFP, as a transfection control, using 20 µl/ml of the cationic polymer polyethylenimine (stock solution 1 mg/ml, Polysciences). After 48h, cells were harvested and prepared for protein extraction and cytometer analysis. Transfection efficiency should be above 40%, usually it is in a range of 50-75%.



B. Whole Cell Protein Extract

Material required: Cell centrifuge (15 ml tubes)
Microcentrifuge (2 ml tubs) at 4°C
Cold Extraction Buffers A and C (***Protocols 1.11*** and ***1.12***)
PMSF and DTT stock solutions (***Protocols 1.9*** and ***1.10***)

DATE: _____
CELL LINE: _____
NUMBER AND SIZE DISHES: _____
TRANSFECTED WITH: _____
TRANSFECTION EFFICIENCY
(% estimated GFP): _____ (minimum acceptable 40%)
CONFLUENCY (%): _____ (typical: 80-100%)

CELL HARVEST approximate time required for 5 dishes: 30 min

Have ready: 10 cm dishes with transfected 293T cells
Open canister for waste, papertowel for dripping
PMSF and DTT solutions
Empty 1.5 ml cups on ice
Prepare 1-2 ml of Buffer A for usage by adding the required PMSF and DTT solutions

1. Put stack of dishes on desk. Beginning at the top, pour medium into waste container, dry edge of dish overhead on paper towel, put dish on desk and add approximately 10 ml PBS. Move dish slightly to distribute PBS. Replace lid and continue with 1. for the next dish.
2. After all dishes are in PBS: begin with lowest dish and release cells softly with scraper. Use 10 ml pipette to aspirate suspension. Flush dish once softly with suspension to collect residual cells. Aspirate and transfer to 15 ml tube.
3. Continue with 2. until cells from all dishes are collected in the 15 ml tubes. Use 500 µl of cell suspension to measure the transfection efficiency by flow cytometry. Transfection efficiency for HEK293T cells is accepted above 30%. Continue with step 4 with the rest of the cell suspension.
4. Centrifuge Tubes: 1000 g room temperature 2 min.
5. Decant supernatants into waste container. Dry edge on papertowel.
6. Centrifuge again: 1000 g room temperature 1 min. Put tubes on ice.
7. Estimate cell pellet volume (packed cell volume, pCV) by comparing it to known volumes of water in reference tube.
   DOCUMENT **pCV:** _____ (typical: 50-100 µl)
8. Remove residual PBS from cell pellet with 200 µl pipette (pellet should be practically "dry").
9. Resuspend cell pellet in **2 x pCV** **Buffer A** (supplemented with PMSF/DTT) with 200 µl pipette and transfer suspension to 1.5 ml cup on ice.
10. Repeat 8. - 9. for all samples.
11. Freeze samples at -80°C or in liquid nitrogen (**cell lysis**).
    Samples may be stored at -80°C for later extraction or may be processed directly.

CELL EXTRACTION approximate time required: 90 min

Have ready: Buffer C, PMSF and DTT solutions. Empty 1.5 ml cups on ice.

12. Thaw cell suspensions on ice or carefully rotate in fingers for thawing.
13. Meanwhile: prepare sufficient volume of **Buffer C** on ice by supplementing DTT and PMSF solutions.
14. Aspirate **2 x pCV** **Buffer C** (supplemented with PMSF/DTT) with 200 µl pipette and mix into cell suspensions by pipetting up and down thoroughly until suspension is homogenous.
15. Incubate under soft agitation (rolling) for 30 min at +4°C.
16. Centrifuge: 20.000 g 4°C 30 min.
17. Transfer supernatant to new 1.5 ml cups on ice.

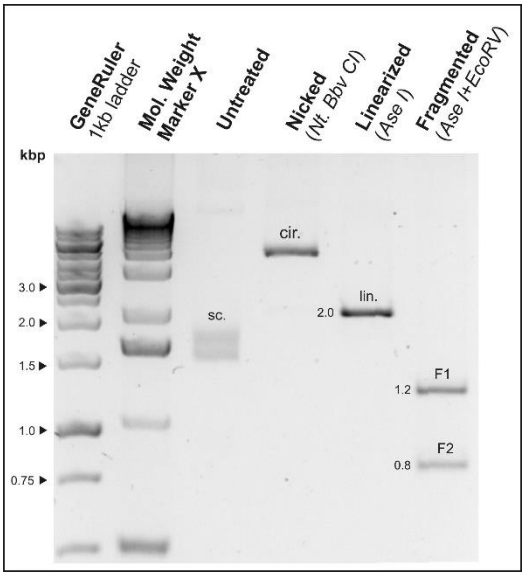
This is the whole cell protein extract. **Protein concentration is acceptable above 3 µg/µl (typically between 3 and 8 µg/µl)**. Extracts may be snap-frozen in liquid nitrogen and stored at -80°C. MutLα proteins in these preparations are very durable and resistant against many cycles of thawing and freezing. They can be used for Western Blotting and for the Mismatch Repair Assay.

Protocol 4 Mismatch Plasmid Substrate Generation

Material required:
- Maxiprep of plasmid pUC19 CPDC Bbv at a concentration of 0.6 – 1.0 µg/µl
- WHCPDPuriAS (5'-gagcgactcgctaccgtcacatt**G**atatccgc-3') at 100 mM and HPLC-purified*
- WHpCPD7 (5'-***Phosphate***-gcggatat**T**aatgtgacggtagcgagtcgctc-3') at 100 mM and HPLC-purified*
- Amicon 50K 500 µl columns.

*HPLC-purified oligomers are critical to improve the quality of the mismatched plasmid.

For identity control, the plasmid may be test-digested and should yield the following patterns (samples run in 2% agarose gel in TAE buffer):



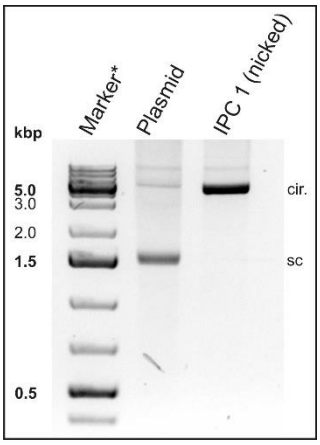
sc. = supercoiled plasmid (original), runs to approxim. 1.5-2 kbp
cir. = circular plasmid (relaxed by nicking), runs to approxim. 5 kbp
lin. = linearized plasmid (cut once), runs to 2 kbp
F1 = Fragmented plasmid fragment 1 (1.2 kbp)
F2 = Fragmented plasmid fragment 2 (0.8 kbp)

A. Nicking (double)

approximate time required: 3.5 hours

1.	Mix:	Water	**to 100 µl:**	**µl**	
		Plasmid solution pUC CPDC Bbv (Maxipreperation),	**50 µg:**	**µl**	
		Date: **Conc.:** .			
		NEB Buffer 3 10x		10	µl
		NEB N.BstNB I Nicking Endonuclease #R0607	40 U	4	µl
		Final volume:		*100*	*µl*

2. Incubate: 55°C for 3 hours
3. Take 0.5 µl sample (*In-Process Control* 1, ***IPC1***)
4. Run ***plasmid*** (250 ng) and ***IPC1*** on a 2% agarose gel.



Nicking causes circularisation of the plasmid ("cir.").
Supercoiled ("sc.") plasmid band should completely disappear.

B. Oligomer removal approximate time required 60 min

5.	Add:	Water	**to 150 µl:**	31	µl
		Oligomer WHCPDPuriAS 100 µM (to 50-fold molar excess):		19	µl

6. Incubate: 85°C for 5 min
7. Let cool to room temperature slowly (over 20 min).
8. Add: Water **to 500 µl**: 350 µl
9. Apply to Amicon 50K 500 µl spin column. Centrifuge: 14.000 g 10 min room temperature. Store flowthrough (***IPC 2***).
   Add 500 µl water. Centrifuge: 14.000 g 10 min room temperature. Store flowthrough (***IPC 3***).
   Add 500 µl water. Centrifuge: 14.000 g 10 min room temperature. Store flowthrough (***IPC 4***).
   Spin column contents into new cup. Centrifuge: 1.000 g 2 min.
10. Fill up sample with water to 90 µl.
11. Take control sample (0.5 µl) (***IPC5***)

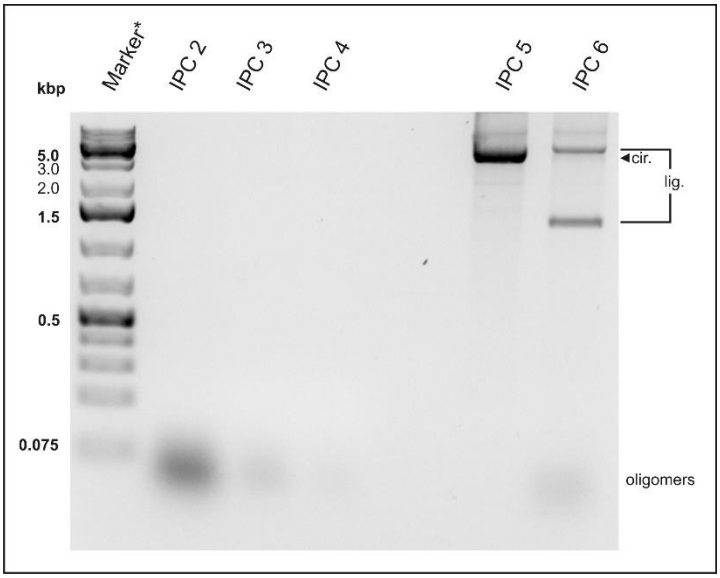
C. Mismatch oligomer annealing and ligation approximate time required: overnight

12.	Add:	Oligomer WHpCPD7 100 µM (to 20fold molar excess):	9	µl
		T4 DNA ligase buffer 10x (NEB)	11	µl
		Final volume:	*110*	*µl*

13. Incubate at 80°C for 5 min
14. Let cool to room temperature slowly (over 20 min).

15.	Add:	DTT 200 mM (Protocol 1.14)		0.5	µl
		ATP 100 mM (Fermentas #R0441)		1	µl
		T4 DNA ligase 400 U/µl (NEB #M0202)	(50 U):	1	µl

16. Incubate at 16°C overnight.
17. Take control sample (0.5 µl) (***IPC6***)
18. Run ***IPC2-IPC6*** on a 2% agarose gel. Apply as much volume of ***IPC2-IPC4*** as fits into the pockets. Oligomers are best visible when detected after just a short running time (10-15 min.). Continue electrophoresis for separation of the plasmid products afterwards.



*Oligomers are washed out successively in **IPC2-IPC4.***
*Band of ligated plasmid ("lig.") appears in **IPC6** versus **IPC5**.*

D. Ethanol precipitation #1 approximate time required: 30 min

19.	Add:	Sodium acetate 3M	(1/10 of total sample volume)	11	µl
		Ethanol 100%	(2.5x of total sample volume)	303	µl

20. Incubate -20°C 10 min
21. Centrifuge 12.000 g 4°C 10 min.
22. Remove supernatant
23. Add Ethanol 70% 50 µl
21. Centrifuge 12.000 g 4°C 2 min.
22. Remove supernatant.
23. Dissolve in water 50 µl
24. Take control sample (0.5 µl; ***IPC7***)

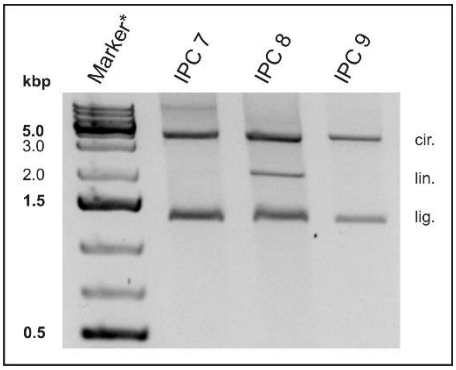
E. Elimination of original plasmid approximate time required: 2.5 hours

25.	Add:	Buffer H 10x	6	µl
		EcoRV (TaKaRa Bio #1042A)	4	µl
		Final volume:	*60*	*µl*

26. Incubate at 37°C 1.5 hour
27. Take control sample (0.5 µl; ***IPC8***)

28.	Add:	Water	16	µl
		NEB Buffer 4 10x	10	µl
		ATP 100 mM (Fermentas #R0441)	2	µl
		Exonuclease V (NEB #M0345)	3	µl
		Final volume:	*100*	*µl*

29. Incubate at 37°C 2 hour. After 1.5 hours, take control sample (0.5 µl; ***IPC9***).
30. Run ***IPC7-IPC9*** on a 2% agarose gel. If the new band (linearized original plasmid without mismatch) appearing in IPC8 versus IPC7 has completely disappeared in IPC9, continue to step 31. Otherwise add further Exonuclease V enzyme.



*EcoRV-sensitive original plasmid (without mismatch) appears linearized ("lin.") in **IPC8** and disappears after Exonuclease V treatment (**IPC9**).*

31. Incubate at 65°C 10 min.
32. Take control sample (1.5 µl; ***IPC10***). **Digest 1 µl of *IPC10* with EcoRV and AseI. If elimination of original plasmid is completed, only the linearized band will appear, corresponding to the digestion of AseI. If fragmented plasmid bands appear, repeat the step "*E. Elimination of original plasmid*".**

F. Ethanol precipitation #2 approximate time required: 30 min

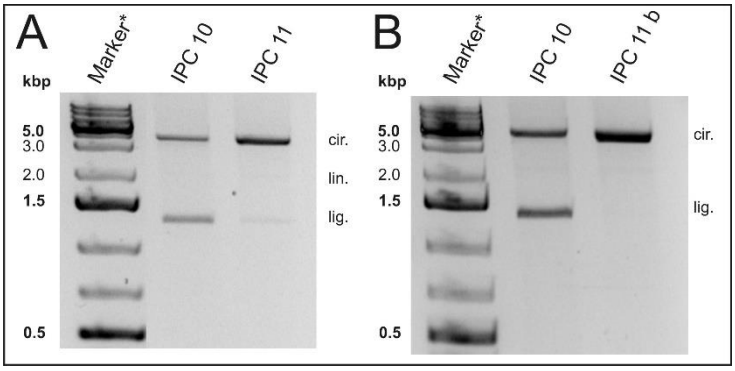
33.	Add:	Sodium acetate 3M	(1/10 of total sample volume)	10	µl
		Ethanol 100%	(2.5x of total sample volume)	275	µl

34. Incubate -20°C 10 min
35. Centrifuge 12.000 g 4°C 10 min.
36. Remove supernatant
37. Add Ethanol 70% 50 µl
38. Centrifuge 12.000 g 4°C 2 min.
39. Remove supernatant.
40. Dissolve in water 86 µl

G. Introduction of 3'-single strand break approximate time required: 3-4 hours

41.	Add:	NEB Buffer 4	10	µl
		Nt.Bbv CI NEB	4	µl
		Final volume:	*100*	*µl*

42. Incubate 37°C for 3 hours.
43. After 1 hour, take control sample (0.5 µl; ***IPC11***).
44. Run ***IPC10-IPC11*** on a 2% agarose gel:



*If nicking is not complete (like in **Panel A**), add further 1-2 µl of Nt.Bbv CI after **step 44** and continue until ligated band (at approximately 1.5 kbp) is not detectable any more (**Panel B, IPC 11 b**).*

45. Incubate at 80°C for 20 min.
46. Centrifuge 12.000 g for 5 min.
47. Remove supernatant and mix with 400 µl water. Apply to Amicon 50K 500 µl spin column.
    Centrifuge: 14.000 g 10 min room temperature.
    Add 500 µl water. Centrifuge: 14.000 g 10 min room temperature.
    Add 500 µl water. Centrifuge: 14.000 g 10 min room temperature.
    Spin column contents into new cup. Centrifuge: 1.000 g 2 min.
48. Fill up sample with water to 100 µl.
49. Determine DNA concentration (typical: 50-200 ng/µl).
    Mw: 1320 kDa=1320 ng/pmol, 1 ng corresponds to 0.76 fmol. Typically, 25-45 ng = 19-34 fmol substrate are used per assay.

Protocol 5: Mismatch Repair Assay (MMR Assay)

Required material:

Nuclear extract (***Protocol 2***)
Whole cell extract (***Protocol 3***), containing transfected MutLα protein:
- Wildtype (for comparison)
- Variant(s) of interest
- Untransfected control

Plasmid substrate (***Protocol 4***)
Mismatch repair reaction buffer 10x (***Protocol 1.18***)
Restriction endonucleases AseI and EcoRV

General considerations:

Always include a wildtype ("positive") control and a "negative" control without MutLα.
Only compare controls and variants that have been prepared in parallel (in one transfection and one whole cell extraction process).
Always use identical preparations of all reagents for parallel experiments.
Identical reagents with identical volumes in all parallel experiments should be included in the master mix.
You may use the "automatic" Excel pipetting scheme by inserting information on your reaction in the places written in red.

Calculations of the pipetting scheme (implicated in the Excel sheet):

Water (I.9):	I.6-Sum(I.10-I.13)
Repair Buffer (I.10):	I.6/10
KCl (I.11):	I.6/500*(I.8-I.12*100/I.6-I.13*420/I.6)
Nuclear extract (I.12):	500/C4
Whole cell extract (I.13):	5/I.7
Total volume control:	Sum I.9 to I.13
Total KCl concentration control:	(I.8*100+I.6*420+I.11*500)/I.6

Mismatch Repair Reaction

1. Mix chilled ingredients on ice as detailed in table:

   Mix components of the **Master mix** on ice and add other components which do not go to the Master mix into the individual sample cups. Aliquot Master mix (volume in D6) into indivdual sample cups. Mix by pipetting, spin down.

	B	C	D	E	F	G	H	I	J	K
2	**28/04/2017**		**Master mix**	enter "x" here if all samples have identical volume	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**
3	**Experiment name**	**comment**	Enter: # of samples (add 5-10% for having some surplus):		**comment**	**comment**	**comment**	**comment**	**comment**	**comment**
4	**Enter:** protein concentration of nuclear extract (mg/ml):	**5,5**	**6,6**							
5	Enter: Volume of plasmid substrate to be used (µl)	**0,8**	Aliquot Master mix to:							
6	Desired total reaction volume (standard: 15 µl)		**12,6**		15	15	15	15	15	15
7	**Enter:** Protein concentration of whole cell extract (mg/ml)				**5**	**5**	**6**	**7,18**	**7,18**	**7,18**
8	Desired KCl concentration (standard: 110 mmol)				110	110	110	110	110	110
9	**Water**		**13,0**	x	**2,0**	**2,0**	**2,0**	**2,0**	**2,0**	**2,0**
10	**Repair buffer 10X (Protocol 1.18)**	**enter batch identifier**	**9,9**	x	**1,5**	**1,5**	**1,5**	**1,5**	**1,5**	**1,5**
11	**KCl 500 mM (Protocol 1.15)**		**individual**		**0,64**	**0,64**	**0,78**	**0,90**	**0,90**	**0,90**
12	**Nuclear extract (50 µg) (Protocol 2)**	**enter batch identifier**	**60,0**	x	**9,1**	**9,1**	**9,1**	**9,1**	**9,1**	**9,1**
13	**Whole cell extract (5 µg) (Protocol 3)**	**enter batch identifier**	**individual**		**1,00**	**1,00**	**0,83**	**0,70**	**0,70**	**0,70**
14	**Plasmid substrate (Protocol 4)**	**enter batch identifier**	**individual**		**0,8**	**0,8**	**0,8**	**0,8**	**0,8**	**0,8**
15	Total volume (control)				15,0	15,0	15,0	15,0	15,0	15,0
16	Total KCl concentration (control)				110,0	110,0	110,0	110,0	110,0	110,0

2. Incubate at 37°C for 15 minutes.
3. In the meantime, prepare appropriate volume of **Stop buffer** at room temperature (do not chill to avoid SDS precipitation):

Stop buffer		Per sample	**Master mix**
Number of samples (+1)	**7**		
Water		16,8	**118**
SDS solution 10% (Protocol 1.17)		2	**14**
EDTA 100 mM solution (Protocol 1.6)		6	**42**
Proteinase K solution NEB #P8102		0,2	**1,4**
Total volume		25	175

4. Add 25 µl **Stop buffer** to each reaction, mix by pipetting.
5. Continue incubation at 37°C for 10 min.

Plasmid Substrate Purification

6. Add **water** to 150 µl total sample volume.
7. Add **phenol** (125 µl to each sample).
8. Vortex for 10 s and centrifuge at 15.000 g, 1 min, at room temperature.
9. Transfer supernatants to new cups.
10. Add **chloroform** (150 µl to each sample).
11. Vortex for 10 s and centrifuge at 15.000 g, 1 min, room temperature.
12. Transfer supernatants to new cups.
13. Ethanol precipitation

Add:	**tRNA solution (1 µg/µl) (Protocol 1.14)**	1 µl	
	Sodium acetate solution 3M pH 5.2	15 µl	(e.g. Merck Chemicals Cat. #567422)
	Ethanol 100%	415 µl	

14. Mix vigorously, store at -20°C for 20 min. *Do not precipitate over night.*
15. Centrifuge: >12.500 g, 4°C, 20 min.
16. Remove supernatants.
17. Add **Ethanol 70%** 50 µl. Shake.
18. Centrifuge: >12.500 g, 4°C, 5 min.
16. Remove supernatants completely.

Plasmid Substrate Analysis

17. Dissolve pellets in 15 µl **Digestion buffer**:

Digestion buffer		
Number of samples (+1):		**7**
	Per sample:	
Water	11,75	**82,3**
Buffer H 10x (TaKaRa Bio)	1,5	**10,5**
RNAse A 1 mg/ml (Protocol 1.16)	1	**7,0**
BSA 10 µg/µl	0,15	**1,1**
Ase I	0,3	**2,1**
EcoRV (TaKaRa Bio #1042A)	0,3	**2,1**
Total volume	15	

18. Vortex, shortly centrifuge and incubate at 37°C for 1 h.
19. Run samples on a 2% agarose gel.
20. Quantification: Determine AUC of all three plasmid peaks. % Absolute Repair is calculated as: AUC (bands of repair products) / AUC (all products).

ARTÍCULO 3

High-sensitivity microsatellite instability assessment for the detection of mismatch repair defects in normal tissue of biallelic germline mismatch repair mutation carriers

Maribel González-Acosta*, Fàtima Marín*, Benjamin Puliafito*, Nuria Bonifaci, Anna Fernández, Matilde Navarro, Héctor Salvador, Francesc Balaguer, Sílvia Iglesias, Àngela Velasco, Èlia Grau, Víctor Moreno, Luis Ignacio Gonzalez-Granado, Pilar Guerra-García, Rosa Ayala9 Benoît Florkin, Christian P. Kratz, Tim Ripperger, Thorsten Rosenbaum, Danuta Januszkiewicz-Lewandowska, Amedeo A. Azizi, Iman Ragab, Michaela Nathrath, Hans-Jürgen Pander, Stephan Lobitz, Manon Suerink, Karin Dahan, Thomas Imschweiler, Ugur Demirsoy, Joan Brunet, Conxi Lázaro, Daniel Rueda, Katharina Wimmer, Gabriel Capellá‡, Marta Pineda‡.

* Ambos autores han contribuido en igual medida a este trabajo y comparten primera autoría.

‡ Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Journal of Medical Genetics, 2019. DOI: 10.1136/jmedgenet-2019-106272

RESUMEN:

El síndrome de Lynch (SL) y el síndrome de Deficiencia Constitucional de Reparación de Apareamientos Erróneos (CMMRD) son síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema de reparación de errores simples de apareamiento (MMR). Los tumores asociados a ellos muestran inestabilidad de microsatélites (MSI), que también se ha detectado a bajos niveles en tejidos no neoplásicos. El objetivo de este trabajo fue evaluar el rendimiento de la evaluación con alta sensibilidad de la MSI (hs-MSI) en tejidos no neoplásicos para la identificación de SL y CMMRD.

Se analizó el DNA de sangre de 131 individuos agrupados en tres cohortes: la utilizada para determinar el nivel basal de MSI (22 controles), la cohorte de entrenamiento (11 CMMRD, 48 SL y 15 controles) y el grupo de validación (18 CMMRD y 18 controles). Para detectar las inserciones y deleciones en marcadores microsatélite se diseñó un panel de microsatélites frecuentemente inestables en tumores y un algoritmo bioinformático propio. El nivel de hs-MSI se calculó representando el porcentaje de marcadores inestables.

La hs-MSI fue significativamente mayor en las muestras de sangre CMMRD cuando se comparó con los controles en la cohorte de entrenamiento ($p<0,001$). Este hallazgo se confirmó con el grupo de validación, alcanzando una especificidad y sensibilidad del 100% para la detección de CMMRD. Además, se detectó un mayor porcentaje de hs-MSI en portadores bialélicos de *MSH2* (n=5) que en los portadores de *MSH6* (n=15). Por otro lado,

el análisis hs-MSI no detectó diferencias entre las muestras de sangre SL y los controles (p=0,564).

Nuestra aproximación hs-MSI podría ser una herramienta valiosa para el diagnóstico de CMMRD, especialmente en los pacientes con sospecha de CMMRD sin mutación identificada o en portadores de VUS en los genes MMR.

SHORT REPORT

High-sensitivity microsatellite instability assessment for the detection of mismatch repair defects in normal tissue of biallelic germline mismatch repair mutation carriers

Maribel González-Acosta,[1] Fátima Marín,[1] Benjamin Puliafito,[1] Nuria Bonifaci,[1] Anna Fernández,[1] Matilde Navarro,[1] Hector Salvador,[2] Francesc Balaguer,[3] Silvia Iglesias,[1] Angela Velasco,[4] Elia Grau Garces,[1] Victor Moreno,[5,6] Luis Ignacio Gonzalez-Granado,[7] Pilar Guerra-García,[8] Rosa Ayala,[9] Benoît Florkin,[10] Christian Kratz,[11] Tim Ripperger,[12] Thorsten Rosenbaum,[13] Danuta Januszkiewicz-Lewandowska,[14] Amedeo A Azizi,[15] Iman Ragab,[16] Michaela Nathrath,[17,18] Hans-Jürgen Pander,[19] Stephan Lobitz,[20] Manon Suerink,[21] Karin Dahan,[22] Thomas Imschweiler,[23] Ugur Demirsoy,[24] Joan Brunet,[1,4] Conxi Lázaro,[1] Daniel Rueda,[25] Katharina Wimmer,[26] Gabriel Capellá,[1] Marta Pineda[1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136jmedgenet-2019-106272).

For numbered affiliations see end of article.

Correspondence to
Dr Gabriel Capellá and Dr Marta Pineda, Hereditary Cancer Program, Catalan Institute of Oncology, L'Hospitalet de Llobregat 08908, Catalunya, Spain; gcapella@iconcologia.net, mpineda@iconcologia.net

GC and MP contributed equally. MG-A, FM and BP contributed equally.

MG-A, FM and BP are joint first authors.
GC and MP are joint senior authors.

Received 7 May 2019
Revised 16 July 2019
Accepted 16 July 2019

To cite: González-Acosta M, Marín F, Puliafito B, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2019-106272

ABSTRACT

Introduction Lynch syndrome (LS) and constitutional mismatch repair deficiency (CMMRD) are hereditary cancer syndromes associated with mismatch repair (MMR) deficiency. Tumours show microsatellite instability (MSI), also reported at low levels in non-neoplastic tissues. Our aim was to evaluate the performance of high-sensitivity MSI (hs-MSI) assessment for the identification of LS and CMMRD in non-neoplastic tissues.

Materials and methods Blood DNA samples from 131 individuals were grouped into three cohorts: baseline (22 controls), training (11 CMMRD, 48 LS and 15 controls) and validation (18 CMMRD and 18 controls). Custom next generation sequencing panel and bioinformatics pipeline were used to detect insertions and deletions in microsatellite markers. An hs-MSI score was calculated representing the percentage of unstable markers.

Results The hs-MSI score was significantly higher in CMMRD blood samples when compared with controls in the training cohort (p<0.001). This finding was confirmed in the validation set, reaching 100% specificity and sensitivity. Higher hs-MSI scores were detected in biallelic *MSH2* carriers (n=5) compared with *MSH6* carriers (n=15). The hs-MSI analysis did not detect a difference between LS and control blood samples (p=0.564).

Conclusions The hs-MSI approach is a valuable tool for CMMRD diagnosis, especially in suspected patients harbouring MMR variants of unknown significance or non-detected biallelic germline mutations.

INTRODUCTION

Lynch syndrome (LS; OMIM #120435), the most prevalent hereditary colorectal and endometrial cancer syndrome, is an autosomal dominant cancer-susceptibility disease caused by inactivating heterozygous germline mutations in mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*).[1] Constitutional mismatch repair deficiency (CMMRD; OMIM #276300) is a rare devastating cancer syndrome caused by biallelic germline mutations in the same genes and mainly characterised by the development of haematological, brain and colorectal tumours during childhood and adolescence.[2 3] Overlapping phenotypes have been described between LS and CMMRD,[4 5] as well as between CMMRD and other cancer syndromes such as neurofibromatosis type 1 (NF1), polymerase proofreading-associated polyposis (PPAP) and Li-Fraumeni.[6 7]

The identification of these inherited conditions has important consequences for the clinical management of carriers.[8 9] Molecular diagnosis of LS and CMMRD is often hampered by the identification of variants of unknown significance (VUS) in about 30% of all identified MMR variants and by difficulties in sequencing *PMS2* due to multiple pseudogenes, which accounts for approximately 60% of CMMRD cases.[3 6]

In LS, somatic inactivation of the MMR wildtype allele initiates an accumulation of errors mainly in repetitive sequences. Consequently, LS-associated tumours are hypermutated (>10 mutations/Mb), exhibit microsatellite instability (MSI) and lose expression of MMR proteins.[1] In CMMRD, the germline inactivation of both MMR alleles together with somatic polymerase exonuclease domain mutations leads to ultra-hypermutated tumours (>100 mutations/Mb).[10] The CMMRD diagnostic hallmark is the loss of MMR protein expression in both tumour and normal tissue.[3 7] However, some missense mutations are associated with conserved expression and MSI may be negative in CMMRD tumours, especially in non-gastrointestinal cancers.[2 3]

Besides the recently reported in vitro repair assay in lymphocytes,[6] tools have been developed to

assess the degree of MSI in CMMRD normal tissues. The germline MSI (gMSI) assay,[11] based on electropherogram analysis of three dinucleotide markers, has demonstrated high specificity but low sensitivity due to its inability to identify biallelic *MSH6* mutation carriers. The ex vivo MSI analysis,[12] based on lymphoblastoid cell lines, in combination with a methylation-tolerance assay, showed higher sensitivity for CMMRD identification. Recently, a next generation sequencing (NGS) approach to detect gMSI has shown high accuracy.[13] None of these techniques are sensitive enough to detect MSI in normal tissues from LS carriers. Nevertheless, low-level MSI has been reported in blood DNA from individuals with LS using laborious single-molecule analyses.[14 15] Notably, MMR deficiency has been detected in apparently normal colonic and endometrial epithelium of LS carriers.[16]

We hypothesised that an assessment of MSI markers at high sensitivity could improve the diagnosis of cancer syndromes associated with MMR deficiency. Our aim was to evaluate the performance of high-sensitivity MSI (hs-MSI) assessment in normal tissues of LS and CMMRD carriers.

MATERIALS AND METHODS

Patients and samples

Samples from 131 individuals were grouped into three cohorts: baseline, training and validation. The baseline cohort comprised 22 healthy control samples; the training cohort included 74 blood samples from healthy controls, patients with CMMRD and individuals with LS (online supplementary table S1); and the validation cohort comprised 36 blinded samples from individuals with clinical diagnosis of CMMRD[3] and healthy controls, kindly provided by the European Consortium C4CMMRD (online supplementary table S2). Some samples were also analysed in a recent study[13] (online supplementary tables S1 and S2). An oral mucosa sample from a patient with CMMRD (online supplementary table S1), four cases with CMMRD-suspected diagnosis and mutation carriers of CMMRD overlapping syndromes were also included (online supplementary table S3). Five DNA samples from frozen tumours were used as controls, two classified as MSI and three as microsatellite stable (MSS), using the MSI Analysis System (Promega). Genomic DNA was obtained using standard protocols.

Assessment of MSI at high sensitivity (hs-MSI)

The analytical sensitivity of variant detection by using a molecular barcoding-based NGS approach was initially assessed with the ClearSeq Cancer HS panel (Agilent Technologies; online supplementary methods).

A custom panel targeting 277 microsatellites, 91% of them mononucleotide repeats, was designed using HaloPlex HS technology (online supplementary figure S1, online supplementary methods). Sequencing of enriched regions was performed in a HiSeq platform at high coverage (20 000×), reaching a mean depth of 1312±447 reads/marker/sample after deduplication. A set of 231 truly monomorphic microsatellites in the baseline were selected. Among them, 186 markers were previously reported as frequently mutated in tumours with high instability (MSI-H). A bioinformatics pipeline for microsatellite indel calling was customised (online supplementary figure S2, online supplementary methods).

To assess the hs-MSI status at each microsatellite locus, the instability level, corresponding to the sum of the frequencies of all allele lengths different from the wildtype (mutational load method), was calculated as (1 − wildtype allele frequency). Alternatively, the frequencies of each alternative microsatellite allele length were used (individual allele method). Whenever the instability level or frequency of alternative allele exceeded the mean value in baseline plus 3 SD and the highest value among the individual samples of the baseline, the microsatellite was considered unstable.

For both methods, an MSI score was calculated per sample, representing the percentage of unstable markers. hs-MSI median score was compared between different training set groups using a Wilcoxon rank-sum test (online supplementary figure S2, online supplementary methods).

Analysis of dinucleotide repeats

gMSI analysis of the dinucleotide markers D17S791, D2S123 and D17S250 was performed as described.[11] Analysis of D2S123 from NGS data was described in online supplementary methods.

RESULTS

The percentage of unstable monomorphic markers frequently mutated in MSI-H tumours included in the hs-MSI panel (n=186; mutational load method) was higher in the DNA from MSI-H than MSS colorectal tumours (online supplementary figure S3A). This MSI score was significantly higher in blood DNA samples from patients with CMMRD (median=23.58%) compared with healthy controls (median=1.10%) (p=1.24e-05) or LS blood samples (median=0.85%) (p=9.49e-08), without overlapping (figure 1A and online supplementary table S1). No evidence of clonal expansion was seen in haematological CMMRD samples. In contrast, no difference was detected between LS and control samples (p=0.564) (figure 1A and online supplementary table S1). Similar results were obtained using the whole set of monomorphic markers (n=231) and when an individual allele method was used irrespective of the absolute values of the thresholds for MSI detection in blood (online supplementary figure S3).

Using an independent blinded set of blood samples, the MSI score accurately distinguished patients with CMMRD (median=26.28%) from controls (median=0.57%) (p=2.784e-07) (figure 1B, online supplementary table S2). In this context, the hs-MSI approach displayed a specificity of 100% (95% CI 89.42% to 100%) and a sensitivity of 100% (95% CI 88% to 100%) (online supplementary table S4). In agreement with the results obtained in the gMSI assay, instability at D2S123 dinucleotide marker was detected in biallelic MMR carriers except for *MSH6* (online supplementary tables S1, S2 and S5).

No correlation between MSI score and age at blood sampling was observed in control, LS or CMMRD samples (figure 2A, online supplementary figures S4 and S5). Moreover, no correlation with age of cancer onset was noted in CMMRD (figure 2B) or LS-affected patients (online supplementary figure S5). In contrast, when CMMRD samples were grouped by germline-affected gene, significant differences were observed between instability levels of *MSH6* and *MSH2* biallelic carriers (p=0.0014) (figure 2C). Furthermore, no dependency of MSI levels and germline affected gene was observed in LS samples (p=0.0523) (online supplementary figure S5A).

An oral mucosa DNA sample (CMMRD-01) displayed similar MSI score to a paired blood sample (figure 2C and online supplementary table S1). Conversely, high hs-MSI score was not detected in the blood from germline *TP53*, *POLE*/*POLD1* and *NF1* mutation carriers, early-onset LS or four cases with a suspected but unconfirmed diagnosis of CMMRD, pointing

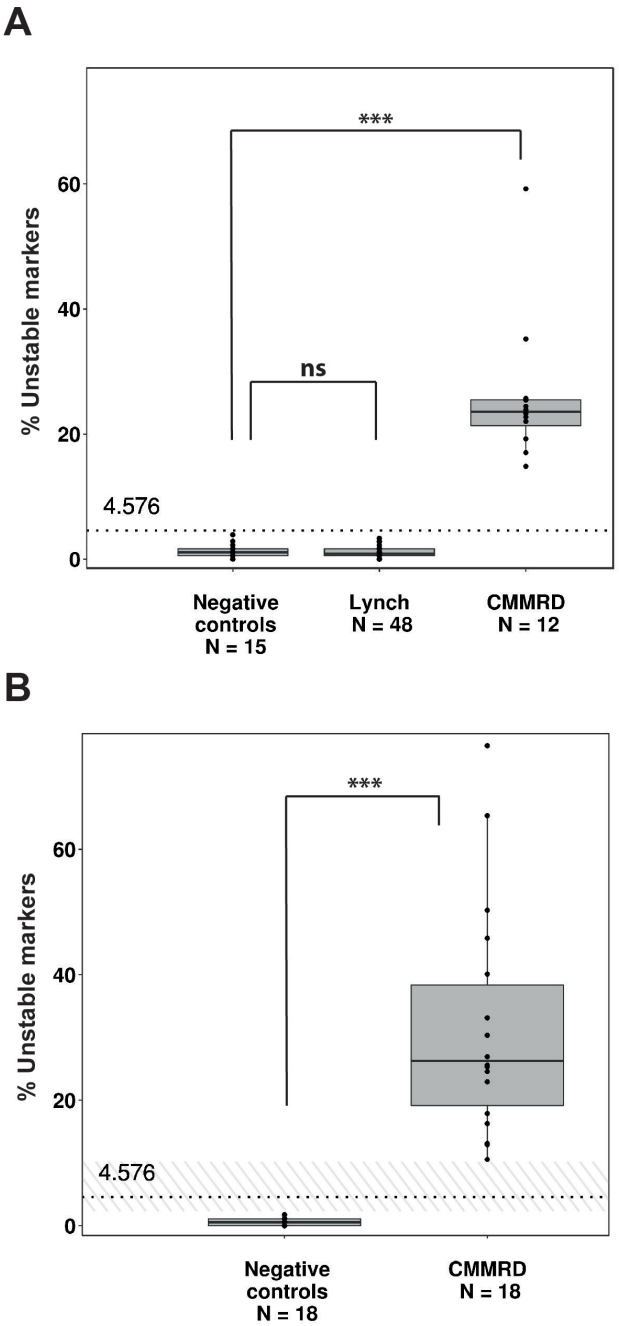


Figure 1 hs-MSI analysis in the training and validation cohorts. Monomorphic microsatellite markers frequently mutated in MSI-H tumours (n=186) analysed using the mutational load analysis method. (A) MSI score in blood DNA samples from LS (median=0.85, IQR=0.55–1.65, range=0.00–3.33), CMMRD (median=23.58, IQR=21.33–25.49, range=14.84–59.22) and healthy individuals (median=1.1, IQR=0.54–1.65, range=0.00–3.89) from the training set. Significant differences were observed between patients with CMMRD and negative controls (***p=1.24e-05), while no differences were found between patients with LS and negative controls (ns, non-significant, p=0.564). Dashed line indicates the threshold for hs-MSI detection in blood samples. (B) MSI score in blinded samples from the validation cohort. Patients with CMMRD (median=26.28, IQR=19.14–38.37, range=10.56–76.50) and negative controls (median=0.57, IQR=0–1.11, range=0–1.79) were discriminated with no overlapping (hatched area) (***p=2.784e-07). Dashed line indicates the threshold for hs-MSI detection. CMMRD, constitutional mismatch repair deficiency; hs-MSI, high-sensitivity microsatellite instability; LS, Lynch syndrome.

to the absence of CMMRD in the latter (online supplementary table S3, online supplementary figure S6).

DISCUSSION

Accurate and prompt diagnosis of CMMRD is essential for therapeutic decisions and surveillance recommendations.[9] Here we report the performance of the novel hs-MSI approach for high-sensitivity gMSI assessment. Our hs-MSI approach based on the analysis of mononucleotide repeats demonstrated higher accuracy to discriminate between controls and CMMRD cases (including *MSH6* biallelic carriers) than previously reported methods,[11–13] requires low DNA input (less than 100 ng), and have an estimated turnaround time of 1 week (online supplementary table S6). In addition, the result obtained with a CMMRD individual's oral mucosa sample suggests its potential for the analysis of MSI in minimally invasive samples, patients with lymphopenia or after allogenic bone marrow transplant. Moreover, the hs-MSI approach is able to robustly discriminate between CMMRD and LS, Li-Fraumeni, NF1 and PPAP, which may assist in classifying cases with overlapping phenotype.[4 5]

The use of a control baseline eliminates the need for paired normal-tumour samples required in other NGS-based MSI analyses.[17] Our method builds on the mSINGS tool[18] using the frequencies of allele lengths different from wildtype allele in contrast to the absolute number of repeat lengths in control baseline, allowing accurate detection of low-level MSI in normal tissues indicating CMMRD. Recently, another NGS-based approach has been developed for MSI detection in blood samples of patients with CMMRD.[13] A good correlation of MSI scores between both approaches was seen in shared samples provided by the C4CMMRD consortium (R^2=0.91; online supplementary figure S7), suggesting that NGS-based hs-MSI assays can reliably detect CMMRD. Interestingly, our MSI score did not overlap between CMMRD samples and controls even in aplastic samples.[13] The improved separation is likely due to the higher number of microsatellite markers analysed, although marker selection, bioinformatics pipeline and analysis method might also be involved.

The high accuracy and suitable turnaround time of the hs-MSI approach, similar to the recently reported in vitro MMR assay in lymphocytes,[6] makes it a valuable CMMRD diagnostic tool. Since CMMRD can present with non-malignant features that overlap with NF1 and Legius syndrome, our approach could be used as a CMMRD indicator in healthy children under suspicion without germline mutations in *NF1* or *SPRED1*, prior to MMR genes analysis, avoiding the potential pitfalls linked to the diagnosis of LS in a minor or VUS identification.[7]

The detection of MMR deficiency or elevated MSI score in lymphocytes may suggest pathogenicity of identified germline MMR variants (online supplementary table S7). However, caution should be taken since other variants in cis (not yet identified) could be responsible of the phenotype. The most intriguing variant in our series is *MLH1* c.2146G>A (p.Val716Met). The presence of an additional causative variant on this *MLH1* allele in patient E was excluded by transcript analysis.[13] Although it was classified as neutral by multifactorial analysis, its identification in trans with a pathogenic *MLH1* mutation in another individual with CMMRD clinical features,[19] and its slightly decreased expression and MMR activity observed in heterologous systems (http://www.insight-database.org/classifications), suggest that its classification should be revisited, particularly since a hypomorphic nature cannot be totally excluded.

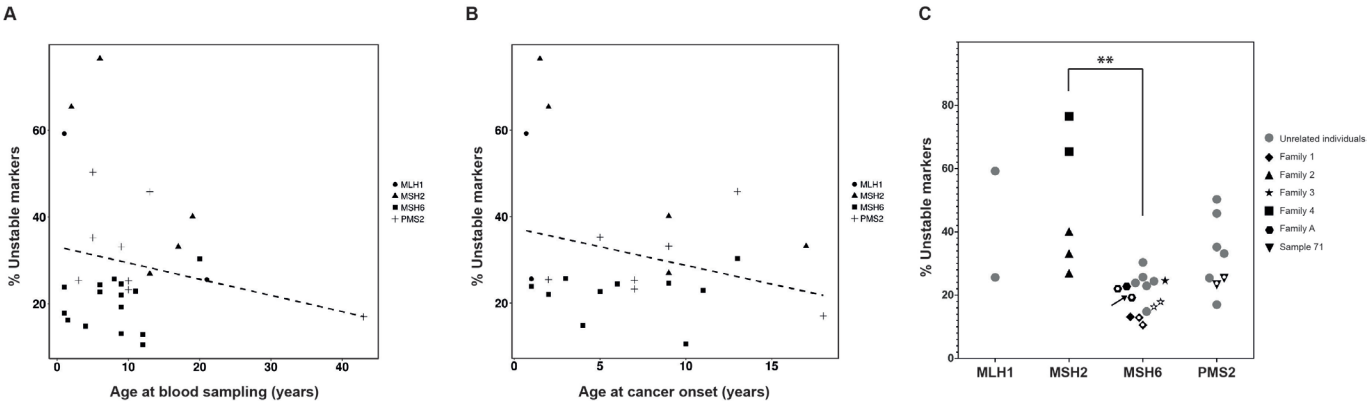


Figure 2 Characterisation of the hs-MSI observed in CMMRD samples. Monomorphic microsatellite markers (selected as frequently mutated in MSI-H tumours) have been analysed (n=186). (A) MSI score in CMMRD blood samples plotted against patient age at blood sampling. No correlation was observed (dashed line, r=−0.04, p=0.823). (B) MSI score in CMMRD blood samples plotted against age of cancer onset. No correlation was observed (dashed line, r=−0.15, p=0.491). (C) MSI score in CMMRD samples plotted against the germline mutated MMR gene. Samples from the same family are indicated by the same symbol. White dots inside symbols indicate samples from the same individual. The buccal mucosa sample is indicated by an arrow. Statistically significant differences between affected genes are indicated (**p<0.005). CMMRD, constitutional mismatch repair deficiency; hs-MSI, high-sensitivity microsatellite instability; MMR, mismatch repair; MSI-H, tumours with high instability.

Mutations in *MSH2* and *MLH1* are associated with a more severe phenotype than *MSH6* and *PMS2* mutations in LS,[1] and this may hold true also in CMMRD,[3] although phenotype/genotype correlation in the latter is complicated by its low prevalence and the presence of hypomorphic MMR mutations.[5] Even though MSI in *MSH6* carriers is more precisely assessed in mononucleotide than dinucleotide repeats, higher instability levels were detected in *MSH2* biallelic carriers than *MSH6* carriers in our hs-MSI approach. Although the limited sample size precludes any conclusion, the MSI level may reflect the intrinsic *MSH6* protein repair capacity of the particular type of markers included in the panel or could be related to disease expressivity. In contrast, no apparent differences by affected gene were observed in CMMRD lymphocytes' MMR assay,[6] which assesses the repair of a 3'-nicked G-T mismatch. Interestingly, with this method intermediate results of MMR activity and complementation were identified in some individuals, suggesting variant hypomorphic nature.[6] The analysis of hs-MSI in these cases would be of particular interest.

In contrast to the absence of significant instability seen in LS samples using the hs-MSI approach, previous works described low-level MSI in blood samples by small-pool PCR[15] and clone analysis.[14] Although those markers were included in our custom panel design, none of them could be analysed due to insufficient coverage, with the exception of D2S123, which did not show instability in LS samples (online supplementary table S5). To improve the sensitivity of MSI assessment, the use of probes with double unique molecular identifiers tagging double-strands[20] may potentially reduce the error rate and increase the sensitivity in MSI detection.

In conclusion, the high performance of the hs-MSI approach in detecting MSI in non-neoplastic tissue from patients with CMMRD is a valuable diagnostic tool which has potential in pretest selection of healthy paediatric patients, as well as in discrimination between CMMRD and other clinically related syndromes. Further evaluation in larger prospective series, including other target tissues and different disease progression stages, is needed to validate the hs-MSI approach in CMMRD diagnostic routine.

Author affiliations
[1]Hereditary Cancer Program, Catalan Institute of Oncology - ICO, Hereditary Cancer Group, Molecular Mechanisms and Experimental Therapy in Oncology Program, Institut d'Investigació Biomèdica de Bellvitge – IDIBELL, Ciber Oncología (CIBERONC) - Instituto de Salud Carlos III, L'Hospitalet de Llobregat, , Spain
[2]Pediatric Oncology Unit, Hospital Sant Joan de Déu, Esplugues, Barcelona, Spain
[3]Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
[4]Hereditary Cancer Program, Catalan Institute of Oncology - ICO, Institut d'Investigació Biomèdica de Girona - IDIBGI, Girona, Spain
[5]Cancer Prevention and Control Program, Catalan Institute of Oncology - ICO, Institut d'Investigació Biomèdica de Bellvitge – IDIBELL, CIBERESP, L'Hospitalet de Llobregat, Barcelona, Spain
[6]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain
[7]Immunodeficiencies Unit, Department of Pediatrics, Doce de Octubre University Hospital, i+12 Research Institute; Complutense University of Madrid, Madrid, Spain
[8]Hematology and Oncology Unit, Department of Pediatrics, Doce de Octubre University Hospital, Madrid, Spain
[9]Department of Hematology, Doce de Octubre University Hospital, i+12 Research Institute, Madrid, Spain
[10]University Department of Pediatrics, CHR Citadelle, Liege, Belgium
[11]Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany
[12]Department of Human Genetics, Hannover Medical School, Hannover, Germany
[13]Department of Pediatrics, Sana Kliniken Duisburg, Duisburg, Germany
[14]Department of Pediatric Oncology, Hematology and Transplantation, Poznań University of Medical Sciences, Poznań, Poland
[15]Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria
[16]Pediatrics Department, Hematology-Oncology Unit, Faculty of Medicine, Ain Shams University, Cairo, Egypt
[17]Pediatric Hematology and Oncology, Klinikum Kassel, Kassel, Germany
[18]Pediatric Oncology Center, Department of Pediatrics, Technische Universität München, Munchen, Germany
[19]Institut für Klinische Genetik, Olgahospital, Stuttgart, Germany
[20]Department of Pediatric Oncology/Pediatric Hematology, Kliniken der Stadt Köln gGmbH, Children's Hospital Amsterdamer Strasse, Koln, Germany
[21]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
[22]Centre de Génétique Humaine, Institut de Pathologie et de Génétique (IPG), Gosselies, Belgium
[23]Pediatric Oncology, Helios-Klinikum, Krefeld, Germany
[24]Department of Pediatric Oncology, Kocaeli Universitesi, Kocaeli, Turkey
[25]Hereditary Cancer Laboratory, Doce de Octubre University Hospital, i+12 Research Institute, Madrid, Spain
[26]Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria

Acknowledgements We thank the participating patients and families.

Contributors MG-A, FM and BP designed and performed the research, analysed the data and wrote the manuscript. NB assisted in bioinformatics analyses. AF assisted in molecular analyses. MaN, HS, FB, SI, AV, EGG, VM, LIG-G, PG-G, BF, CK, TiR, ThR, DJ-L, AAA, IR, MiN, H-JP, SL, MS, KD, TI, UD and JB provided samples and clinical data. VM also contributed to the statistical analysis. RA assisted in molecular analyses. CL, DR and KW provided samples, supervised the study and wrote the manuscript. MP and GC conceived the project, supervised the study, analysed the data and wrote the manuscript. All authors revised and approved the manuscript. MP and GC shared the last authorship.

Funding This work was funded by the Spanish Ministry of Economy and Competitiveness and cofunded by FEDER funds - a way to build Europe (grant SAF2015-68016-R), CIBERONC and the Government of Catalonia (grants 2017SGR1282 and PERIS SLT002/16/0037), and the AECC fellowship to MG-A. AF was supported by a grant from the Catalonian Health Department (SLT002/16/00409). FM was supported by CIBERONC. We thank the CERCA Programme for institutional support.

Competing interests None declared.

Patient consent for publication Obtained.

Ethics approval The study was approved by the ethics committee of the Bellvitge University Hospital (HUB) (file no PR255/15). Written informed consent was obtained from all individuals.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* 2015;15:181–94.
2 Bakry D, Aronson M, Durno C, Rimawi H, Farah R, Alharbi QK, Alharbi M, Shamvil A, Ben-Shachar S, Mistry M, Constantini S, Dvir R, Qaddoumi I, Gallinger S, Lerner-Ellis J, Pollett A, Stephens D, Kelies S, Chao E, Malkin D, Bouffet E, Hawkins C, Tabori U. Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency Consortium. *Eur J Cancer* 2014;50:987–96.
3 Wimmer K, Kratz CP, Vasen HFA, Caron O, Colas C, Entz-Werle N, Gerdes A-M, Goldberg Y, Ilencikova D, Muleris M, Duval A, Lavoine N, Ruiz-Ponte C, Slavc I, Burkhardt B, Brugieres L, EU-Consortium Care for CMMRD (C4CMMRD). Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'care for CMMRD' (C4CMMRD). *J Med Genet* 2014;51:355–65.
4 Ahn DH, Rho JH, Tchah H, Jeon I-S. Early onset of colorectal cancer in a 13-year-old girl with Lynch syndrome. *Korean J Pediatr* 2016;59:40–2.
5 Li L, Hamel N, Baker K, McGuffin MJ, Couillard M, Gologan A, Marcus VA, Chodirker B, Chudley A, Stefanovici C, Durandy A, Hegele RA, Feng B-J, Goldgar DE, Zhu J, De Rosa M, Gruber SB, Wimmer K, Young B, Chong G, Tischkowitz MD, Foulkes WD. A homozygous PMS2 founder mutation with an attenuated constitutional mismatch repair deficiency phenotype. *J Med Genet* 2015;52:348–52.
6 Shuen AY, Lanni S, Panigrahi GB, Edwards M, Yu L, Campbell BB, Mandel A, Zhang C, Zhukova N, Alharbi M, Bernstein M, Bowers DC, Carroll S, Cole KA, Constantini S, Crooks B, Dvir R, Farah R, Hijiya N, George B, Laetsch TW, Larouche V, Lindhorst S, Luiten RC, Magimairajan V, Mason G, Mason W, Mordechai O, Mushtaq N, Nicholas G, Oren M, Palma L, Pedroza LA, Ramdas J, Samuel D, Wolfe Schneider K, Seeley A, Semotiuk K, Shamvil A, Sumerauer D, Toledano H, Tomboc P, Wierman M, Van Damme A, Lee YY, Zapotocky M, Bouffet E, Durno C, Aronson M, Gallinger S, Foulkes WD, Malkin D, Tabori U, Pearson CE. Functional repair assay for the diagnosis of constitutional mismatch repair deficiency from non-neoplastic tissue. *J Clin Oncol* 2019:JCO1800474.
7 Suerink M, Ripperger T, Messiaen L, Menko FH, Bourdeaut F, Colas C, Jongmans M, Goldberg Y, Nielsen M, Muleris M, van Kouwen M, Slavc I, Kratz C, Vasen HF, Brugières L, Legius E, Wimmer K. Constitutional mismatch repair deficiency as a differential diagnosis of neurofibromatosis type 1: consensus guidelines for testing a child without malignancy. *J Med Genet* 2019;56:53–62.
8 Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, Church JM, Dominitz JA, Johnson DA, Kaltenbach T, Levin TR, Lieberman DA, Robertson DJ, Syngal S, Rex DK. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task force on colorectal cancer. *Am J Gastroenterol* 2014;109:1159–79.
9 Vasen HFA, Ghorbanoghli Z, Bourdeaut F, Cabaret O, Caron O, Duval A, Entz-Werle N, Goldberg Y, Ilencikova D, Kratz CP, Lavoine N, Loeffen J, Menko FH, Muleris M, Sebille G, Colas C, Burkhardt B, Brugieres L, Wimmer K, EU-Consortium Care for CMMR-D (C4CMMR-D). Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D). *J Med Genet* 2014;51:283–93.
10 Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, Van Loo P, Tarpey PS, Coupland P, Behjati S, Pollett A, Lipman T, Heidari A, Deshmukh S, Avitzur Na'ama, Meier B, Gerstung M, Hong Y, Merino DM, Ramakrishna M, Remke M, Arnold R, Panigrahi GB, Thakkar NP, Hodel KP, Henninger EE, Göksenin AY, Bakry D, Charames GS, Druker H, Lerner-Ellis J, Mistry M, Dvir R, Grant R, Elhasid R, Farah R, Taylor GP, Nathan PC, Alexander S, Ben-Shachar S, Ling SC, Gallinger S, Constantini S, Dirks P, Huang A, Scherer SW, Grundy RG, Durno C, Aronson M, Gartner A, Meyn MS, Taylor MD, Pursell ZF, Pearson CE, Malkin D, Futreal PA, Stratton MR, Bouffet E, Hawkins C, Campbell PJ, Tabori U, Biallelic Mismatch Repair Deficiency Consortium. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet* 2015;47:257–62.
11 Ingham D, Diggle CP, Berry I, Bristow CA, Hayward BE, Rahman N, Markham AF, Sheridan EG, Bonthron DT, Carr IM. Simple detection of germline microsatellite instability for diagnosis of constitutional mismatch repair cancer syndrome. *Hum Mutat* 2013;34:847–52.
12 Bodo S, Colas C, Buhard O, Collura A, Tinat J, Lavoine N, Guilloux A, Chalastanis A, Lafitte P, Coulet F, Buisine M-P, Ilencikova D, Ruiz-Ponte C, Kinzel M, Grandjouan S, Brems H, Lejeune S, Blanché H, Wang Q, Caron O, Cabaret O, Svrcek M, Vidaud D, Parfait B, Verloes A, Knappe UJ, Soubrier F, Mortemousque I, Leis A, Auclair-Perrossier J, Frébourg T, Fléjou J-F, Entz-Werle N, Leclerc J, Malka D, Cohen-Haguenauer O, Goldberg Y, Gerdes A-M, Fedhila F, Mathieu-Dramard M, Hamelin R, Wafaa B, Gauthier-Villars M, Bourdeaut F, Sheridan E, Vasen H, Brugières L, Wimmer K, Muleris M, Duval A, European Consortium "Care for CMMRD". Diagnosis of constitutional mismatch Repair-Deficiency syndrome based on microsatellite instability and lymphocyte tolerance to methylating agents. *Gastroenterology* 2015;149:1017–29.
13 Gallon R, Mühlegger B, Wenzel S-S, Sheth H, Hayes C, Aretz S, Dahan K, Foulkes W, Kratz CP, Ripperger T, Azizi AA, Baris Feldman H, Chong A-L, Demirsoy U, Florkin B, Imschweiler T, Januszkiewicz-Lewandowska D, Lobitz S, Nathrath M, Pander H-J, Perez-Alonso V, Perne C, Ragab I, Rosenbaum T, Rueda D, Seidel MG, Suerink M, Taeubner J, Zimmermann S-Y, Zschocke J, Borthwick GM, Burn J, Jackson MS, Santibanez-Koref M, Wimmer K. A sensitive and scalable microsatellite instability assay to diagnose constitutional mismatch repair deficiency by sequencing of peripheral blood leukocytes. *Hum Mutat* 2019;40:649–55.
14 Alazzouzi H, Domingo E, González S, Blanco I, Armengol M, Espín E, Plaja A, Schwartz S, Capella G, Schwartz S. Low levels of microsatellite instability characterize MLH1 and MSH2 HNPCC carriers before tumor diagnosis. *Hum Mol Genet* 2005;14:235–9.
15 Coolbaugh-Murphy MI, Xu J-P, Ramagli LS, Ramagli BC, Brown BW, Lynch PM, Hamilton SR, Frazier ML, Siciliano MJ. Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. *Hum Mutat* 2010;31:317–24.
16 Staffa L, Echterdiek F, Nelius N, Benner A, Werft W, Lahrmann B, Grabe N, Schneider M, Tariverdian M, von Knebel Doeberitz M, Bläker H, Kloor M. Mismatch repair-deficient crypt foci in Lynch syndrome--molecular alterations and association with clinical parameters. *PLoS One* 2015;10:e0121980.
17 Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 2014;30:1015–6.
18 Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem* 2014;60:1192–9.
19 Marcos I, Borrego S, Urioste M, García-Vallés C, Antiñolo G. Mutations in the DNA mismatch repair gene MLH1 associated with early-onset colon cancer. *J Pediatr* 2006;148:837–9.
20 Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* 2012;109:14508–13.

SUPPLEMENTARY METHODS

ASSESSMENT OF THE ANALYTICAL SENSITIVITY OF SNV AND MSI DETECTION BY USING A MOLECULAR BARCODING BASED NGS APPROACH

Cell lines

A HCT15 cell line was cultured in RPMI 1640 medium, whereas RKO and SW480 lines were cultured in DMEM, both supplemented with 10% Fetal Bovine Serum (Thermo Fisher Scientific, USA) and 1% PenStrep (100 U/mL penicillin and 100 µg/mL streptomycin; Thermo Fisher Scientific), maintained in humidified 37°C 5% CO_2 incubators. Genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) according to manufacturer´s instructions. DNA quality was assessed using a Nanondrop ND 1000 Spectrophotometer (Thermo Fisher Scientific) and agarose gel electrophoresis. Extracted DNA was quantified by Qubit Flurometer with the dsDNA BR Assay (Invitrogen, Carlsbad, CA, USA).

ClearSeq Cancer HS panel

The ClearSeq Cancer HS panel (Agilent Technologies, USA) was used to assess the analytical sensitivity of single nucleotide variants (SNV) and microsatellite instability (MSI) detection. It is a 79.5 kbp commercial panel targeting COSMIC hotspots in 47 genes. The panel utilizes HaloPlexHS, an amplicon-based target enrichment method designed to detect low-frequency allelic variants by the addition of a degenerate 10-nucleotide-long molecular barcode index to the captured DNA fragments.

Sequencing libraries were prepared by using serial dilutions (100%, 50%, 10%, 2%, 0.4%, 0%) of DNA extracted from the microsatellite unstable cell lines (RKO or HCT15) in DNA extracted from the microsatellite-stable SW480 cell line. Library preparation of the target regions in the cell line DNA mixtures was performed using HaloPlex HS Target Enrichment kit (Agilent Technologies, USA), according to the HaloPlex HS Target Enrichment System For Illumina Sequencing Protocol (Version C0, December 2015). Library pools were prepared with four samples for a final concentration of 4 nM. Pools were quantified by TapeStation (Agilent Technologies, USA), denatured, and sequenced on the Illumina MiSeq sequencer using the 150 bp paired-end sequencing protocol using 15 pM seeding concentration and 10% PhiX spike-in with v3 cartridges and 10 pM seeding concentration and 10% PhiX spike-in with v2 cartridges.

Bioinformatic pipeline and estimated limit of detection

Alignment (SureCall pipeline): Pair-ended sequence reads in FASTQ files from Illumina MiSeq were initially processed with Agilent software Surecall v3.5. Afterward, the reads were aligned to the hg19 human reference genome (February 2009 assembly) using the default parameters of bwa-mem. SureCall performs the molecular barcode analysis and the deduplication process.

SNV Calling (SureCall pipeline): The SNPPET algorithm uses sorted BAM files to perform the SNV calling. Variants were filtered for quality differences between alternate and reference alleles, proximity to 3’ ends, and other quality filters. The parameters were set to detect low frequency variants while minimizing false positives.

SNV and microsatellite (MS) indel calling (custom pipeline): Consensus reads generated from 1 to 3 original reads per molecular barcode were filtered out from the sorted BAM file. Sample-level, fully local indel realignment was performed using the Genome Analysis Toolkit- 3.7.0 (GATK). In a first step we created the intervals to be realigned with the "RealignerTargetCreator" tool using as input the sample BAM file, the reference genome and a gold standard for known indels (Mills_and_1000G_gold_standard.indels.b37.sites.vcf). In a second step the reads are realigned over the intervals created in the previous step using the "IndelRealigner tool". In both cases, the default parameters were used. After filtering and realigning, SAMtools version 1.3.1 "mpileup" command generated mpileup files for each of the final BAM files.

SNV (substitutions) were called using VarScan2 version 2.4.3 (http://varscan.sourceforge.net)[1] "mpileup2snp" command. We set the minimum variant frequency to 0.001, the minimum read depth to make a call to 30, and the minimum number of reads supporting the variant to 3. The variants identified by VarScan2 were further filtered using the "fpfilter" tool, a false positive filter included in the variant caller software, for different quality and position parameters. Only variants in target regions were annotated.

Twelve and 9 unique SNVs respectively, not present in SW480 cell line, were called in HCT15 and RKO cell lines. One out of the 12 SNV in 100% HCT15 was only detected by VarScan2 at a low frequency. The frequencies of identified SNVs decreased proportionally to the factor of dilution (Table S8). The limit of detection was established at 0.004 since possible false variants, only present in one point of the series, were detected at very low frequencies (0.001-0.003) (data not shown). Taking into account the two replicas performed with DNA mixtures from HCT15 and SW480, all variants were detected at a frequency below 0.01 (Table S9).

The MSIseq package in R was used to locate MS (only mononucleotide repeat (MNR) type) greater than 5 bp across all sequences included in the panel. This analysis revealed 499 MNRs of a maximum length of 18 bp mostly located in low-covered regions that were further analyzed for MSI. MNR indel calling was performed following a previously described approach.[2] Absolute frequencies for each allele of the MNR were calculated using a custom R script. Seven unstable or polymorphic microsatellites loci fulfilling previous criteria were identified in DNA mixtures from HCT15 and SW480 DNA cell lines (Figure S8A) that were validated by DNA fragment analysis and fluorescent capillary electrophoresis (Figure S8B). Stutter peaks caused by polymerase slippage during PCR or sequencing errors were observed in electropherograms but they were not always present in the analysis by Haloplex NGS, which demonstrated higher specificity. Although some unstable alleles were detected at a very low frequency (below 0.01), the limit of detection needed to be established independently for each microsatellite.

ASSESSMENT OF MICROSATELLITE INSTABILITY AT HIGH SENSITIVITY (hs-MSI) BY USING A CUSTOM PANEL

Custom NGS panel design and sequencing

Two hundred seventy-seven microsatellite targets (91% MNRs) were included in the custom panel. Selected microsatellite target regions included: 1) MS markers frequently mutated in MSI-H tumors: MNR with increased mutation frequency in MSI-H colorectal cancer, gastric cancer, endometrial cancer and colon cell culture tumors according to SelTar*base* (http://seltarbase.org, release 201307)[3], and additional microsatellites that were present in >15% of MSI-H CRC by whole-exome sequencing[4]; 2) MS markers from MSI diagnostic panels: microsatellite loci from the Bethesda panel and Promega MSI Analysis System version 1.2[5], 59 MNR sites included in the reported MSI-detection panel by Zhao et al.[6], and MT1X (T)20[6]; 3) MS within antigen presentation genes according to SelTar*base*; 4) Additional published MS targets of interest.[7, 8, 9, 10]

The custom panel was designed for the HaloPlex HS target enrichment using SureDesign (Agilent, Santa Clara, v4.0.0.18). Target regions included 50bp flanking regions around the microsatellite. Only probes for "on-target" amplicons that covered the entire microsatellite region with 10bp flanks were considered. Microsatellite target regions not covered by more than one "on-target" amplicon were excluded. Library pools were prepared with 26 samples for a final concentration of 25 nM. Pools were quantified by TapeStation (Agilent Technologies, USA), denatured, and sequenced at high depth (20.000x) on an Illumina HiSeq2500 sequencer using the 150 bp paired-end sequencing protocol, 9.5 pM seeding concentration and 25% PhiX spike-in. After deduplication process, the mean depth (±SD) was 1312 ± 447 reads/marker/sample.

Optimization of the bioinformatics pipeline for microsatellite indel calling

The bioinformatics pipeline described above for microsatellite indel calling was further optimized. After SureCall alignment, all reads that did not cover completely the microsatellite were filtered out from the BAM file using SAMtools prior to indel calling. The PCR and sequencing error and/or basal instability for each MS locus was assessed in 22 healthy control blood DNA samples. The minimum read depth for each microsatellite loci was set to 100 and no minimum of reads supporting indels was established in order to capture all PCR and sequencing errors. Only microsatellites with valid data in at least 5 out of 22 control samples (256 out of 277) were considered. Six MS were additionally excluded because the wildtype allele in controls did not match the reference genome. Mean frequencies plus 3 SD were calculated for each microsatellite allele, including the wildtype, and they were used as reference values for case sample analysis.

In order to increase the sensitivity in MSI assessment, 231 out of 256 MS, truly monomorphic in the baseline (with a mean frequency of wildtype allele above 0.94) were selected. Among them, 186 markers (93% MNR 5-14bp of lenght) were previously reported as frequently mutated in MSI-H tumors.[3, 4]

Indel calling in case samples was performed by using the pipeline described above for the analysis of baseline control samples with minor modifications to add more restrictive

parameters: the minimum frequency was set to 0.004 (since we had previously established this limit of detection for SNV), and the minimum number of reads supporting the indel was set to 3.

MSI classification system

To assess the hs-MSI status in each microsatellite locus, the instability level, corresponding to the sum of the frequencies of all allele lengths different from the wildtype (mutational load method), was calculated as (1 – wildtype allele frequency). This value was compared with the reference value (1 – mean wildtype frequency in baseline). Whenever the instability level exceeded the mean value in baseline plus 3 SD and the highest value among the individual samples of the baseline the MS was considered unstable.

As an alternative, the individual allele method was evaluated. Here the frequencies of each alternative microsatellite allele length (instead of the sum of them) were compared against their respective length frequency in the baseline. If at least one of the individual allele length frequencies in a case sample exceeded the baseline mean frequency plus 3 SD and the highest value among the individual samples of the baseline, then the MS was considered unstable.

For both methods, an hs-MSI score was calculated per case sample, representing the percentage of unstable microsatellites out of the total number of valid microsatellite markers (minimum read depth of 100 per microsatellite locus). A threshold for identifying MSI status (positive/negative) in case samples was set to the mean hs-MSI score of the 15 healthy controls included in the training set plus 3 SD.

Hs-MSI score median was compared between different training set groups using a Wilcoxon Rank Sum Test and according to germline affected gene using a Kruskal-Wallis test followed by a Dunn's multiple pairwise comparisons test (Bonferroni correction). MSI score association with age at blood sampling and age at onset was analyzed using the Spearman's correlation coefficient (rho, rs). All the analyses were performed in R software (http://www.R-project.org).

ANALYSIS OF DINUCLEOTIDE REPEATS

Germline MSI (gMSI) analysis

Germline microsatellite instability (gMSI) analysis was performed as described in Ingham et al. 2013[9] in training cohort CMMRD and 22 baseline control blood samples. PCR amplification of the dinucleotide microsatellites D17S791, D2S123, and D17S250 was performed. PCR products were analyzed on an Applied Biosystems 3130XL Genetic Analyzer using GeneMapper software (Applied Biosystems, Forster City, California, USA). The gMSI ratios were determined by dividing the height of an allele's trailing "stutter" peak (n+1) by the height of the allele's major peak (n). A gMSI ratio threshold was chosen for each microsatellite that ensured a high specificity and sensitivity.[9] On the basis of the sensitivity and specificity data, elevated gMSI ratios were defined as follows: >0.060 for D2S123 (mean + 4 SDs), >0.069 for D17S250 (mean + 4 SDs) and >0.117 for D17S791 (mean + 3 SDs). In CMMRD samples, observed gMSI ratio minus the marker-specific threshold was calculated; positive values represented ratios above the

threshold. If two or more markers were above the threshold, the sample was classified as gMSI positive.

Analysis of dinucleotide repeats from NGS data

Analysis of dinucleotide D2S123 was carried out by read counting in IGV v.2.4.10 platform using hg19 reference genome. The zygosity of each marker was previously determined in the gMSI analysis (see above). The level of instability was calculated as the percentage of the sum of reads of all allele lengths different from the wild-type, divided by the wildtype reads x 100. The instability threshold value for D2S123 was defined as >8.23 (mean + 3 SDs).

CLONALITY TESTING OF BLOOD SAMPLES FROM CMMRD PATIENTS

Presence of lymphoproliferative clones in CMMRD patients' blood samples was evaluated using standard BIOMED-2 assay.[11] Briefly, clonally rearranged immunoglobulin and T-cell receptor genes were assayed by multiplex PCR with fluorescence primer sets for IGH VH-JH, IGH DH-JH, IGK, IGL, TCRB, TCRG, and TCRD rearrangements and size-resolved by capillary electrophoresis.

REFRENCES

1 Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22(3):568-76.

2 Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem* 2014;60(9):1192-9.

3 Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med* 2016;22(11):1342-50.

4 Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, Bahl S, Cao Y, Amin-Mansour A, Yamauchi M, Sukawa Y, Stewart C, Rosenberg M, Mima K, Inamura K, Nosho K, Nowak JA, Lawrence MS, Giovannucci EL, Chan AT, Ng K, Meyerhardt JA, Van Allen EM, Getz G, Gabriel SB, Lander ES, Wu CJ, Fuchs CS, Ogino S, Garraway LA. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* 2016;17(4):1206.

5 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2(5):401-4.

6 Zhao H, Thienpont B, Yesilyurt BT, Moisse M, Reumers J, Coenegrachts L, Sagaert X, Schrauwen S, Smeets D, Matthijs G, Aerts S, Cools J, Metcalf A, Spurdle A, Amant F, Lambrechts D. Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks. *Elife* 2014;3:e02725.

7 Coolbaugh-Murphy M, Maleki A, Ramagli L, Frazier M, Lichtiger B, Monckton DG, Siciliano MJ, Brown BW. Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. *Genomics* 2004;84(2):419-30.

8 He S, Zhao Z, Yang Y, O'Connell D, Zhang X, Oh S, Ma B, Lee JH, Zhang T, Varghese B, Yip J, Dolatshahi Pirooz S, Li M, Zhang Y, Li GM, Ellen Martin S, Machida K, Liang C. Truncating mutation in the autophagy gene UVRAG confers oncogenic properties and chemosensitivity in colorectal cancers. *Nat Commun* 2015;6:7839.

9 Ingham D, Diggle CP, Berry I, Bristow CA, Hayward BE, Rahman N, Markham AF, Sheridan EG, Bonthron DT, Carr IM. Simple detection of germline microsatellite instability for

diagnosis of constitutional mismatch repair cancer syndrome. *Hum Mutat* 2013;34(6):847-52.

10 Ripberger E, Linnebacher M, Schwitalle Y, Gebert J, von Knebel Doeberitz M. Identification of an HLA-A0201-restricted CTL epitope generated by a tumor-specific frameshift mutation in a coding microsatellite of the OGT gene. *J Clin Immunol* 2003;23(5):415-23.

11 van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003;17(12):2257-317.

Supplemental Table S1. Clinico-molecular characteristics of patients included in the training cohort and summary of the obtained results. Abbreviations: preT-LBL: precursor T-cell lymphoblastic lymphoma; T-LBL: T-cell lymphoblastic lymphoma; ALL: acute lymphoblastic leukaemia; sPNET, supratentorial primitive neuroectodermal tumor; B-ALL, B-cell acute lymphoblastic leukaemia; T-NHL, T-cell non-Hodgkin lymphoma; EC, endometrial cancer; CRC: colorectal cancer; NHL: non-Hodgkin lymphoma; OC, ovarian cancer; LC, liver cancer; BC, breast cancer; y, years; m, months; NP, not performed; gMSI, germline microsatellite instability; gMSI detection: -, no detection; +, detection; hs-MSI, high-sensitivity microsatellite instability; hs-MSI detection: -, no detection; +, detection. (*) Samples previously analyzed in Gallon et al., 2019. (‡) Another sample from the same patient was previously analyzed in Gallon et al., 2019.

Patient ID	Family ID	MMR affected gene	Germline variant (cDNA)	Germline variant (protein)	Available sample	Age at sampling	Cancer diagnosis before blood sampling	Disease stage at blood sampling	Tumours (age of onset)	Result of clonality testing	gMSI (according to Ingham et al., 2013)				hs-MSI score		
											D2S123	D17S250	D17S791	gMSI detection	186 MS	231 MS	hs-MSI detection
CMMRD:																	
CMMRD-01 (91‡)	Family A	*MSH6*	c.[2653A>T];[2653A>T]	p.[Lys885*];[Lys885*]	Blood	9y	Yes	Under chemotherapy	Burkitt lymphoma (2y), preT-LBL (3y and 8y)	Polyclonal	-0.01	-0.02	-0.04	-	22.03	16.25	+
					Buccal mucosa	9y	Yes	Under chemotherapy			NP	NP	NP	NP	19.23	14.34	+
CMMRD-02	Family A	*MSH6*	c.[2653A>T];[2653A>T]	p.[Lys885*];[Lys885*]	Blood	6y	Yes	Under chemotherapy	Wilms tumour (5y), T-LBL (8y)	NP	-0.02	-0.02	-0.02	-	22.73	17.87	+
CMMRD-03 (49‡)		*MLH1*	c.[332C>T]; [332C>T]	p.[Ala111Val];[Ala111Val]	Blood	15m	Yes	Under chemotherapy	T-LBL (7m)	Polyclonal	0.09	0.11	-0.03	+	59.22	45.80	+
CMMRD-04 (51‡)		*PMS2*	c.[2007-2A>G];[2007-2A>G]		Blood	3y	Yes	Cancer affected	ALL (2y), Glioblastoma (3y)	NP	0.02	0.17	0.34	+	25.42	19.49	+
CMMRD-05 (71*)		*PMS2*	c.[862C>T];[862C>T]	p.[Gln288*];[Gln288*]	Blood	10y	Yes	Cancer affected	sPNET (7y), Sigmoidal adenocarcinoma (10y)	NP	0.03	0.13	0.21	+	23.26	18.34	+
CMMRD-06 (A*)		*PMS2*	Complex rearrangement	Fusion transcript from *PMS2* exon 10 into a sequence derived from intron 14 of *CCZ1B* gene	Blood	5y	Yes	Cancer affected	Glioblastoma (5y)	NP	0.11	0.26	0.36	+	35.20	27.50	+
CMMRD-07 (43‡)		*MSH6*	c.[467C>G];[1316A>G]	p.[Ser156*];[Asp439Gly]	Blood	4y	Yes	Cancer affected	Anaoplastic medulloblastoma (4y)	NP	-0.03	-0.01	-0.03	-	14.84	11.43	+
CMMRD-08 (83*)		*MSH6*	c.[1135_1139del];[2277_2293del]	p.[Arg379*];[Glu760Profs*6]	Blood	8y	Yes	Healthy	B-ALL (3y), T-NHL (7y)	NP	-0.02	-0.02	-0.02	-	25.70	19.83	+
CMMRD-09 (B‡)		*MSH6*	c.[2238dup];[2980T>A]	p.[Leu747Serfs*9];[Tyr994Asn]	Blood	6y	Yes	Cancer affected	Medulloblastoma (6y)	NP	-0.01	-0.01	-0.02	-	24.44	18.11	+
CMMRD-10		*PMS2*	c.[24-2A>G];[24-2A>G]	p.[Ser8Argfs*4];[Ser8Argfs*4]	Blood	43y	Yes	Cancer affected	Lymphoma (18y), Lymphoma (37y), EC (39y), 3 independent CRC (43y), Lymphoma (43y)	NP	0.05	0.05	0.08	+	17.03	13.52	+
CMMRD-11		*MSH6*	c.[742C>T];[742C>T]	p.[Arg248*];[Arg248*]	Blood	1y	Yes	Cancer affected	NHL (1y)	Polyclonal	0.02	-0.01	-0.05	-	23.89	18.18	+
Lynch syndrome:																	
Lynch-01		MLH1	c.1590_1598dup	p.(Gly532_Val534dup)	Blood	54y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
Lynch-02		MLH1	c.1731G>A	p.Ser556Argfs*14	Blood	43y	No	Healthy	No tumor	NP	NP	NP	NP	NP	2.78	2.49	-
Lynch-03	Family B	MLH1	c.1865T>A	p.Leu622His	Blood	43y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.11	0.83	-
Lynch-04		MLH1	c.1865T>A	p.Leu622His	Blood	31y	No	Healthy	No tumor	NP	NP	NP	NP	NP	2.81	2.50	-
Lynch-05	Family C	MLH1	c.199G>A	p.(Gly67Arg)	Blood	25y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
Lynch-06		MLH1	c.208-?_306+?del	p.(?)	Blood	40y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.56	0.41	-
Lynch-07	Family D	MLH1	c.350C>T	p.(Thr117Met)	Blood	35y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
Lynch-08	Family D	MLH1	c.350C>T	p.(Thr117Met)	Blood	52y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.56	0.41	-
Lynch-09	Family E	MLH1	c.676C>T	p.(Arg226*)	Blood	26y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.60	0.45	-
Lynch-10		MSH2	c.2635-5_2635-3inv	p.(?)	Blood	33y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.66	1.23	-
Lynch-11		MSH2	c.1980_1981del	p.(Asp660Glufs*15)	Blood	41y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.55	0.41	-
Lynch-12		MSH2	c.2021dup	p.(Lys675*)	Blood	30y	No	Healthy	No tumor	NP	NP	NP	NP	NP	3.33	2.48	-
Lynch-13		MSH2	c.2222_2223del	p.(Lys741Argfs*8)	Blood	40y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.55	0.41	-
Lynch-14		MSH2	c.367-6_370dup	p.(Ser124Tyrfs*12)	Blood	37y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.55	0.41	-
Lynch-15	Family F	MSH2	c.746_747del	p.(Lys249Argfs*6)	Blood	36y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.56	0.41	-
Lynch-16		MSH6	c.1477G>T	p.(Glu493*)	Blood	37y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.17	0.88	-
Lynch-17	Family G	MSH6	c.1739C>A	p.(Ser580*)	Blood	65y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.10	0.82	-
Lynch-18		MSH6	c.2188dup	p.(Tyr730Leufs*26)	Blood	24y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.10	0.82	-

Lynch-19		MSH6	c.236C>A	p.(Ser79*)	Blood	52y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.56	0.41	-
Lynch-20		MSH6	c.3238_3239del	p.(Leu1080Valfs*12)	Blood	59y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.11	0.83	-
Lynch-21		MSH6	c.3261del	p.(Phe1088Serfs*2)	Blood	41y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.66	1.64	-
Lynch-22		MSH6	c.3261del	p.(Phe1088Serfs*2)	Blood	53y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.68	1.26	-
Lynch-23		PMS2	c.1A>G	p.(?)	Blood	41y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
Lynch-24		PMS2	c.538-3C>G	p.(Glu180Glnfs*5, Glu180_Gln235del)	Blood	38y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
Lynch-25		PMS2	c.780del	p.(Asp261Metfs*46)	Blood	26y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
Lynch-26		PMS2	c.989-2A>G	p.Glu330_Gly381del	Blood	61y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.55	0.41	-
Lynch-27		MLH1	c.1852_1854del	p.(Lys618del)	Blood	39y	Yes	Unknown	CRC (38y)	NP	NP	NP	NP	NP	1.73	1.30	-
Lynch-28		MLH1	c.1865T>A	p.Leu622His	Blood	40y	Yes	Unknown	CRC (39y)	NP	NP	NP	NP	NP	0.57	0.43	-
Lynch-29	Family B	MLH1	c.1865T>A	p.Leu622His	Blood	34y	Yes	Healthy	CRC (33y), EC (39y)	NP	NP	NP	NP	NP	1.11	0.83	-
Lynch-30	Family C	MLH1	c.199G>A	p.(Gly67Arg)	Blood	54y	Yes	Healthy	CRC (39y)	NP	NP	NP	NP	NP	0	0	-
Lynch-31		MLH1	c.306+5G>A	p.[Lys70_Glu102del, Glu102Phefs*18]	Blood	64y	Yes	Healthy	CRC (50y), CRC (55y)	NP	NP	NP	NP	NP	1.63	1.21	-
Lynch-32		MLH1	c.306+5G>A	p.[Lys70_Glu102del, Glu102Phefs*18]	Blood	51y	Yes	Unknown	CRC (40y)	NP	NP	NP	NP	NP	0.56	0.41	-
Lynch-33	Family E	MLH1	c.676C>T	p.(Arg226*)	Blood	51y	Yes	Healthy	EC (39y), CRC (39y), CRC (46y)	NP	NP	NP	NP	NP	2.78	2.07	-
Lynch-34		MSH2	c.518T>G	p.(Leu173Arg)	Blood	44y	Yes	Healthy	OC (43y), CRC (45y)	NP	NP	NP	NP	NP	2.84	2.53	-
Lynch-35		MSH2	c.689_691delinsTT	p.(Ala230Valfs*16)	Blood	46y	Yes	Unknown	OC (42y)	NP	NP	NP	NP	NP	1.11	0.83	-
Lynch-36	Family F	MSH2	c.746_747del	p.(Lys249Argfs*6)	Blood	30y	No	Unknown	LC (34y)	NP	NP	NP	NP	NP	0	0	-
Lynch-37		MSH2	c.942+3A>T	p.(Val265_Gln314del)	Blood	41y	Yes	Unknown	EC (40y)	NP	NP	NP	NP	NP	2.20	1.65	-
Lynch-38		MSH2	Deletion E7-E16	p.(?)	Blood	53y	Yes	Unknown	BC (51y), EC (51y)	NP	NP	NP	NP	NP	1.10	1.22	-
Lynch-39		MSH2	Duplication E11-E16	p.(?)	Blood	64y	Yes	Healthy	CRC (54y)	NP	NP	NP	NP	NP	1.65	1.23	-
Lynch-40	Family G	MSH6	c.1739C>A	p.(Ser580*)	Blood	69y	Yes	Healthy	CRC (69y)	NP	NP	NP	NP	NP	1.10	0.82	-
Lynch-41		MSH6	c.2147_2148del	p.(Thr716Serfs*39)	Blood	47y	Yes	Healthy	CRC (41y), CRC (52y)	NP	NP	NP	NP	NP	0	0.41	-
Lynch-42		MSH6	c.255dup	p.(Thr86Hisfs*4)	Blood	44y	Yes	Unknown	CRC (42y)	NP	NP	NP	NP	NP	0.55	0.4	-
Lynch-43		MSH6	c.2731C>T	p.(Arg911*)	Blood	70y	Yes	Healthy	CRC (68y), PC (73y)	NP	NP	NP	NP	NP	1.10	0.82	-
Lynch-44		MSH6	c.3202C>T	p.(Arg1068*)	Blood	58y	Yes	Healthy	CRC (50y)	NP	NP	NP	NP	NP	2.78	2.06	-
Lynch-45		PMS2	c.164-2A>G	p.(Asp55Alafs*2)	Blood	67y	Yes	Cancer affected	CRC (66y)	NP	NP	NP	NP	NP	1.70	1.70	-
Lynch-46		PMS2	c.2444C>T	p.(Ser815Leu)	Blood	59y	Yes	Healthy	OC (54y)	NP	NP	NP	NP	NP	0	0.41	-
Lynch-47		PMS2	c.903G>T	p.(Tyr268*)	Blood	68y	Yes	Healthy	CRC (68y), CRC (68y)	NP	NP	NP	NP	NP	0.55	0.41	-
Lynch-48		PMS2	C.943C>T	p.(Arg315*)	Blood	53y	Yes	Healthy	CRC (49y)	NP	NP	NP	NP	NP	0	0	-
Healthy controls:																	
CONTROL_01					Blood	65y	No	Healthy	No tumor	NP	NP	NP	NP	NP	2.23	1.79	-
CONTROL_02					Blood	52y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.66	1.33	-
CONTROL_03					Blood	59y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.64	1.32	-
CONTROL_04					Blood	62y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.12	0.9	-
CONTROL_05					Blood	52y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.09	0.87	-
CONTROL_06					Blood	56y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.54	0.44	-
CONTROL_07					Blood	61y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.1	0.88	-
CONTROL_08					Blood	56y	No	Healthy	No tumor	NP	NP	NP	NP	NP	3.89	3.11	-
CONTROL_09					Blood	56y	No	Healthy	No tumor	NP	NP	NP	NP	NP	2.86	2.28	-
CONTROL_10					Blood	51y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.1	0.88	-
CONTROL_11					Blood	52y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0.55	0.44	-

CONTROL_12	Blood	65y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
CONTROL_13	Blood	50y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
CONTROL_14	Blood	50y	No	Healthy	No tumor	NP	NP	NP	NP	NP	0	0	-
CONTROL_15	Blood	50y	No	Healthy	No tumor	NP	NP	NP	NP	NP	1.65	1.32	-

Supplemental Table S2. Clinico-molecular characteristics of patients included in the validation cohort and summary of the obtained results. Abbreviations: T-LBL: T-cell lymphoblastic lymphoma; T-NHL, T-cell non-Hodgkin lymphoma; B-NHL, B-cell non-Hodgkin lymphoma; sPNET, supratentorial primitive neuroectodermal tumor; CRC: colorectal cancer; y, years; m, months; NA, not available; NP, not performed; gMSI, germline microsatellite instability; gMSI detection: -, no detection; +, detection; hs-MSI, high-sensitivity microsatellite instability; hs-MSI detection: -, no detection; +, detection. All these samples were previously analyzed in Gallon et al., 2019.

Sample ID	Patient ID (Family ID)	Affected Gene	Germline variant (cDNA)	Germline variant (protein)	Age at sampling	Cancer diagnosis before blood sampling	Disease stage at blood sampling	Tumours (age)	Result of clonality testing	gMSI (extracted from Gallon et al., 2019)				hs-MSI score		
										D2S123	D17S250	D17S791	gMSI detection	186 MS	231 MS	hs-MSI detection
CMMRD:																
C	1	*PMS2*	c.[1515del];[1515del]	p.[Phe506Serfs*89];[Phe506Serfs*89]	13y	Yes	Cancer affected	B-cell burkitt lymphoma (13y), T-LBL (ND)	Polyclonal	0.30	0.07	0.06	+	45.83	36.07	+
D	3	*MSH6*	c.[3838C>T];[3838C>T]	p.[Gln1280*];[Gln1280*]	20y	Yes	Unknown	Duodenal adenocarcinoma and CRC (13y)	NP	-0.04	-0.02	-0.05	-	30.36	23.29	+
E	5	*MLH1*	c.[62C>A];[2146G>A]	p.[Ala21Glu];[Val716Met]	21y	Yes	Under treatment	T-NHL (14 months), B-NHL (12y), borderline Phyloides tumor (16y), Glioblastoma (21y)	NP	0.02	0.05	0.06	+	25.6	20.18	+
65	6	*MSH6*	c.[691del];[691del]	p.[Val231Tyrfs*15];[Val231Tyrfs*15]	11y	Yes	Cancer affected	Glioblastoma (11y)	NP	-0.07	-0.06	-0.05	-	22.94	17.73	+
93	7	*PMS2*	c.[1145-31_1145-13del];[1145-31_1145-13del]	p.[Asn383*;Gly382Valfs*19];[Asn383*;Gly382Valfs*19]	9y	Yes	Cancer affected	Glioblastoma (9y)	NP	0.04	0.23	0.29	+	33.14	26.09	+
99	8 (F1)	*MSH6*	c.[3557-1G>C];[(457+1_458-1)_(3646+1_3646-1)del]	p.[Glu1187_Gly1216del];[Ser154_Gly1216del]	12y	Yes	Cancer affected (relapse)	T-cell lymphoma (10y), Lymphoma relaps (12y)	Polyclonal	-0.07	-0.05	-0.08	-	10.56	7.85	+
105	8 (F1)				12y	Yes	Cancer affected (relapse)			-0.07	-0.05	-0.07	-	12.92	9.62	+
82	9 (F1)	*MSH6*	c.[3557-1G>C];[(457+1_458-1)_(3646+1_3646-1)del]	p.[Glu1187_Gly1216del];[Ser154_Gly1216del]	9y	No	Healthy	No Tumor	NP	-0.07	-0.05	-0.08	-	13.14	9.75	+
56	10	*PMS2*	c.[2444C>T];[2444C>T]	p.[Ser815Leu];[Ser815Leu]	5y	No	Healthy	No Tumor	NP	0.18	0.08	0.33	+	50.29	39.64	+
71	18	*PMS2*	c.[862C>T];[862C>T]	p.[Gln288*];[Gln288*]	10y	Yes	Cancer affected	sPNET (7y), Sigmoidal adenocarcinoma (10y)	NP	-0.02	0.10	0.29	+	25.32	19.81	+
58	19 (F2)	*MSH2*	c.[1667T>C];[1667T>C]	p.[Leu556Ser];[Leu556Ser]	13y	Yes	Cancer affected	Carcinoma in situ of the coecum and a poorly cohesive gastric carcinoma (no-signet cells) (9y), Glioblastoma (13y)	NP	0.07	0.06	0.26	+	26.95	21.50	+
87	20 (F2)	*MSH2*	c.[1667T>C];[1667T>C]	p.[Leu556Ser];[Leu556Ser]	17y	Yes	Cancer affected	Glioblastoma (17y)	NP	0.05	0.08	0.23	+	33.15	24.69	+
104	23 (F2)	*MSH2*	c.[1667T>C];[1667T>C]	p.[Leu556Ser];[Leu556Ser]	19y	Yes	Cancer affected	CRC (9y), Glioblastoma (19y)	NP	0.05	0.08	0.26	+	40.11	30.25	+
101	21 (F3)	*MSH6*	c.[3261dup];[3261dup]	p.[Phe1088Leufs*5];[Phe1088Leufs*5]	9y	Yes	Cancer affected	Glioblastoma (9y), Polyposis coli (9y)	NP	-0.04	-0.06	-0.06	-	24.57	18.30	+
107	22 (F3)	*MSH6*	c.[3261dup];[3261dup]	p.[Phe1088Leufs*5];[Phe1088Leufs*5]	13m	No	Healthy	No Tumor	NP	-0.04	-0.06	-0.11	-	17.88	13.22	+
109	22 (F3)				15m	No	Healthy			-0.04	-0.06	-0.11	-	16.29	12.13	+
116	24 (F4)	*MSH2*	c.[958_959dup];[958_959dup]	p.[Thr321Profs*11];[Thr321Profs*11]	2y	Yes	Under chemotherapy	Medulloepithelioma (2y)	NP	0.07	0.04	0.20	+	65.36	50.62	+
132	25 (F4)	*MSH2*	c.[958_959dup];[958_959dup]	p.[Thr321Profs*11];[Thr321Profs*11]	6y	Yes	Healthy	T-cell lymphoma (19 months), Brain tumor (no pathology confirmation) (7y)	NP	0.16	0.15	0.40	+	76.50	59.18	+
Healhy controls:																
7	Control07				NA	No	Healthy		NP	-0.06	-0.05	-0.05	-	0.59	0.45	-
32	Control32				NA	No	Healthy		NP	-0.08	-0.06	-0.08	-	0	0	-
41	Control41				NA	No	Healthy		NP	-0.07	-0.06	-0.09	-	1.20	1.4	-
42	Control42				NA	No	Healthy		NP	-0.08	-0.05	-0.08	-	0.60	0.46	-
50	Control48				NA	No	Healthy		NP	-0.05	-0.04	-0.07	-	0	0	-
55	Control50				NA	No	Healthy		NP	-0.08	-0.07	-0.09	-	1.79	1.4	-
66	Control57				NA	No	Healthy		NP	-0.07	-0.05	-0.07	-	0	0	-
72	Control61				NA	No	Healthy		NP	-0.07	-0.04	-0.09	-	0.58	0.45	-
74	Control63				NA	No	Healthy		NP	-0.08	-0.07	-0.05	-	1.15	0.88	-
79	Control67				NA	No	Healthy		NP	-0.07	-0.07	-0.07	-	0	0	-
80	Control68				NA	No	Healthy		NP	-0.08	-0.07	-0.07	-	1.72	1.28	-
89	Control74				NA	No	Healthy		NP	-0.08	-0.07	-0.07	-	0	0	-
95	Control78				NA	No	Healthy		NP	-0.09	-0.05	-0.06	-	1.12	0.83	-
111	Control81				NA	No	Healthy		NP	-0.08	-0.04	-0.10	-	0	0	-
119	Control85				NA	No	Healthy		NP	-0.08	-0.06	-0.09	-	0.56	0.42	-
123	Control88				NA	No	Healthy		NP	-0.07	-0.04	-0.06	-	0	0	-

126	Control89	NA	No	Healthy	NP	-0.07	-0.05	-0.11	-	1.09	0.81	-
131	Control92	NA	No	Healthy	NP	-0.06	-0.07	-0.08	-	0	0	-

Supplemental Table S3. Clinico-molecular characteristics and summary of the obtained results from patients with a suspected diagnosis of CMMRD, and confirmed diagnosis of Li-Fraumeni, NF1 and PPAP and early-onset LS (age of onset of the first tumor ≤25 years). Abbreviations: NA, not available; NP, not performed; IHC, immunohistochemistry; gMSI, germline microsatellite instability; gMSI detection: -, no detection; +, detection; hs-MSI, high-sensitivity microsatellite instability; hs-MSI detection: -, no detection; +, detection; CRC, colorectal cancer.

Case ID	Clinical Criteria	Germline screening results	Age at sampling	Tumours (age)	Tumor characteristics	**gMSI** (according to Ingham et al., 2013)				**hs-MSI score**		
						D2S123	**D17S250**	**D17S791**	**gMSI detection**	186 MS	231 MS	**hs-MSI detection**
Suspected diagnosis of CMMRD:												
Suspected-01	CALMs	No MMR mutations	9y	Hepatoblastoma (9y)	NA	-0.03	-0.06	-0.05	-	1.65	1.23	-
Suspected-02	Father affected of LS: *MSH2* c.924_925dup, p.(Ala309Glufs*23)	Not tested	14y	Melanoma (14y)	IHC: MSH6 loss	-0.03	-0.01	-0.03	-	1.1	0.83	-
Suspected-03	Mother affected of LS: *MSH6* c.1153_1155del, p.(Arg385del)	Heterozygous *MSH6* c.1153_1155del, p.(Arg385del)		Ependimoma (2y)	IHC: Conserved protein expression	-0.03	-0.02	-0.04	-	1.66	1.65	-
Suspected-04	Carcinoma from the LS spectrum at age <25 years and 1 CALM	No MMR mutations	16y	Medulloblastoma (6y), CRC (16y)	NA	NP	NP	NP	NP	0.55	0.44	-
Li-Fraumeni:												
LiFrau-01		Heterozygous *TP53* c.1010G>A, p.(Arg337His)	30y	No Tumor		NP	NP	NP	NP	0	0	-
LiFrau-02		Heterozygous *TP53* c.742C>T, p.(Arg248Trp)	23y	Leukemia (14y)		NP	NP	NP	NP	0	0	-
LiFrau-03		Heterozygous *TP53* c.724T>G, p.(Cys242Gly)	27y	Breast (27y)		NP	NP	NP	NP	0.55	0.44	-
LiFrau-04		Heterozygous *TP53* c.473G>A; p.(Arg158His)	38y	Lung (38y)		NP	NP	NP	NP	0.55	0.44	-
Neurofibromatosis type 1:												
NF1-01		Heterozygous *NF1* c.2446C>T, p.Arg816*	20y			NP	NP	NP	NP	1.15	0.94	-
NF1-02		Heterozygous *NF1* c.1318C>T, p.Arg440*	22y			NP	NP	NP	NP	0	0	-
NF1-03	>10 CALMs, lisch nodules	Heterozygous *NF1* c.5648dup, p.(Asn1883Lysfs*9)	34y	>50 Neurofibromas of <0.5 cm		NP	NP	NP	NP	0	0	-
Polymerase proofreading-associated polyposis:												
PPAP-01		Heterozygous *POLD1* c.946G>C; p.(Asp316His)	65y	CRC (58y)		NP	NP	NP	NP	0.57	0.47	-
PPAP-02		Heterozygous *POLE* c.833C>A; p.(Thr278Lys)	67y	CRC (54y)		NP	NP	NP	NP	0	0	-
PPAP-03		Heterozygous *POLE* c.833C>A; p.(Thr278Lys)	57y	CRC (52y), CRC (52y), CRC (53y)		NP	NP	NP	NP	1.16	0.94	-
PPAP-04		Heterozygous *POLE* c.1270C>G; p.(Leu424Val)	30y	CC (28y), Brain (31y)		NP	NP	NP	NP	0.58	0.47	-
PPAP-05		Heterozygous *POLD1* c.946G>C; p.(Asp316His)	47y			NP	NP	NP	NP	0	0	-
Early-onset Lynch syndrome:												
early-LS-01		Heterozygous MSH2 c.735_736insTGTT, p.(Lys246fs*)	28y	Thyroid (25y), Ovary (41y)		NP	NP	NP	NP	0.58	0.47	-
early-LS-02		Heterozygous MLH1 c.1459C>T, p.Arg487*	25y	CRC (24y)		NP	NP	NP	NP	0.58	0.47	-
early-LS-03		Heterozygous MLH1 c.1865T>A; p.(Leu622His)	47y	CRC (22y), CRC (43y)		NP	NP	NP	NP	2.31	1.9	-
early-LS-04		Heterozygous MLH1 c.199G>A; p.(Gly67Arg)	25y	CRC (25y)		NP	NP	NP	NP	0.58	0.48	-
early-LS-05		Heterozygous MSH2 c.732del; p.(Leu244Phefs*2)	45y	CRC (24y), CRC (31y)		NP	NP	NP	NP	0.57	0.46	-
early-LS-06		Heterozygous MLH1 c.884+4A>G; p.(His264_Ser295delinsPhefs*2)	26y	Leukemina (3y), brain (28y)		NP	NP	NP	NP	1.18	0.97	-
early-LS-07		Heterozygous MSH2 E1-E2 deletion	41y	Hodgkins Disease (22y), Uterine (38y)		NP	NP	NP	NP	0	0	-
early-LS-08		Heterozygous MSH2 c.2081T>C; p.(Phe694Ser)	25y	CRC (24y)		NP	NP	NP	NP	1.16	0.94	-
early-LS-09		Heterozygous MLH1 c.794G>C, p.(Arg265Pro)	16y	CRC (16y)		NP	NP	NP	NP	0	0	-

Supplemental Table S4. Sensitivity and specificity for the detection of CMMRD by using hs-MSI and gMSI approaches. CMMRD cases from training and validation cohort were included in the analysis. Negative controls from training and validation sets were included in the analysis of the hs-MSI approach; negative controls from the validation cohort were used in the analysis of gMSI approach. Abbreviations: MSI, microsatellite instability; hs-MSI, high-sensitivity microsatellite instability; gMSI, germline microsatellite instability.

MSI analysis method	CMMRD affected	Negative control
hs-MSI panel (186 MS)		
% MSI above threshold	29	0
% MSI below threshold	0	33
Sensitivity *(95% IC)*	100.00% *(88.06-100%)*	
Specificity *(95% IC)*	100.00% *(89.42-100%)*	
gMSI Ingham *et al.*		
ratio above threshold	15	0
ratio below threshold	14	18
Sensitivity *(95% IC)*	*51.72% (32.53-70.55%)*	
Specificity *(95% IC)*	*100.00% (81.47-100%)*	

Supplemental Table S5. Analysis of dinucleotide D2S123 from NGS data. Analysis of D2S123 marker from NGS was carried out in CMMRD samples (training cohort), a subset of LS patient samples and baseline samples. Abbreviations: NA, not applicable; Homo, homozygous allele; Hetero, heterozygous allele; MS, microsatellite; MSI, microsatellite instability. Gray shading indicates detection of MSI above the established threshold (8.23): -, no detection; +, detection.

Sample	Affected Gene	Number of reads by allele lenght													MS zygosity	wildtype allele lenght	wildtype allele reads	sum of non-wildtype allele reads	Ratio non-wildtype /wildtype	% MS instability	MSI detection
		(AC)18	(AC)19	(AC)20	(AC)21	(AC)22	(AC)23	(AC)24	(AC)25	(AC)26	(AC)27	(AC)28	(AC)29	(AC)30							
CMMRD-01	*MSH6*	0	0	12	235	5	0	0	0	0	0	0	0	0	Homo	(AC)21	235	17	0.07	7.23	-
CMMRD-01 BM	*MSH6*	0	0	3	274	4	0	0	0	0	0	0	0	0	Homo	(AC)21	274	7	0.03	2.55	-
CMMRD-02	*MSH6*	0	0	0	38	0	0	0	0	0	0	0	0	0	Homo	(AC)21	38	0	0.00	0.00	-
CMMRD-03	*MLH1*	1	0	8	44	6	0	0	0	0	0	0	0	0	Homo	(AC)21	44	15	0.34	34.09	+
CMMRD-04	*PMS2*	0	1	33	9	1	0	0	0	0	0	0	18	2	Hetero	AC)20 / (AC)2	51	13	0.25	25.49	+
CMMRD-05	*PMS2*	0	0	1	10	2	0	0	0	0	0	0	7	0	Hetero	AC)21 / (AC)2	17	3	0.18	17.65	+
CMMRD-06	*PMS2*	0	0	3	127	23	0	0	0	0	0	0	0	0	Homo	(AC)21	127	26	0.20	20.47	+
CMMRD-07	*MSH6*	0	0	0	0	0	0	0	0	0	1	5	177	1	Homo	(AC)29	177	7	0.04	3.95	-
CMMRD-08	*MSH6*	0	0	7	93	0	0	0	0	0	0	1	83	0	Hetero	AC)21 / (AC)2	176	8	0.05	4.55	-
CMMRD-09	*MSH6*	0	4	134	118	1	0	0	0	0	0	0	0	0	Hetero	AC)20 / (AC)2	252	5	0.02	1.98	-
CMMRD-10	*PMS2*	0	0	5	176	11	2	0	0	0	0	0	0	0	Homo	(AC)21	176	18	0.10	10.23	+
CMMRD-11	*MSH6*	0	0	0	0	0	4	115	4	0	0	0	0	0	Homo	(AC)24	115	8	0.07	6.96	-
Lynch-02	*MLH1*	0	1	6	128	0	0	0	0	0	0	1	70	0	Hetero	AC)22 / (AC)2	86	2	0.02	2.33	-
Lynch-12	*MSH2*	0	1	0	8	0	0	0	0	0	0	0	0	0	Hetero	AC)23 / (AC)2	219	13	0.06	5.94	-
Lynch-33	*MLH1*	0	0	9	234	0	0	0	0	0	0	0	0	0	Hetero	AC)21 / (AC)2	198	8	0.04	4.04	-
Lynch-14	*MSH2*	0	1	1	151	128	0	0	0	0	0	0	0	0	Homo	(AC)21	234	10	0.04	4.27	-
Lynch-01	*MLH1*	0	0	0	0	6	88	0	0	0	0	4	59	0	Homo	(AC)21	212	7	0.03	3.30	-
Lynch-31	*MLH1*	0	0	6	96	0	0	0	0	0	0	0	0	40	Hetero	AC)21 / (AC)2	279	2	0.00	0.72	-
Lynch-44	*MSH6*	0	0	1	6	271	0	0	0	0	0	0	0	0	Hetero	AC)23 / (AC)2	147	10	0.07	6.80	-
Lynch-39	*MSH2*	0	0	0	2	60	3	23	0	0	0	0	0	0	Hetero	AC)21 / (AC)3	136	6	0.04	4.41	-
Lynch-21	*MSH6*	0	0	0	0	0	0	0	0	0	0	0	0	0	Homo	(AC)22	271	7	0.03	2.58	-
Lynch-11	*MSH2*	0	0	0	0	0	0	0	0	0	0	0	0	0	Hetero	AC)22 / (AC)2	83	5	0.06	6.02	-
BL-1	NA	0	0	2	143	0	0	0	0	0	0	1	105	0	Hetero	AC)21 / (AC)2	248	3	0.01	1.21	-
BL-2	NA	0	0	6	162	0	0	0	0	0	0	0	0	0	Homo	(AC)21	162	6	0.04	3.70	-
BL-3	NA	0	0	2	72	0	0	0	0	0	0	1	27	0	Hetero	AC)21 / (AC)2	99	3	0.03	3.03	-
BL-4	NA	0	0	0	0	0	0	1	13	0	0	1	46	0	Hetero	AC)21 / (AC)2	59	2	0.03	3.39	-
BL-5	NA	0	0	7	181	0	0	0	0	0	0	0	0	0	Homo	(AC)21	181	7	0.04	3.87	-
BL-6	NA	0	0	0	2	89	0	0	0	0	0	1	49	0	Hetero	AC)22 / (AC)2	138	3	0.03	2.17	-
BL-7	NA	0	0	0	3	41	0	0	0	0	0	0	1	14	Hetero	AC)22 / (AC)3	55	4	0.07	7.27	-
BL-8	NA	0	0	0	0	0	2	49	0	0	0	0	2	26	Hetero	AC)24 / (AC)3	75	4	0.05	5.33	-
BL-9	NA	0	0	3	159	0	0	0	0	0	0	0	0	0	Homo	(AC)21	159	3	0.02	1.89	-
BL-10	NA	0	0	2	137	0	0	0	0	0	0	0	0	0	Homo	(AC)21	137	2	0.02	1.46	-
BL-11	NA	0	2	97	0	0	0	0	0	0	0	0	79	0	Hetero	AC)20 / (AC)2	176	2	0.01	1.14	-
BL-12	NA	0	0	0	145	0	0	49	0	0	0	0	0	0	Hetero	AC)21 / (AC)2	194	0	0.00	0.00	-
BL-13	NA	0	0	5	103	80	0	0	0	0	0	0	0	0	Hetero	AC)21 / (AC)2	183	5	0.03	2.73	-
BL-14	NA	0	0	0	0	0	0	0	0	0	1	2	184	0	Homo	(AC)29	184	3	0.02	1.63	-
BL-15	NA	0	1	87	0	0	0	0	0	0	0	2	93	0	Hetero	AC)20 / (AC)2	180	3	0.02	1.67	-
BL-16	NA	0	0	0	0	53	16	0	0	0	0	0	0	0	Hetero	AC)22 / (AC)2	69	0	0.00	0.00	-
BL-17	NA	0	0	0	20	0	0	0	0	0	0	0	10	0	Hetero	AC)21 / (AC)2	30	0	0.00	0.00	-
BL-18	NA	0	1	3	89	1	0	0	0	0	0	0	46	0	Hetero	AC)21 / (AC)2	135	5	0.04	3.70	-
BL-19	NA	0	0	0	0	0	0	0	0	0	0	4	130	1	Homo	(AC)29	130	5	0.04	3.85	-
BL-20	NA	0	0	1	0	2	91	1	0	0	0	2	72	0	Hetero	AC)23 / (AC)2	163	6	0.04	3.68	-
BL-21	NA	0	0	7	139	0	0	0	0	0	0	0	0	0	Homo	(AC)21	139	7	0.05	5.04	-
BL-22	NA	0	0	2	79	103	0	0	0	0	0	0	0	0	Hetero	AC)21 / (AC)2	182	2	0.01	1.10	-

Supplemental Table S6. Turn-around time for the hs-MSI approach.

Blood DNA extraction and DNA quality control							
Kit:			**Time required:**				
Wizard® Genomic DNA Purification Kit, Promega			1 day				
Library Preparation							
Enrichment approach:	**Sequenceable Design Size:**	**Read lenght:**	**Input DNA:**	**Fragmentation:**	**Capture method:**	**Targeting method:**	**Library prep. time:**
Custom panel HaloPlex HS Target Enrichment System for Illumina Sequencing, Agilent	0.114 Mbases	2x150 bp	57.6 ng	Enzymatic	Amplicon	DNA molecular inversion probes	3 days
Run							
Illumina Platform:	**Run type:**	**Kit:**			**Throughput (3.46 GB/sample required):**		**Run time:**
HiSeq 2500*	Rapid Run with OBCG	TruSeq Rapid Cluster Kit - PE and SR, TruSeq Rapid SBS Kits - 200 Cycles and 50 Cycles			26 samples/flowcell		40h
Alternative sequencing options:							
MiSeq	v2 Reagents	MiSeq Reagents Kits v2			1 sample/flowcell		24h
	v3 Reagents	MiSeq Reagents Kits v3			2 sample/flowcell		56h
MiniSeq	High Output	MiniSeq Reagent Kit			2 samples/flowcell		24h
NextSeq 550	Mid Output	NextSeq 500 kit			11 samples/flowcell		26h
	High Output	NextSeq 500 kit			35 patients/flowcell		29h
Bioinformatics analysis							
Pipeline:	**Software:**			**Human genome reference:**		**Time**	
Trimming and alignment	SureCall v3.5, Agilent			hg19		1h/sample (1 day/26 samples)	
Indel calling	Custom pipeline			hg19		2h/sample (1 day/26 samples)	

Supplemental Table S7. MMR variants identified in CMMRD patients from training and validation cohorts. Mutations are described in accordance with the Human Genome Variation Society (http://www.hgvs.org/mutnomen) guidelines using the following mRNA sequences as references: MLH1: NM_000249.3, MSH2: NM_000251.2, MSH6: NM_000179.2 and PMS2: NM_000535.5. Abbreviations: NR, not reported. (*) Classification according to the InSIGHT classification guidelines.

Case ID	Previous reports	Family ID	MMR affected gene	MMR variants	Variant (cDNA)	Variant (protein)	Protein domain (for missense variants)	InSIGHT classification	ClinVar classification
CMMRD-01 (DNA91)	Tesch et al, 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)	Family A	*MSH6*	c.[2653A>T];[2653A>T]	c.2653A>T	p.(Lys885*)	_	Class 5*	Class 5
CMMRD-02	Unpubllished								
CMMRD-03 (DNA49)	Tesch et al, 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*MLH1*	c.[332C>T]; [332C>T]	c.332C>T	p.(Ala111Val)	ATPase domain	Class 4	Class 4
CMMRD-04 (DNA51)	Tesch et al, 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*PMS2*	c.[2007-2A>G];[2007-2A>G]	c.2007-2A>G	p.(?)	_	Class 4*	NR
CMMRD-05 / DNA71	Gallon et al., 2019 (PMID: 30740824)		*PMS2*	c.[862C>T];[862C>T]	c.862C>T	p.(Gln288*)	_	Class 5*	NR
CMMRD-06 (DNAA)	Vogt et al, 2016 (PMID: 27329736) Gallon et al., 2019 (PMID: 30740824)		*PMS2*	complex rearrangement	fusion transcript from *PMS2* exon 10 into a sequence derived from intron	p.(?)	_	Class 5*	NR
CMMRD-07 (DNA43)	Tesch et al, 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*MSH6*	c.[467C>G];[1316A>G]	c.467C>G	p.(Ser156*)	_	Class 5	Class 5
					c.1316A>G	p.(Asp439Gly)	Connector and MSH2 interaction domains	Class 3	Class 3
CMMRD-08 (DNA83)	Tesch et al, 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*MSH6*	c.[1135_1139del];[2277_2293del]	c.1135_1139del	p.(Arg379*)	_	Class 5	Class 5
					c.2277_2293del	p.(Glu760Profs*6)	_	Class 5*	NR
CMMRD-09 (DNAB)	Tesch et al, 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*MSH6*	c.[2238dup];[2980T>A]	c.2238dup	p.(Leu747Serfs*9)	_	Class 5*	Class 5
					c.2980T>A	p.(Tyr994Asn)	Lever domain	Class 3*	NR
CMMRD-10	Unpublished		*PMS2*	c.[24-2A>G];[24-2A>G]	c.24-2A>G	p.Ser8Argfs*4	_	Class 4*	NR
CMMRD-11	Unpublished		*MSH6*	c.[742C>T];[742C>T]	c.742C>T	p.(Arg248*)	_	Class 5	Class 5
DNAC	Tesch et al., 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*PMS2*	c.[1515del];[1515del]	c.1515del	p.(Phe506Serfs*89)	_	Class 5*	NR
DNAD	Gallon et al., 2019 (PMID: 30740824)		*MSH6*	c.[3838C>T];[3838C>T]	c.3838C>T	p.(Gln1280*)	_	Class 5	Class 5
DNAE	Tesch et al., 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*MLH1*	c.[62C>A];[2146G>A]	c.62C>A	p.(Ala21Glu)	ATPase and MutSα interaction domains	Class 5	Class 5
					c.2146G>A	p.(Val716Met)	PMS2 interaction domain	Class 1	Class 1
DNA65	Gallon et al., 2019 (PMID: 30740824)		*MSH6*	c.[691del];[691del]	c.691del	p.(Val231Tyrfs*15)	_	Class 5*	NR
DNA93	Tesch et al., 2018 (PMID: 30013564) Gallon et al., 2019 (PMID: 30740824)		*PMS2*	c.[1145-31_1145-13del];[1145-31_1145-13del]	c.1145-31_1145-13del	p.(Asn383*; Gly382Valfs*19)	_	Class 5*	NR
DNA99/DNA105	Gallon et al., 2019 (PMID: 30740824)	Family 1	*MSH6*	c.[3557-1G>C];[(457+1_458-1)_(3646+1_3646-1)del]	c.3557-1G>C	p.(Glu1187_Gly1216del)	_	Class 4*	Class 4
DNA82	Gallon et al., 2019 (PMID: 30740824)								
DNA56	Suerink et al., 2018 (PMID: 28503822) Tesch et al., 2018 (PMID: 30013564)		*PMS2*	c.[2444C>T];[2444C>T]	c.2444C>T	p.(Ser815Leu)	MLH1 interaction domain	Class 4	Class 3/Class 4
DNA58	Gallon et al., 2019 (PMID: 30740824)	Family 2	*MSH2*	c.[1667T>C];[1667T>C]	c.1667T>C	p.(Leu556Ser)	Lever, MSH6 interaction and EXO1 stabilisation domains	Class 3	Class 3
DNA87	Gallon et al., 2019 (PMID: 30740824)								
DNA104	Gallon et al., 2019 (PMID: 30740824)								
DNA101	Gallon et al., 2019 (PMID: 30740824)	Family 3	*MSH6*	c.[3261dup];[3261dup]	c.3261dup	p.(Phe1088Leufs*5)	_	Class 5	Class 5
DNA107/DNA109	Gallon et al., 2019 (PMID: 30740824)								
DNA116	Gallon et al., 2019 (PMID: 30740824)	Family 4	*MSH2*	c.[958_959dup];[958_959dup]	c.958_959dup	p.(Thr321Profs*11)	_	Class 5*	NR
DNA132	Gallon et al., 2019 (PMID: 30740824)								

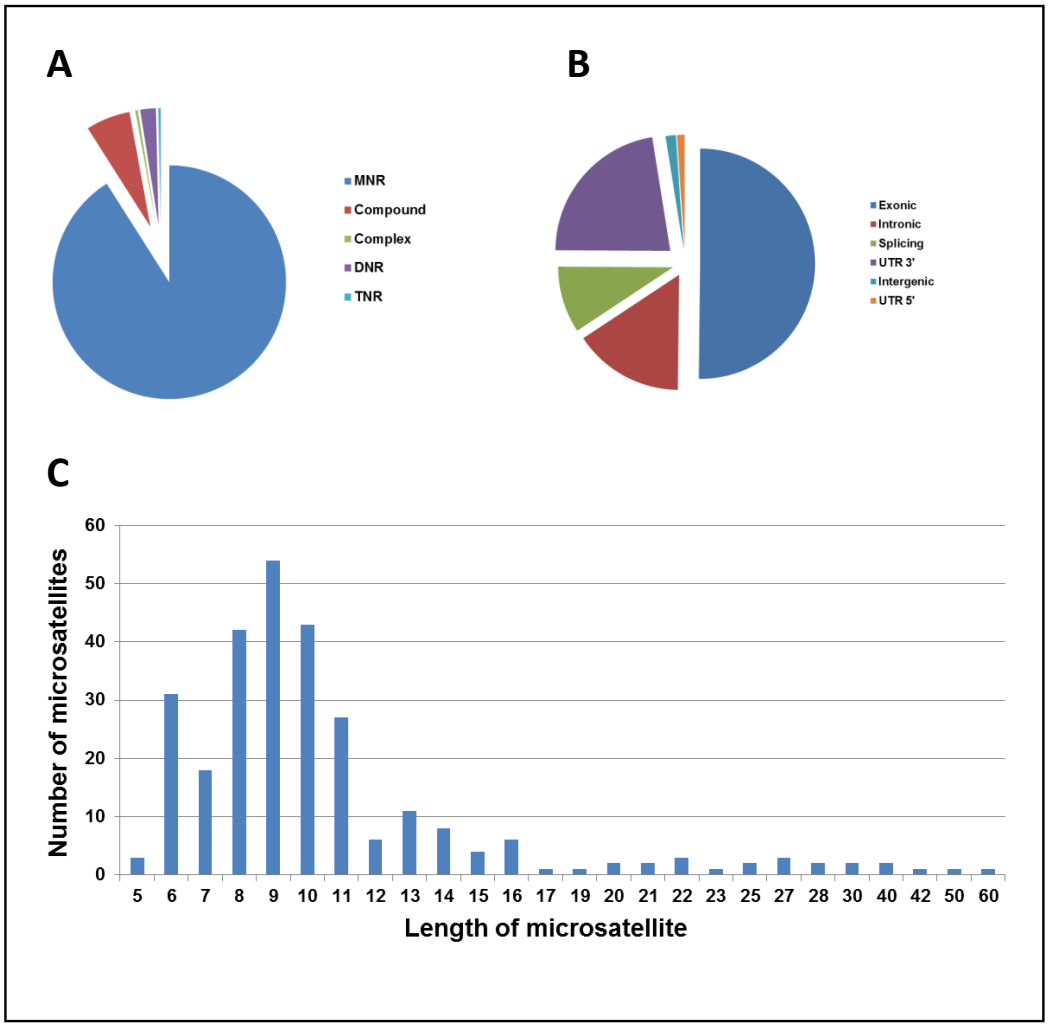
Supplemental Table S8. Example of true SNVs identified in serial dilutions of HCT15 and SW480 DNA. (1) Alternate allele frequency. (2) Expected allele frequency taking into account the dilution factor and the 100% frequency.

Chrom	Position	Ref	Alt	Gene	% HCT15	Alt Freq (1)	Expect Freq (2)	Read Depth	Nº Alt reads
chr4	55593568	C	A	KIT	100	49.62%	49.62%	1316	653
					50	28.13%	24.81%	2293	645
					10	4.93%	4.96%	1542	76
					2	0.85%	0.99%	470	4
					0.4	0.52%	0.20%	572	3
					0	0.00%	0.00%	213	0
chr7	128851593	A	G	SMO	100	45.90%	45.90%	2209	1014
					50	26.15%	22.95%	2122	555
					10	4.20%	4.59%	1356	57
					2	0.79%	0.92%	2025	16
					0.4	0.42%	0.18%	956	4
					0	0.00%	0.00%	1462	0

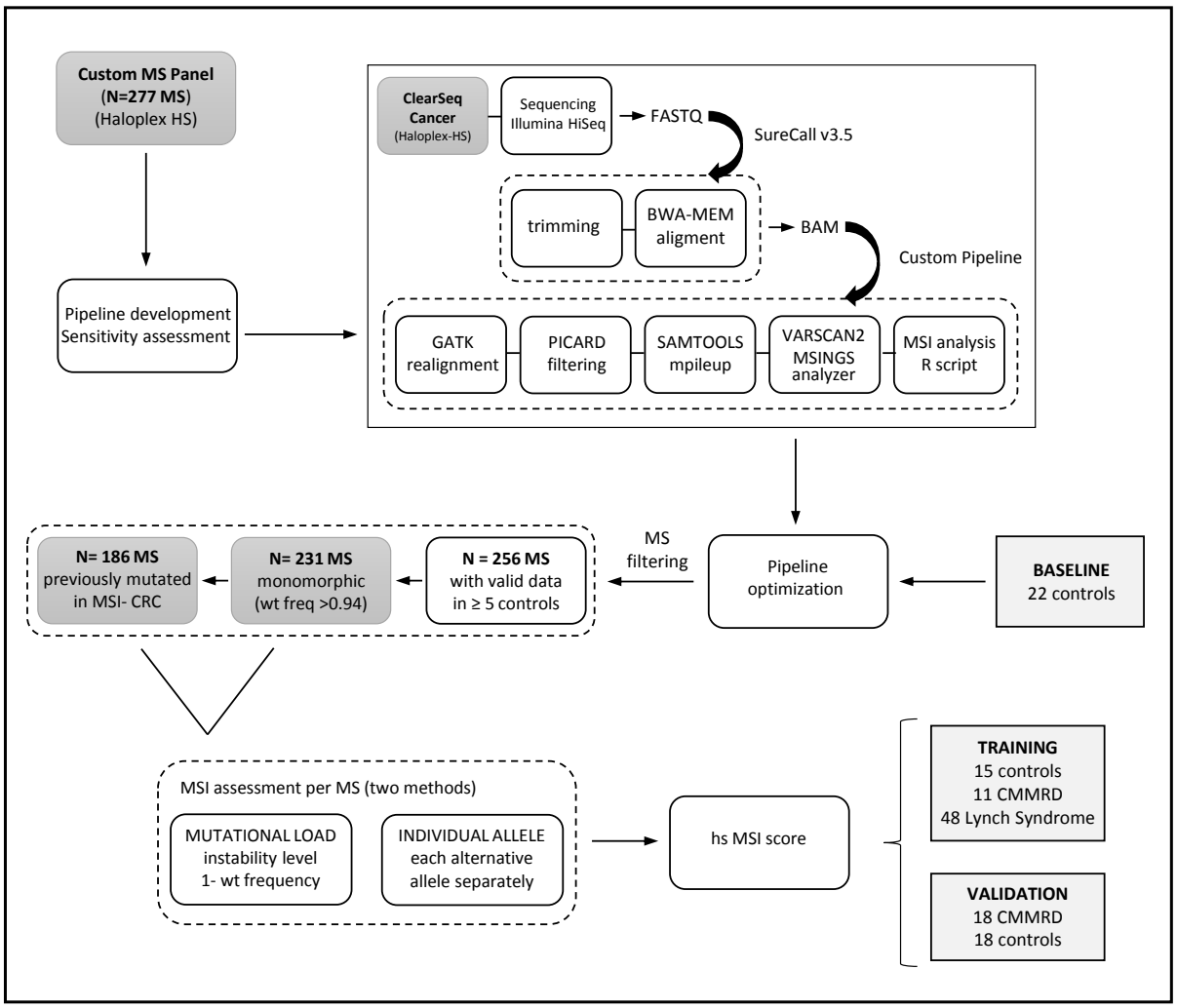
Supplemental Table S9. Allele frequencies of SNVs in 100% HCT15 and their limit of detection. All SNV listed above were identified in both independent replicas of serial dilutions of HCT15 cell line DNA. (1) Highest frequency of the alternative allele that was detected in 100% HCT15. (2) Lowest frequency detected for the alternative allele above 0.4%.

Chrom	Position	Ref	Alt	Gene	HGVS (Coding)	HGVS (Protein)	Alt Freq (1)	Limit Freq (2)	Read Depth	Nº Alt reads	Next Freq	Read Depth	Nº Alt reads
chr9	139399401	G	A	NOTCH1	NM_017617.4:c.4742C>T	NP_060087.3:p.Pro1581Leu	48.93%	**0.50%**	**1803**	**9**	0.26%	762	2
chr17	7577559	G	A	TP53	NM_001126115.1:c.326C>T	NP_001119587.1:p.Ser109Phe	49.85%	**0.94%**	**1064**	**10**	0.00%	1228	0
chr4	55593568	C	A	KIT	NM_000222.2:c.1648-14C>A	NA	49.62%	**0.52%**	**572**	**3**	0.00%	1220	0
chr12	25398281	C	T	KRAS	NM_004985.4:c.38G>A	NP_004976.2:p.Gly13Asp	49.55%	**0.87%**	**805**	**7**	0.00%	737	0
chr7	128851593	A	G	SMO	NM_005631.4:c.1918A>G	NP_005622.1:p.Thr640Ala	45.90%	**0.42%**	**956**	**4**	0.00%	1462	0
chr11	32449639	G	A	WT1	NM_000378.4:c.770-35C>T	NA	48.70%	**0.75%**	**532**	**4**	0.00%	831	0
chr5	67588105	C	T	PIK3R1	NM_181524.1:c.35C>T	NP_852665.1:p.Pro12Leu	45.23%	**0.91%**	**440**	**4**	0.19%	535	1
chr3	178936091	G	A	PIK3CA	NM_006218.3:c.1633G>A	NP_006209.2:p.Glu545Lys	33.24%	**0.51%**	**971**	**5**	0.00%	538	0
chr21	36206721	T	C	RUNX1	NM_001001890.2:c.710A>G	NP_001001890.1:p.Gln237Arg	4.70%	**0.51%**	**1958**	**10**	0.00%	419	0
chr5	67591208	G	T	PIK3R1	NM_001242466.1:c.657-40G>T	NA	0.67%	**0.43%**	**702**	**3**	0.00%	486	0
chr9	133748335	G	A	ABL1	NM_005157.5:c.996G>A	NP_005148.2:p.Arg332=	1.29%	**0.48%**	**2917**	**14**	0.10%	2015	2
chr2	212812155	C	T	ERBB4	NM_005235.2:c.421C>T	NP_005226.1:p.Glu141Lys	1.49%	**0.74%**	**2347**	**17**	0.15%	1902	7

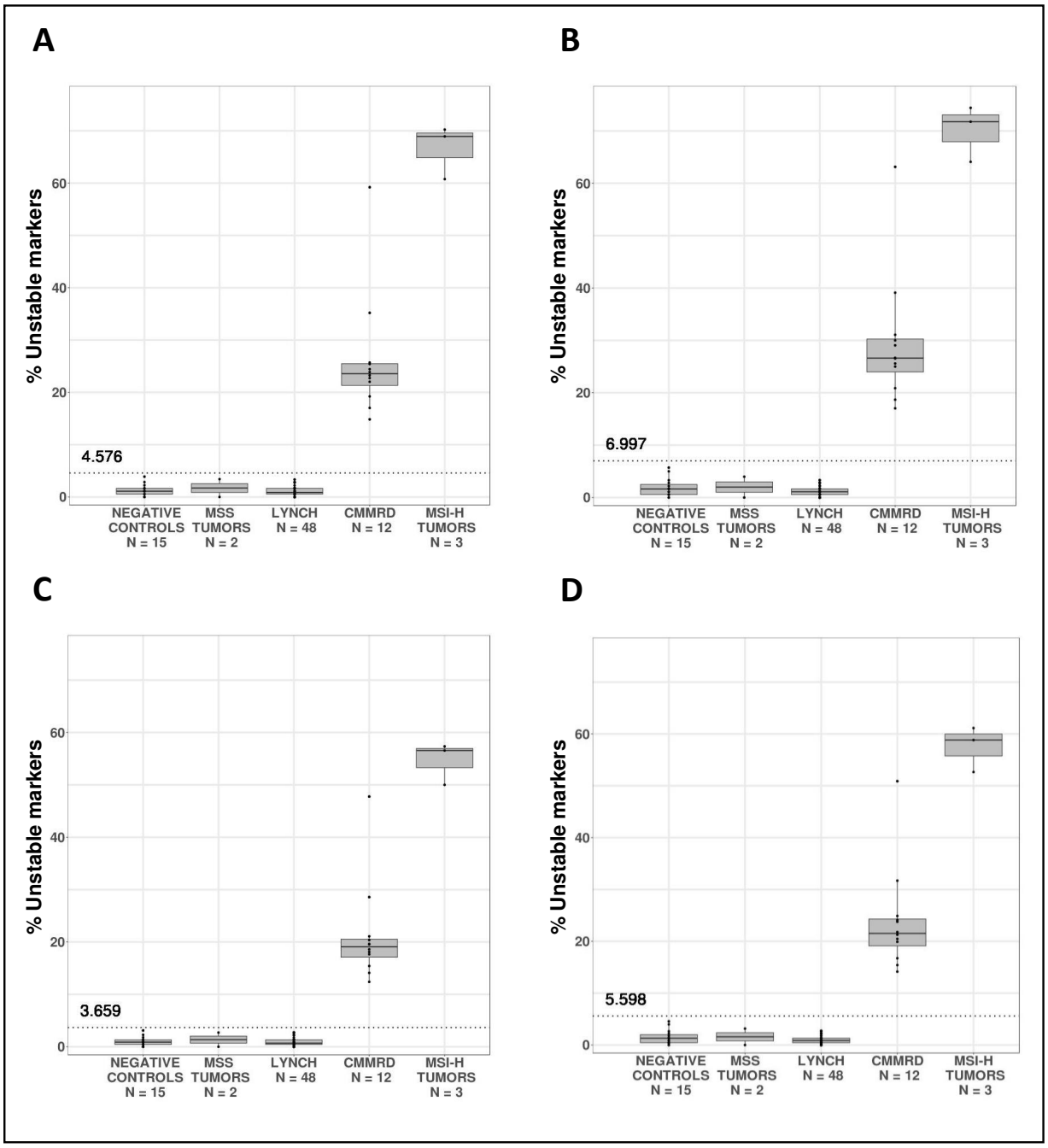
Supplemental Figure S1. Characterization of the 277 microsatellites included in the hs-MSI panel. A) Distribution of microsatellite markers by type. MNR, mononucleotide repeat; DNR, dinucleotide repeat; TNR, trinucleotide repeat. **B)** Localization of the included microsatellite markers. **C)** Length of the included microsatellite markers.



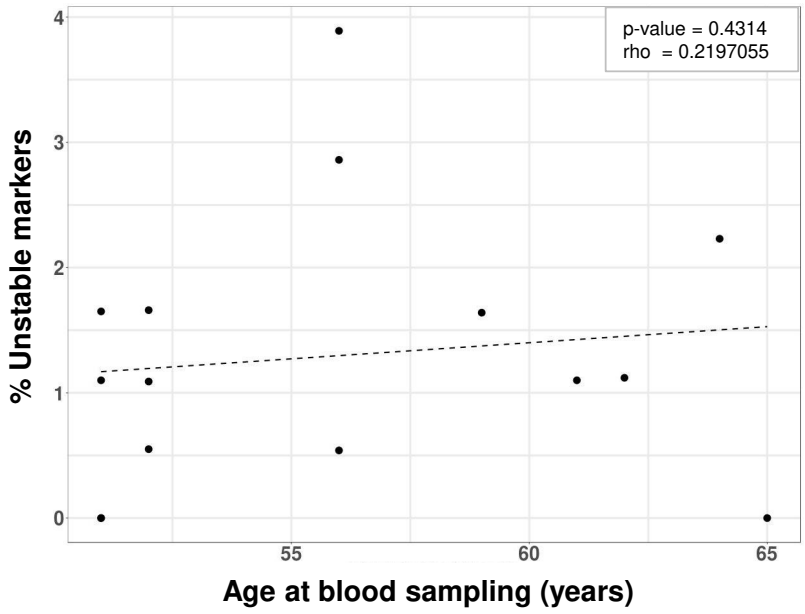
Supplemental Figure S2. Flowchart of analysis procedure.



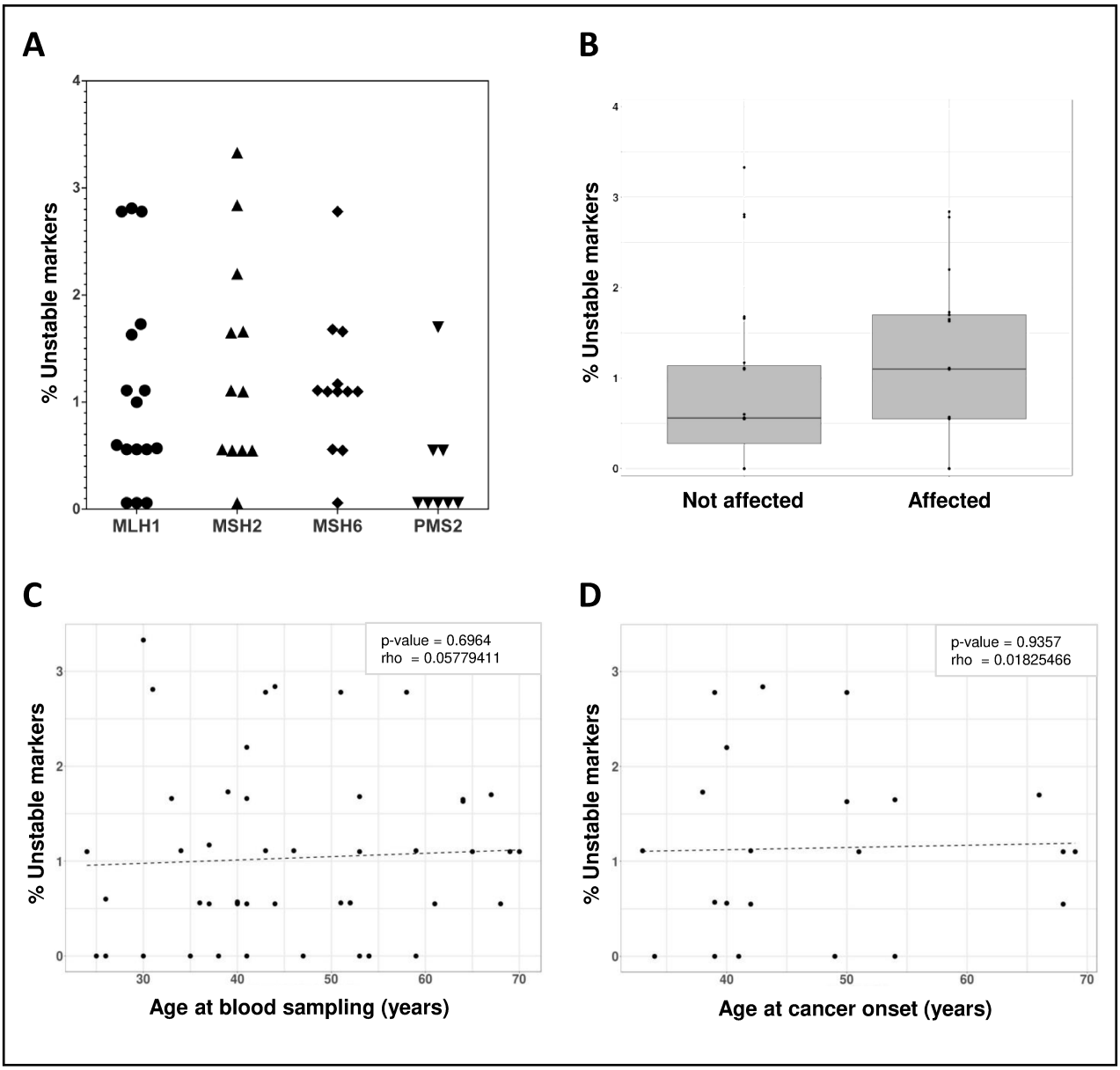
Supplemental Figure S3. MSI score in tumor DNA samples and blood DNA samples included in the training cohort. A) The set of monomorphic markers selected as frequently mutated in MSI-H tumors (n=186) analyzed by using the mutational load analysis method. **B)** The set of monomorphic markers selected as frequently mutated in MSI-H tumors (n=186) analyzed by using the single allele analysis method. **C)** The whole set of monomorphic markers (n=231) analyzed by using the mutational load analysis method. **D)** The whole set of monomorphic markers (n=231) analyzed by using the single allele analysis method. Dashed gray line indicates the threshold for hs-MSI detection in blood samples.



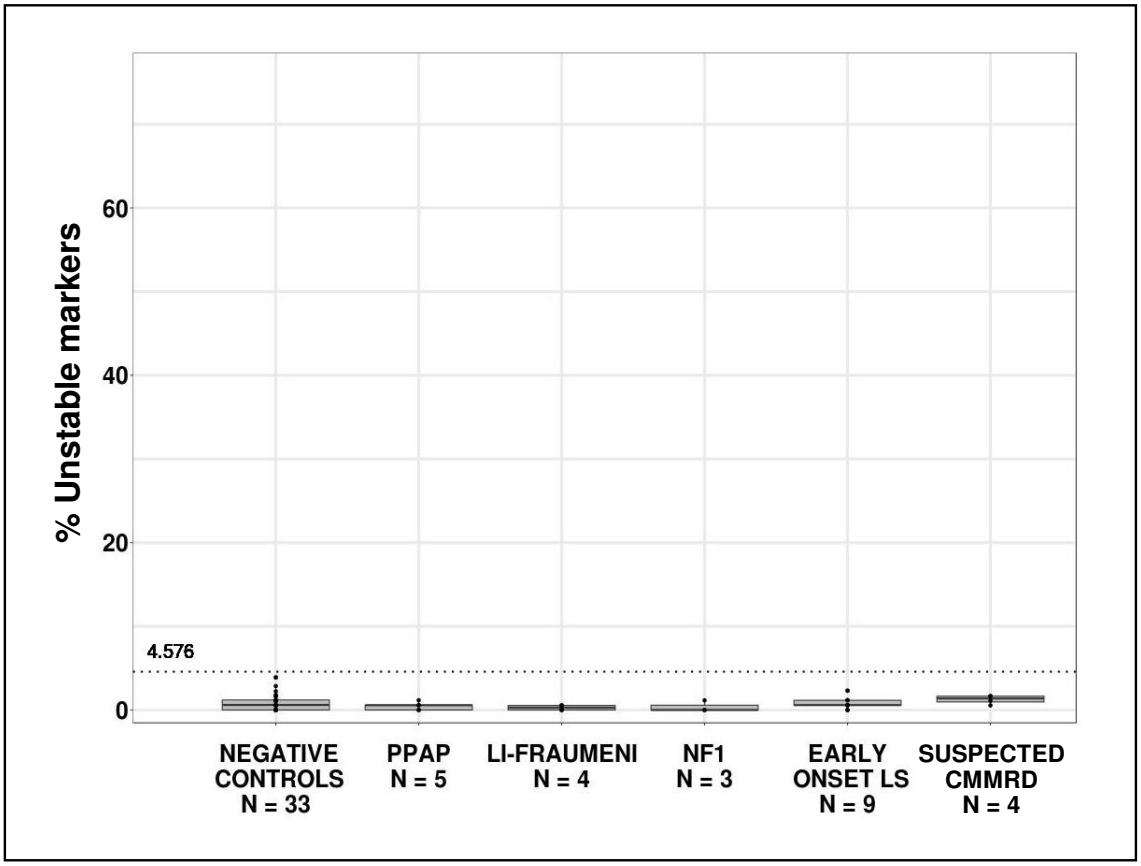
Supplemental Figure S4. MSI score in blood samples from healthy controls plotted against age at blood sampling. No correlation was observed (dashed line).



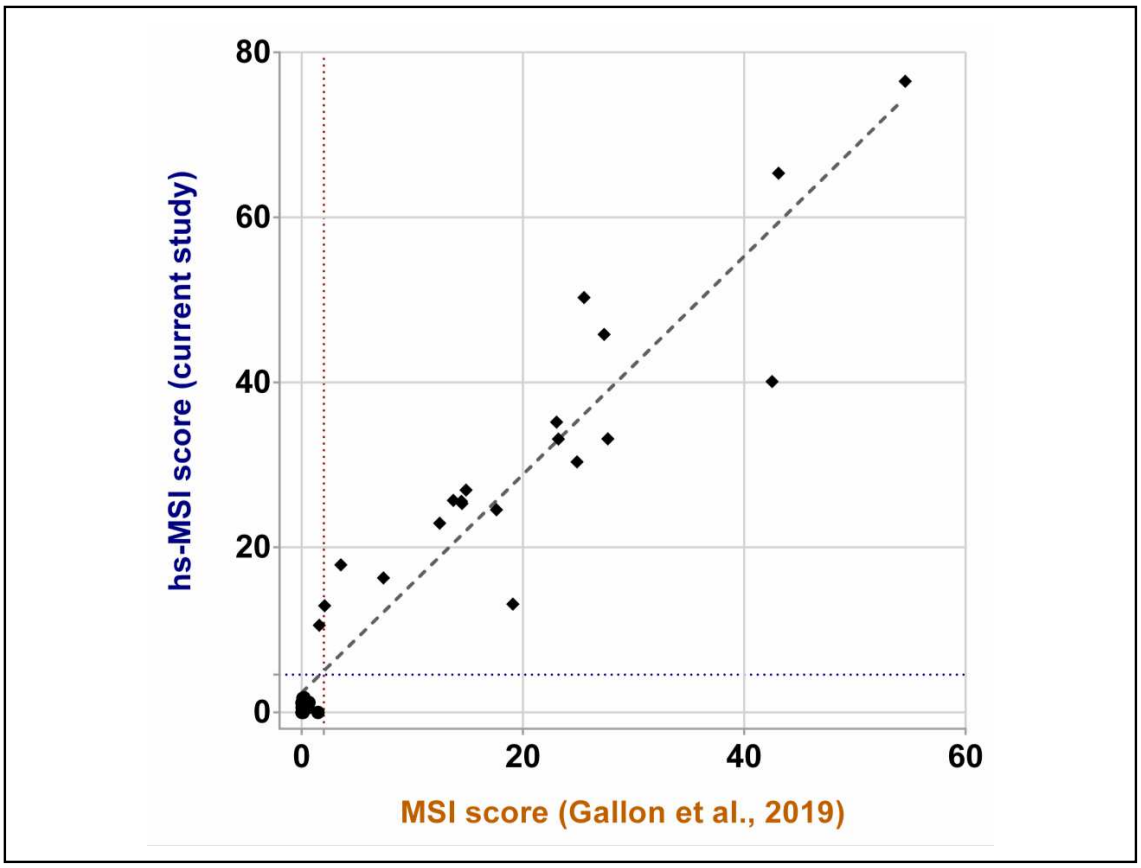
Supplemental Figure S5. Characterization of the MSI levels in blood samples from Lynch syndrome individuals. A) MSI score plotted against the germline mutated MMR gene. **B)** MSI score plotted against cancer diagnosis before blood sampling. Lynch patients are divided into individuals not affected with cancer or cancer affected. **C)** MSI score plotted against individual age at blood sampling. No correlation was observed (dashed line). **D)** MSI score plotted against individual age of cancer onset. No correlation was observed (dashed line).



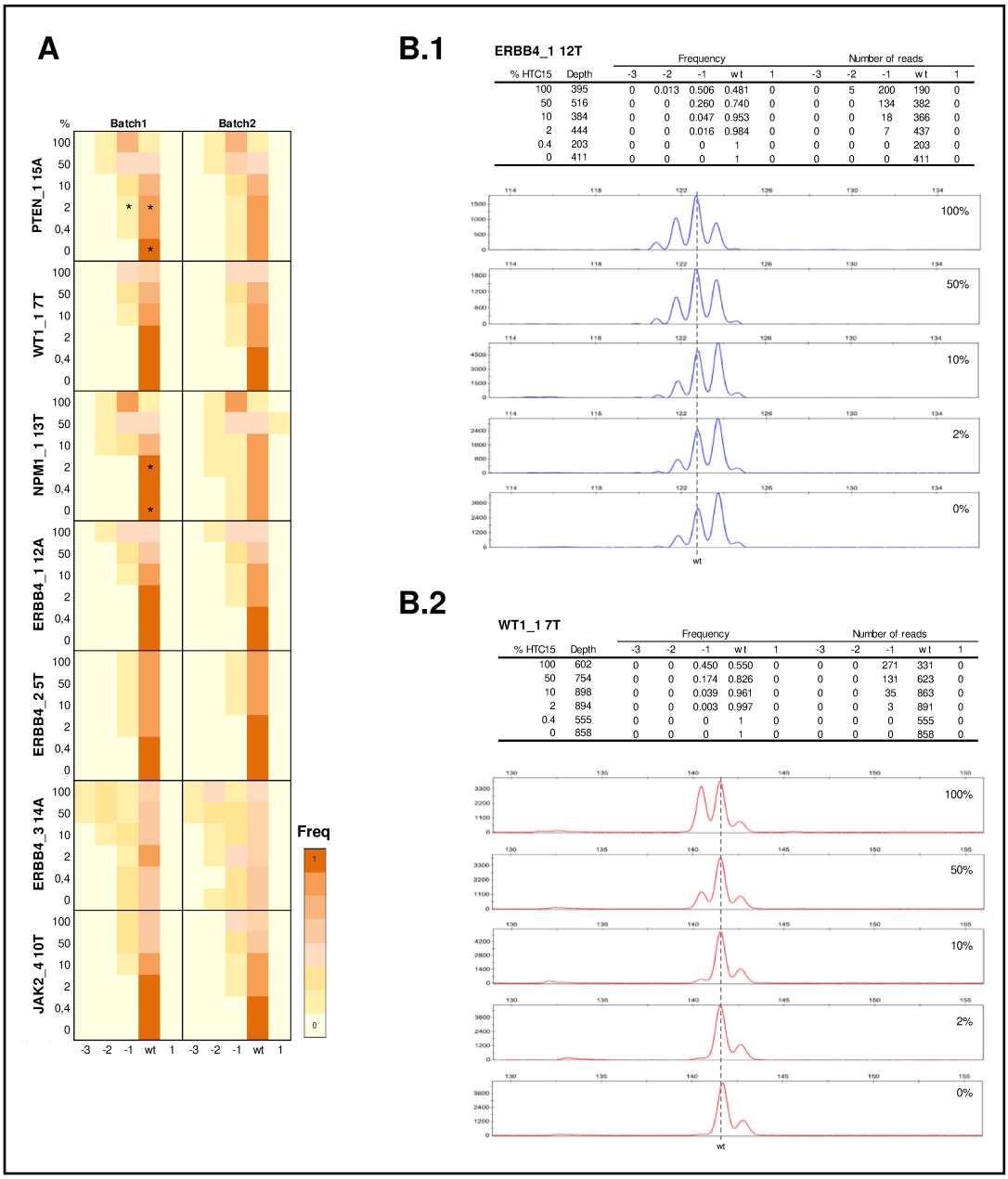
Supplemental Figure S6. MSI score in blood DNA samples from polymerase proofreading-associated polyposis (PPAP), Li-Fraumeni and neurofibromatosis type 1 (NF1) patients, early-onset Lynch syndrome (LS) patients and cases with a suspected diagnosis of CMMRD. Dashed line indicates the threshold for hs-MSI detection.



Supplemental Figure S7. Correlation of MSI scores between Gallon et al., 2019 and the current study in shared samples. Diamonds represent samples from CMMRD patients; circles, negative controls. Dashed blue and orange lines indicate the thresholds for MSI detection stablished in each study. Correlation between MSI scores was observed ($p<0.0001$, $R^2=0.91$) (dashed grey line).



Supplemental Figure S8. Unstable or polymorphic microsatellite markers obtained from sequencing the ClearSeq Cancer HS target regions in serial dilutions of HCT15 and SW480 DNA. A) Allele frequencies of the identified seven unstable or polymorphic microsatellites loci: PTEN (chr10:89720634) 15A, WT1 (chr11:32413476) 7T, NPM1 (chr5:170837514) 13T, ERBB4 (chr2:212289143) 12A, ERBB4 (chr2:212530143) 5T, ERBB4 (chr2:212578380) 14A, JAK2 (chr9:5073682) 10T. DNA mixtures refer to serial dilutions (100%, 50%, 10%, 2%, 0.4%, 0%) of DNA extracted from HCT15 and SW480 cell lines. Two independent replicas were performed (batch 1 and batch 2). Each color represents a frequency range from 0 (light yellow) to 1 (orange). The symbol (*) represents points of the series that did not reach a minimum coverage of 30. **B)** Results obtained in the analysis of markers ERBB4_1 12A (chr2:212289143) **(B.1)**, in blue, and WT1_1 7A (chr11:32413476) **(B.2)**, in red. Tables show frequencies and number of reads obtained for each of the microsatellite alleles in all points of the series. Alleles were validated by DNA fragment analysis and electropherograms are shown. Stutter peaks observed in electropherograms are not present the analysis by NGS, which demonstrated higher sensitivity.



ERBB4_1 12T

% HTC15	Depth	Frequency -3	-2	-1	wt	1	Number of reads -3	-2	-1	wt	1
100	395	0	0.013	0.506	0.481	0	0	5	200	190	0
50	516	0	0	0.260	0.740	0	0	0	134	382	0
10	384	0	0	0.047	0.953	0	0	0	18	366	0
2	444	0	0	0.016	0.984	0	0	0	7	437	0
0.4	203	0	0	0	1	0	0	0	0	203	0
0	411	0	0	0	1	0	0	0	0	411	0

WT1_1 7T

% HTC15	Depth	Frequency -3	-2	-1	wt	1	Number of reads -3	-2	-1	wt	1
100	602	0	0	0.450	0.550	0	0	0	271	331	0
50	754	0	0	0.174	0.826	0	0	0	131	623	0
10	898	0	0	0.039	0.961	0	0	0	35	863	0
2	894	0	0	0.003	0.997	0	0	0	3	891	0
0.4	555	0	0	0	1	0	0	0	0	555	0
0	858	0	0	0	1	0	0	0	0	858	0

DISCUSIÓN

1. Evaluación de la patogenicidad de las variantes MMR

1.1 Reclasificación de variantes en genes reparadores mediante su caracterización exhaustiva

En esta tesis doctoral se ha evaluado la patogenicidad de un total de 19 variantes en genes reparadores de 28 familias con sospecha de SL: 2 en *MLH1*, 2 en *PMS2*, 12 en *MSH2* y 3 en *MSH6* (**Tabla 11**) (González-Acosta *et al.,* 2017 – Artículo 1 y publicaciones adicionales en Anexo 1: Pineda, González-Acosta *et al.,* 2014; Vargas-Parra *et al.,* 2017; Dámaso *et al.,* en preparación).

1.1.1 Estrategias de evaluación de la patogenicidad de las variantes

Como se ha mencionado en la Introducción, para determinar la patogenicidad de una variante es necesario integrar múltiples líneas de evidencias, tanto cualitativas como cuantitativas. Con el objetivo de estudiar de forma exhaustiva las 19 variantes identificadas en genes MMR, en esta tesis doctoral se ha aplicado una combinación de estudios funcionales a nivel de RNA y proteína, predicciones *in silico* y cálculo multifactorial de probabilidad.

Estudios a nivel de RNA

Una importante proporción de las variantes deletéreas identificadas en los genes MMR son variantes que afectan la expresión del RNA mensajero. Esto puede ser debido a la alteración de su expresión debido a variantes en el promotor, o bien afectando la estabilidad del tránscrito o su maduración, que conducirá a la alteración del patrón normal de splicing y generará, en consecuencia, tránscritos aberrantes no funcionales (Baralle et al., 2009; Hinrichsen et al., 2015; Thompson et al., 2015; Thompson et al., 2014).

Tabla 11. Análisis funcional de las variantes estudiadas en los genes MMR. En rojo se indican las mutaciones que se han reclasificado como Patogénica (Clase 5), en verde las Neutrales (Clase 1) y en gris las VUS (Clase 3). NR, no reportado; ND, no disponible; VUS, *variant of unknown significance*; pPAT, probablemente patogénica; PPP, probabilidad posterior de patogenicidad.

Gene	Variante	Frecuencia en controles (%) (ESP-EA/ExAC/ 1000G)	Clasificación inicial	Estudio	Análisis mRNA		Análisis proteína			Cálculo multifactorial (PPP)	Clasificación final	Justificación
					Predicción *in silico*	Procesamiento y estabilidad del cDNA	Predicción *in silico*	Actividad MMR (% del WT)	Expresión MLH1 (% del WT)			
MLH1	**c.121G>C; p.Asp41His**	NR	VUS (Clase 3)	Pineda, González-Acosta et al., 2014	Sin efecto	Sin efecto	Deletérea	22,75 ± 5,77	50,39 ± 9,41	0,9999	**Patogénica (Clase 5)**	**PPP > 0,99 y actividad MMR defectuosa**
MLH1	**c.2128A>G; p.(Asn710Asp)**	NR	VUS (Clase 3)	Pineda, González-Acosta et al., 2014	ESE inconcluso	ND	Neutral	ND	ND	1,92E-06	**Neutral (Clase 1)**	**PPP < 0,001**
PMS2	**c.2149G>A; p.Val717Met**	0,02/0,09/ 0,02	VUS (Clase 3)	González-Acosta et al., 2017	Sin efecto	Sin efecto	Inconcluso	94,86 ± 19,74	68,83 ± 11,63	0,7965	**VUS (Clase 3)**	**PPP 0,05-0,949**
PMS2	**c.2444C>T; p.Ser815Leu**	NR/0,0008/ NR	VUS (Clase 3)	González-Acosta et al., 2017	Inconcluso	Sin efecto	Deletérea	10,11 ± 7,14	18,53 ± 9,25	0,993	**Patogénica (Clase 5)**	**PPP > 0,99 y actividad MMR y expresión defectuosa**
MSH2	**c.211G>C; p.Tyr66Serfs*10**	NR	pPAT (Clase 4)	Vargas-Parra et al., 2017	Inconcluso	Tránscrito aberrante: r.195_211del	ND	ND	ND	0,6575	**Patogénica (Clase 5)**	**RNA aberrante**
MSH2	**c.518T>G; p.Leu173Arg**	NR	VUS (Clase 3)	Vargas-Parra et al., 2017	ESE aberrante	Sin efecto	Deletérea	ND	ND	0,992	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH2	**c.989T>C; p.Leu330Pro**	NR	pPAT (Clase 4)	Vargas-Parra et al., 2017	ESE inconcluso	Sin efecto	Deletérea	ND	ND	0,9957	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH2	**c.1276G>A; p.Ile411_Gly426del**	NR	pPAT (Clase 4)	Vargas-Parra et al., 2017	SS inconcluso	Tránscrito aberrante: r.1230_1277del	Deletérea	ND	ND	0,6649	**Patogénica (Clase 5)**	**RNA aberrante**

Tabla 11. Continuación.

MSH2	**c.1511-1G>A; p.(?)**	NR	pPAT (Clase 4)	Vargas-Parra et al., 2017	SS aberrante	ND	ND	ND	ND	0,9959	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH2	**c.1787A>G; p.Asn596Ser**	0,0002/NR/ NR	VUS (Clase 3)	Dámaso et al., en preparación	Sin efecto	Sin efecto	Neutral	ND	ND	ND	**VUS (Clase 3)**	**Datos insuficientes**
MSH2	**c.2069A>G; p.Gln690Arg**	NR	VUS (Clase 3)	Vargas-Parra et al., 2017	ESE inconcluso	Sin efecto	Deletérea	ND	ND	0,9917	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH2	**c.2074G>C; p.(Gly692Arg)**	NR	pPAT (Clase 4)	Vargas-Parra et al., 2017	Sin efecto	ND	Deletérea	ND	ND	0.9967	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH2	**c.[2635-3C>T; 2635-5T>C]; p.(?)**	NR	pPAT (Clase 4)	Vargas-Parra et al., 2017	SS aberrante	ND	ND	ND	ND	0,9998	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH2	**Duplicación E8; p.Val463Glufs*11**	NR	VUS (Clase 3)	Dámaso et al., en preparación	ND	Tránscrito aberrante: r.1277_1387dup	ND	ND	ND	ND	**Patogénica (Clase 5)**	**RNA aberrante**
MSH2	**Duplicación E11; p.Gly587Alafs*3**	NR	VUS (Clase 3)	Vargas-Parra et al., 2017	ND	Tránscrito aberrante: r.1662_1759dup	ND	ND	ND	0,6568	**Patogénica (Clase 5)**	**RNA aberrante**
MSH2	**Duplicación E11-16; p.(?)**	NR	VUS (Clase 3)	Vargas-Parra et al., 2017	ND	Tránscrito aberrante: r.1662_*23dup	ND	ND	ND	0,9994	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH6	**c.1153_1155del; p.Arg385del**	NR	VUS (Clase 3)	Dámaso et al., en preparación	Sin efecto	Sin efecto	Deletérea	ND	ND	0,9906	**Patogénica (Clase 5)**	**PPP > 0,99**
MSH6	**c.1618_1620del; p.Leu540del, p.Ser536_Asp1058 delinsAsn**	NR	VUS (Clase 3)	Dámaso et al., en preparación	Sin efecto	Tránscrito WT y aberrante: r.1618_1620del r.1607_3172del. Desequilibrio: 0,69 ±0,03 (P) / 0,65 ±0,06 (NP)	Deletérea	ND	ND	0,9914	**Patogénica (Clase 5)**	**PPP > 0,99 y RNA aberrante**
MSH6	**c.3150_3161dup; p.Val1051_Ile1054 dup**	NR	VUS (Clase 3)	Dámaso et al., en preparación	Sin efecto	Sin efecto	Deletérea	ND	ND	0,9999	**Patogénica (Clase 5)**	**PPP > 0,99**

Aunque existen multitud de ensayos dirigidos al estudio del impacto de una variante en RNA, la mayoría de laboratorios siguen la estrategia convencional de RT-PCR (del inglés "*Reverse Transcription PCR*") y secuenciación (Spurdle et al., 2008; Thompson et al., 2014). Esta estrategia permite estudiar un amplio número de variantes de *splicing* gracias a la secuenciación del producto de PCR y, si se combina otras técnicas, también permite determinar si existe expresión alélica diferencial o ASE del tránscrito con la variante (Thompson et al., 2014; Tricarico et al., 2017). Alternativamente, las variantes de *splicing* también se pueden estudiar mediante ensayos *in vitro* mediante *minigenes*. Estos ensayos consisten en clonar dentro de un vector el DNA genómico del paciente, o un DNA control con la variante de interés creada mediante mutagénesis dirigida, y después transfectarlo de forma transitoria en una línea celular para posteriormente extraer el RNA y analizarlo mediante RT-PCR y secuenciación (Gaildrat et al., 2010; Tournier et al., 2008). Los ensayos con *minigenes* son especialmente útiles cuando no se dispone de RNA del paciente y han demostrado tener una buena correlación con los resultados obtenidos a partir RT-PCRs que sí utilizan RNA del paciente (van der Klift et al., 2015).

En esta tesis doctoral hemos seguido una estrategia de RT-PCR utilizando RNA del paciente para estudiar el efecto de las variantes sobre el *splicing* y la expresión del tránscrito. Sin embargo, existen multitud de variaciones técnicas para realizar estos ensayos. Por un lado, se puede analizar mediante RT-PCR el RNA extraído de sangre fresca del paciente, cultivar previamente los linfocitos durante 3-5 días (como en nuestro caso), o cultivar durante periodos largos líneas linfoblastoides establecidas a partir de linfocitos (Houdayer et al., 2012; Thompson et al., 2015; Whiley et al., 2014). Cultivar los linfocitos permite obtener una mayor cantidad de RNA y ofrece la posibilidad de inhibir el sistema NMD. Nosotros recomendamos su utilización siempre que sea posible para incrementar la sensibilidad de la detección de tránscritos aberrantes, sobre todo si la visualización de los productos de PCR se hace a través de geles de agarosa, mucho menos sensibles que la electroforesis capilar, por ejemplo (Borras et al., 2012; Borras et al., 2013; Whiley et al., 2014; Wimmer & Wernstedt, 2014).

Otra fuente de variaciones en el protocolo de estudio del *splicing* de una variante es el diseño de la RT-PCR. Como recomiendan varios estudios, nosotros hemos utilizado la *SuperScript II reverse transcriptase* para la generación del cDNA, ya que presenta una buena eficiencia para la copia de tránscritos largos, y hemos incluído el análisis en paralelo de 3 a 5 muestras de individuos control para poder distinguir los tránscritos fruto del splicing alternativo, y que se dan de forma natural en las células (Thompson et al., 2015; Whiley et al., 2014). Por otro lado, conocer la secuencia del gen que se va a analizar, los tránscritos alternativos, así como la posible presencia de pseudogenes, es esencial para diseñar correctamente los *primers* de amplificación, ya que un mal diseño puede llevar a la amplificación sesgada de ciertos tránscritos y falsear los resultados. Una de las principales limitaciones de las RT-PCRs convencionales es que analizan los tránscritos de RNA mediante amplicones cortos, por lo que eventos que afectan a exones diferentes del que contiene la variante son difíciles de observar. Las estrategias de *long-range PCR*, que se basan en la obtención de amplicones largos que cubren varios exones del gen, resuelven este posible problema (de Jong et al., 2017; Duraturo et al., 2013; Guarinos et al., 2010; Monika Morak et al., 2019; Plaschke et al., 2003; van der Klift et al., 2015). En el caso de *PMS2*, es recomendable realizar siempre esta estrategia para evitar la amplificación de pseudogenes (Blount & Prakash, 2018; Borras et al., 2013; Brea-Fernandez et al., 2014; Ganster et al., 2010; van der Klift et al., 2016). Uno de los diseños más utilizados es la generación dos amplicones grandes, solapantes entre ellos, que cubren toda la extensión del gen. Para evitar la amplificación de los pseudogenes, los *primers* para la amplificación se encuentran localizados en regiones que no se encuentran en el pseudogen, como por ejemplo el exón 10, y que tampoco tienen polimorfismos conocidos (Clendenning et al., 2006; Wimmer & Wernstedt, 2014). En esta tesis se ha seguido esta estrategia para el análisis del gen *PMS2,* pero no para los otros genes MMR.

Por último, los estudios de ASE son una herramienta efectiva para detectar las diferencias de expresión entre el tránscrito con la variante y el *wild-type*. Estos ensayos se basan en la cuantificación relativa de los dos alelos del RNA mensajero gracias a la extensión

de un único nucleótido en la posición dónde se encuentra la variante o, en caso de que no se haya encontrado variante causal o para variantes intrónicas, un polimorfismo exónico en heterocigosis (Damaso et al., 2018). El ASE se puede determinar por diferentes técnicas, como la pirosecuenciación (Kwok et al., 2010), extensión del *primer* y desnaturalización en cromatografía líquida de alta eficacia (Aceto et al., 2009), o secuenciación de la reacción de extensión del nucleótido (SNuPE, del inglés "*Single NUcleotide Primer Extension*") (Castellsague et al., 2010; Tricarico et al., 2017), entre otros. En esta tesis doctoral hemos utilizado la metodología de SNuPE ya que tiene una complejidad técnica baja y una elevada sensibilidad (Borras et al., 2012; Borras et al., 2013; Castellsague et al., 2010; Damaso et al., 2018).

Estudios a nivel de proteína

Como se ha comentado en la Introducción de esta tesis, existe una amplia variedad de ensayos funcionales que se pueden realizar a nivel de proteína y cada uno de ellos evalúa el impacto de las variantes sobre un aspecto mole. Así, se han desarrollado ensayos para evaluar la actividad reparadora, la expresión y estabilidad, la localización subcelular, la interacción con otras proteínas, la resistencia a agentes alquilantes y la unión del ATP, entre otros (Borras et al., 2013; Guerrette et al., 1998; Heinen & Rasmussen, 2012; Heinen et al., 2002; Hinrichsen et al., 2013; Ollila et al., 2008; Rasmussen et al., 2012). Sin embargo, alrededor del 75% de las variantes patogénicas reportadas se asocian a defectos en la actividad reparadora o su impacto en la expresión y estabilidad de la proteína, mientras que los defectos en la localización subcelular y el resto de características son menos frecuentes (Hinrichsen et al., 2015). Por este motivo, los actuales algoritmos para interpretar las variantes a nivel funcional recomiendan testar primero la proteína a nivel de actividad reparadora y expresión y, sólo cuando no se ha detectado defecto en estos aspectos, realizar ensayos complementarios (Couch et al., 2008; Hinrichsen et al., 2015; Peña-Diaz & Rasmussen, 2016; Thompson et al., 2014).

En esta tesis doctoral se han realizado estudios funcionales de proteína a nivel de actividad reparadora y de expresión utilizando ensayos basados en sistemas *in vitro* con extractos celulares de proteínas humanas. Aunque estos tests se pueden realizar sobre levadura o células de mamífero, se recomiendan los ensayos con extractos celulares porque éstos permiten evaluar todo tipo de variantes independientemente de su grado de conservación evolutiva y recrean las condiciones genéticas y fisiológicas que se dan en las células *in vivo* (Peña-Diaz & Rasmussen, 2016; Richards et al., 2015; Thompson et al., 2014)

Los métodos para testar la actividad reparadora se discutirán en el apartado 1.2 de esta discusión. En cuanto a la expresión de la proteína, ésta se ha evaluado mediante la técnica de *Western Blot* y se han utilizado los mismos extractos celulares de proteínas humanas que en paralelo se han testado para la actividad reparadora. Los extractos se han conseguido mediante transfección transitoria de la proteína humana de interés en una línea celular humana deficiente en el gen MMR correspondiente. Si bien esta metodología presenta una baja complejidad técnica, deben tenerse en cuenta algunas consideraciones. Ciertas líneas no se transfectan de forma eficiente y la propia naturaleza del método implica que, en ocasiones, exista cierta heterogeneidad en los niveles de las proteínas exógenas expresadas, por lo que es importante asegurar una buena y homogénea eficiencia de transfección (Heinen & Rasmussen, 2012; Thompson et al., 2014). Por otro lado, este ensayo mide la expresión relativa de la proteína con la variante en relación a los niveles de expresión de la proteína *wildt-type*, así que es importante añadir en un mismo experimento varias réplicas del extracto con la proteína *wild-type* así como utilizar un control de carga (Thompson et al., 2014). Del mismo modo, dada la variabilidad de la técnica de *Western-Blot*, se recomienda hacer siempre varias réplicas experimentales.

Predicciones *in silico*

Actualmente existe una gran variedad de predictores *in silico* del efecto de las variantes sobre el RNA y la proteína, pero se ha demostrado que su sensibilidad y especificidad es moderada, de modo que se recomienda utilizarlos como apoyo a otras

evidencias para la clasificación de variantes, siempre y cuando se utilicen varios predictores y los resultados sean concordantes entre ellos (Choi et al., 2012; Moles-Fernandez et al., 2018; Richards et al., 2015; Thompson et al., 2014; Thusberg et al., 2011). Asimismo, las predicciones *in silico* también pueden utilizarse para priorizar qué ensayos funcionales realizar (Couch et al., 2008; Peña-Diaz & Rasmussen, 2016) o incluirse en los cálculos multifactoriales (Thompson, Goldgar, et al., 2013; Thompson, Greenblatt, et al., 2013; Thompson et al., 2014). Las normas de InSIGHT para la clasificación de variantes en genes MMR contemplan la utilización de las predicciones *in silico* en el cálculo multifactorial de probabilidad y para apoyar los resulados de los ensayos funcionales de RNA que se han realizado sobre RNA extraído de una muestra biológica del paciente. De este modo, si una variante ha demostrado generar un *splicing* aberrante en este tipo de muestra y el resultado está respaldado por varios predictores, la variante puede clasificarse como patogénica (Clase 5) (Thompson et al., 2014) (2018-06_InSiGHT_VIC_v2.4 https://www.insight-group.org/criteria/). De forma más laxa, las guías ACMG-AMP también contemplan las predicciones *in silico* como un criterio secundario que se puede sumar a otras evidencias para clasificar las variantes (Richards et al., 2015).

Cálculo multifactorial de probabilidad

Las evidencias cualitativas, como por ejemplo la cosegregación de la variante con la enfermedad o la presencia de MSI en los tumores, se pueden calibrar como razones de verosimilitud, o LR, para incorporarse a algoritmos multifactoriales Bayesianos y, así, obtener una probabilidad posterior de patogenicidad. El análisis multifactorial presenta la ventaja, por lo tanto, de proporcionar estimaciones cuantitativas de la patogenicidad de una variante, lo que permite clasificarla de forma más reproducible.

El modelo multifactorial para los genes MMR que se ha utilizado en esta tesis doctoral combina, concretamente, las LR de las predicciones *in silico*, la cosegregación y las características del tumor (presencia de MSI) para calcular esta probabilidad posterior de patogenicidad.

Inicialmente, el sistema del cálculo multifactorial fue desarrollado para los genes *BRCA1* y *BRCA2* (Goldgar et al., 2004; Lindor et al., 2012) y posteriormente fue modificado para aplicarse a los genes MMR con algunas variaciones (Thompson, Greenblatt, et al., 2013). Por ejemplo, el modelo para los genes MMR combina las predicciones *in silico* de los programas MAPP y PolyPhen-2.1 para determinar las probabilidades a priori de patogenicidad de las mutaciones *missense*, en vez de utilizar el algoritmo Align-GVGD como se hace para *BRCA1* y *BRCA2*. Para las variantes intrónicas, en cambio, el modelo MMR aplica las mismas LR calculadas para *BRCA1* y *BRCA2*, ya que en su momento no se disponía de una serie suficientemente grande para calibrar esta evidencia. Por otro lado, el modelo inicial para variantes MMR tiene también en cuenta las evidencias de cosegregación y la MSI y BRAF como características moleculares de los tumores, mientras que el modelo *BRCA1* y *BRCA2* también incluye las LR derivadas de la co-ocurrencia en *trans* con otra variante patogénica y la historia familiar y personal del paciente.

Con el tiempo, los modelos multifactoriales han evolucionado y, actualmente, el modelo multifactorial para *BRCA1* y *BRCA2* también incluye los datos de la frecuencia poblacional y los resultados funcionales a nivel de RNA y de proteína, aunque aún es necesario validar estos resultados en una serie de variantes más amplia para refinar los límites de confianza (Parsons et al., 2019). Del mismo modo, recientemente también se han calibrado los resultados del ensayo *in vitro* de actividad reparadora llamado CIMRA (del inglés "*Cell-free In vitro MMR Activity*"), para incluirlo en el cálculo multifactorial de probabilidad para los genes MMR (Drost et al., 2018) (ver apartado 1.2.1 y 1.2.3).

Como se ha indicado, el modelo multifactorial para los genes MMR solo incluye como característica tumoral la MSI y BRAF, pero no los datos de IHC. Esto se debe a que, en su inicio, se decidió no incluir los datos de IHC de los tumores para evitar posibles sobreestimaciones de la patogenicidad (Pastrello et al., 2011; Thompson, Goldgar, et al., 2013). Como esta decisión suponía una limitación del modelo, se ha empezado a trabajar en la inclusión de los datos de IHC. Así, la calibración de las características tumorales tendría en

cuenta varios escenarios: sólo datos de IHC, sólo MSI o ambos si los resultados son concordantes entre ellos. De darse esta última situación, las LR de IHC y MSI no se sumarían, sino que se combinarían para evitar sobreestimaciones **(Tabla 12)** (datos no publicados facilitados por B. A. Thompson).

Tabla 12. Razones de verosimilitud o LR (*Likelihood Ratios*) calculadas para la pérdida o conservación de la expresión de la proteína MMR en el tumor, presencia o ausencia de MSI y combinación de ambas características. Datos no publicados facilitados por B. A. Thompson.

Gen donde se localiza la variante	Resultados IHC		Resultados MSI			Concordancia entre IHC y MSI	
	Pérdida de expresión	Expresión conservada	MSI-H	MSI-L	MSS	MSI-H / MSI-L y pérdida IHC	MSI-L / MSS y IHC conservada
MLH1	5,02	0,19	4,07	0,13	0,03	5,32	0,03
MSH2	25,2	0,07	3,9	0,14	0,07	39,7	0,08
MSH6	5,48	0,38	4,19	0,26	0,18	7,13	0,26

Tanto la calibración del ensayo funcional CIMRA como la de los datos de IHC suponen una mejora substancial del modelo multifactorial para los genes MMR. No obstante, este modelo aún presenta limitaciones a tener en cuenta, principalmente derivadas de algunas asunciones que hace. Por ejemplo, debido a que la probabilidad a priori se basa en el efecto en proteína, ciertas mutaciones de *splicing* pueden presentar probabilidades de patogenicidad a priori muy bajas que dificultará su clasificación mediante este cálculo (Lindor et al., 2012; Thompson, Goldgar, et al., 2013; Thompson et al., 2014). Por este motivo, recomendamos recopilar la máxima información disponible sobre la variante y los individuos portadores, así como evaluar múltiples evidencias, para generar una clasificación robusta de las variantes.

1.1.2 Sistemas de clasificación de las variantes

Tras un análisis exhaustivo las variantes estudiadas en esta tesis doctoral se han clasificado aplicando las guías elaboradas por el InSIGHT (2018-06_InSiGHT_VIC_v2.4 https://www.insight-group.org/criteria/). En estas guías se indica cómo aplicar cada evidencia y qué criterios ha de cumplir una variante para entrar en cada categoría de

clasificación. Además, al tratarse de unas guías específicas para genes MMR, también recoge los resultados del cálculo multifactorial de probabilidad para el modelo MMR e indica, para cada categoría de clasificación, la probabilidad posterior de patogenicidad. Además, también incluye cómo interpretar los resultados de los ensayos funcionales para variantes MMR (Thompson et al., 2014).

A pesar de su especificidad, las guías de InSIGHT son más conservadoras a la hora de asignar una clasificación de patogenicidad o benignidad que las guías ACMG-AMP (Richards et al., 2015), de carácter general para la clasificación de variantes genéticas asociadas a enfermedades mendelianas, tienen en cuenta evidencias similares que las guías de InSIGHT, aunque no incluyen evidencias específicas asociadas a los genes MMR (características de los tumores, estudios funcionales específicos, cálculo multifactorial...). A pesar de ello, debido al uso extendido de las guías ACMG-AMP, el comité de interpretación de variantes de InSIGHT está trabajando para adaptar las guías ACMG-AMP para la clasificación de variantes en genes MMR, como ya se ha hecho para otros genes como *PTEN o CDH1* (K. Lee et al., 2018; Mester et al., 2018).

Además, recientemente se ha publicado un estudio que escala los criterios para la clasificación de variantes de la guía ACMG-AMP y modela su combinación para generar probabilidades de patogenicidad y así poder aplicar estas reglas dentro de un modelo Bayesiano (Tavtigian et al., 2018). Esta transformación de los criterios ACMG-AMP, antes cualitativos, permite una clasificación cuantitativa de las variantes y brinda la oportunidad de refinar los criterios de clasificación y combinar tanto los que son a favor como en contra de la patogenicidad para obtener una probabilidad final de patogenicidad. Además, abre la puerta a combinarse con los otros cálculos multifactoriales ya existentes.

1.1.3 Rendimiento de la caracterización exhaustiva

En el Programa de Cáncer Hereditario del Instituto Catalán de Oncología, los estudios de reclasificación de variantes en genes MMR comenzaron en 2008 con la

identificación de 2 variantes recurrentes en *MLH1*, que resultaron ser mutaciones fundadoras en la población española (Borras et al., 2010). Posteriormente, el análisis de 15 VUS adicionales, 8 en *MLH1*, 6 en *PMS2* y 1 en *MSH2*, permitió reclasificar 12 de ellas a patogénicas o probablemente neutras (Borras et al., 2012; Borras et al., 2013; Menendez et al., 2010). Actualmente, el número de VUS identificadas por la Unidad de Diagnóstico del Programa ha crecido considerablemente, sobre todo desde la implantación de los paneles NGS en el diagnóstico (Feliubadaló et al., 2019).

En la presente tesis se analizaron 19 variantes inicialmente clasificadas como VUS (Clase 3) o probablemente patogénicas (Clase 4) **(Tabla 11)**. Primero, se realizaron dos trabajos que caracterizaban de forma exhaustiva 2 variantes en *MLH1* y otras 2 en *PMS2* (González-Acosta et al., 2017 – Artículo 1; Pineda, González-Acosta et al., 2014 – Anexo 1) y, después, se decidió estudiar las VUS dentro del contexto de los pacientes con Síndrome Lynch-Like (Vargas-Parra et al., 2017 – Anexo 1; Dámaso et al., en preparación – Anexo 1). La serie de pacientes SLL estaba formada por 156 individuos con sospecha de SL debido a las características de sus tumores. Tras el análisis exhaustivo de estos individuos, se identificaron 6 variantes probablemente patogénicas (Clase 4) en *MSH2* y 19 VUS (Clase 3) (2 en *MLH1*, 1 en *PMS2*, 9 en *MSH2* y 7 en *MSH6*). Una de las variantes de *MSH6*, la c.3226C>T (p.Arg1076Cys), fue reclasificada a probablemente patogénica (Clase 4) por el InSIGHT al inicio del estudio por lo que se desestimó realizar estudios adicionales. De las variantes restantes, sólo se dispuso de material biológico y/o datos de cosegregación de 15 de ellas (12 en *MSH2* y 3 en *MSH6*) que fueron el objetivo de esta tesis **(Tabla 11).**

Por lo tanto, de las 29 variantes MMR identificadas en los 4 trabajos anteriores y susceptibles a estudio de reclasificación, se pudieron analizar 19 (González-Acosta et al., 2017 – Artículo 1; Pineda, González-Acosta et al., 2014 – Anexo 1; Vargas-Parra et al., 2017 – Anexo 1 y Dámaso et al., en preparación – Anexo 1). Mediante la integración del cálculo multifactorial de patogenicidad (donde se incluye la información de los datos clínico-patológicos y de cosegregación) y el resultado de los estudios funcionales a nivel de RNA y

proteína, 17 de las variantes se han reclasificado a patogénica (Clase 5) o neutra (Clase 1) **(Tabla 11)**, lo que nos da una tasa de reclasificación de las variantes del 59% (17/29).

El cálculo multifactorial de probabilidad permitió reclasificar el 68% (13/19) de las variantes estudiadas (**Tabla 11**). Además, para dos de ellas, las variantes *MLH1* c.121G>C (p.Asp41His) y *PMS2* c.2444C>T (p.Ser815Leu), los resultados de los estudios funcionales realizados elucidaron el mecanismo de patogenicidad, ya que ambas variantes presentaron una actividad reparadora deficiente. Por otro lado, la variante en *MSH6* c.1618_1620del (r.1618_1620del, r.1607_3172del; p.Leu540del, p.Ser536_Asp1058delinsAsn) presentó un efecto deletéreo parcial sobre el RNA al producir tránscrito wild-type y aberrante. No obstante, como no se estudió la variante a nivel de proteína, no se puede descartar que la patogenicidad se deba únicamente al efecto en RNA.

Por otro lado, el 21% (4/19) de las variantes pudieron reclasificarse gracias a los resultados de los estudios de RNA. Las variantes de *MSH2* c.211G>C (r.195_211del, p.Tyr66Serfs*10), c.1276G>A (r. r.1230_1277del, p.Ile411_Gly426del), la duplicación del exón 8 (r. r.1277_1387dup, p.Val463Glufs*11) y la duplicación del exón 11 (r.1662_1759dup, p.Gly587Alafs*3) presentaron un *splicing* aberrante y se reclasificaron a mutaciones patogénicas (Clase 5) debido a que generaban un tránscrito aberrante con un codón de parada prematuro o una deleción dentro de un dominio funcional **(Tabla 11)** (Thompson et al., 2014).

El 11% (2/19) restante de las variantes permaneció como VUS (Clase 3). No obstante, aunque no han podido reclasificarse, las predicciones *in silico* y los resultados funcionales obtenidos hasta la fecha para las variantes *PMS2* c.2149G>A (p.Val717Met) y *MSH2* c.1787A>G (p.Asn596Ser) sugieren su neutralidad **(Tabla 11)**. La variante *PMS2* c.2149G>A presentó unos niveles de actividad reparadora y expresión de la proteína similares al *wild-type* y tampoco presentó defectos a nivel de *splicing* del RNA. De forma similar, la variante *MSH2* c.1787A>G no presentó defectos a nivel de RNA y los programas *in silico* la predicen

como neutra a nivel de proteína. Con todo, serían necesarias evidencias adicionales para poder reclasificar estas variantes mediante las guías InSIGHT.

En conclusión, gracias al análisis exhaustivo de las variantes, éstas se han podido reclasificar en una proporción significativa, confirmando así el SL en las familias portadoras de las mutaciones patogénicas lo que mejora su manejo clínico. Nuestros resultados confirman la importancia de acumular diferentes evidencias para garantizar la clasificación sólida de las variantes (Amendola et al., 2016; Thompson et al., 2014; Tricarico et al., 2017; van der Klift et al., 2016; Zuntini et al., 2018). Basándonos en nuestras observaciones, proponemos una modificación del algoritmo propuesto en Borràs *et al.*, 2012 para el estudio de las VUS en genes MMR (Borras et al., 2012) que incluye el cálculo multifactorial de probabilidad como paso a realizar en paralelo a los estudios funcionales debido a su buena tasa de reclasificación. En cuanto a los estudios funcionales, mantenemos la priorización en base a las predicciones *in silico*, sobre todo por el buen rendimiento que presentan los estudios de *splicing* al combinarse con las predicciones *in silico* (Houdayer et al., 2012; Moles-Fernandez et al., 2018; Thompson et al., 2014) y recomendamos realizar primero el ensayo *in vitro* de actividad reparadora junto al análisis de la expresión de las proteínas, ya que ambos utilizan los mismos extractos celulares **(Figura 17)**.

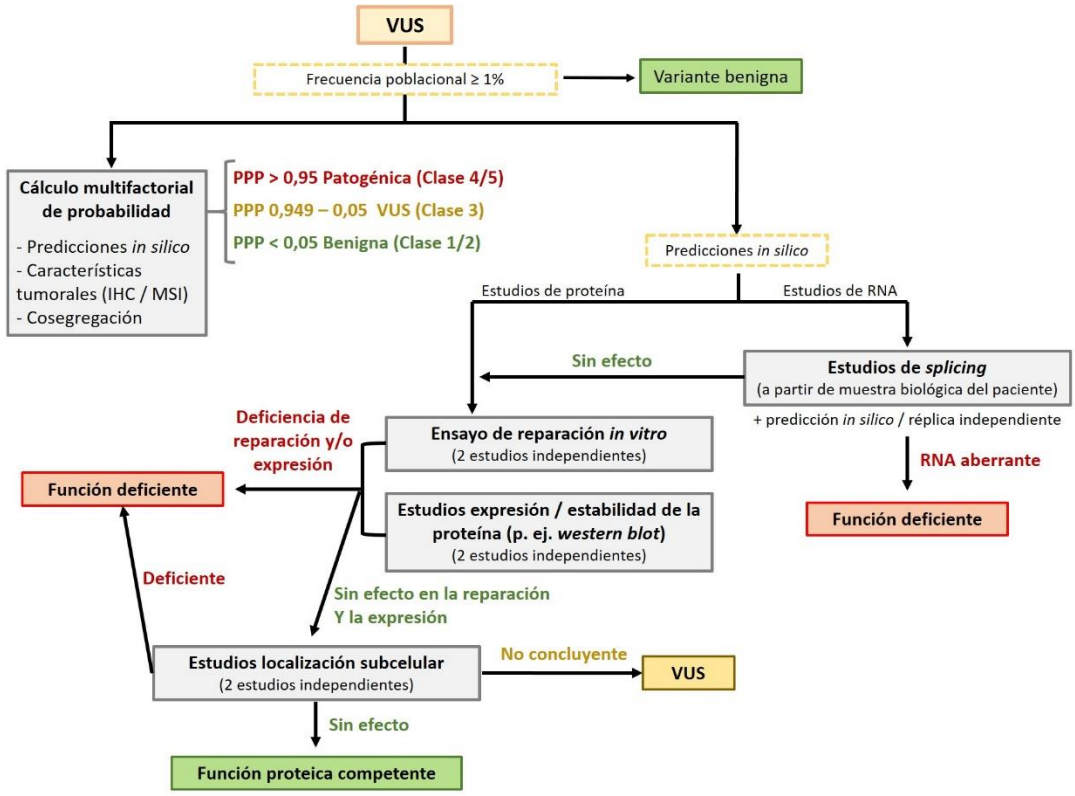


Figura 17. Algoritmo propuesto para el estudio de las VUS en genes MMR.

1.2 Validación del ensayo *in vitro* de actividad reparadora

1.2.1 Tipos de ensayos *in vitro* de actividad reparadora con extractos celulares

Como se ha comentado anteriormente, en esta tesis doctoral se ha utilizado un ensayo *in vitro* de actividad reparadora que está basado en la utilización de extractos celulares de proteínas humanas. La utilización de este tipo de sistema permite estudiar todo tipo de variantes y recrea las condiciones genéticas y fisiológicas que se dan en las células *in vivo* humanas (Peña-Diaz & Rasmussen, 2016).

Dentro de los ensayos *in vitro* de actividad reparadora con extractos celulares, existen variaciones técnicas. Una de las más importantes es a nivel de cómo se obtiene la proteína mutada que se analizará en el ensayo. Nosotros (González-Acosta *et al.*, en primera

revisión en J Mol Diagn - Artículo 2) hemos utilizado un sistema de transfección transitoria en el que la proteína con la variante se clona en un plásmido de expresión y se transfecta en una línea celular humana deficiente en esa proteína, para luego purificar la proteína expresada y utilizarla en el ensayo de reparación. Sin embargo, existen otras aproximaciones que obtienen dicha proteína mediante expresión en células de insecto S*f*9 por infección con baculovirus (Ollila et al., 2006; Raevaara et al., 2005), o mediante transcripción y traducción *in vitro* (Drost et al., 2013; Drost et al., 2018; Drost et al., 2010; Drost et al., 2012). Ambos métodos permiten obtener grandes cantidades de proteína, pero los niveles de expresión de la proteína obtenidos no suelen correlacionar con los que se darían en las células humanas, por lo que estos extractos de proteína no se pueden utilizar para testar después los niveles de expresión (por ejemplo, mediante *western blot*), la interacción con otras proteínas o la localización subcelular. De querer estudiar alguna de estas características, será necesario producirla de nuevo mediante transfección transitoria en células humanas (Raevaara et al., 2005). En comparación con otros estudios, nuestra aproximación permite evaluar tanto la actividad reparadora como la expresión de la proteína utilizando los mismos extractos proteicos obtenidos mediante transfección transitoria, y presenta la optimización tanto de los reactivos utilizados como de las condiciones de reacción y su validación (Borras et al., 2012; Borras et al., 2013; Hinrichsen et al., 2013; Plotz et al., 2006; Takahashi et al., 2007). En la **Tabla 13** se resumen las principales ventajas y limitaciones de cada enfoque.

Recientemente, se ha adaptado un ensayo *in vitro* de actividad reparadora parecido al nuestro para la identificación de individuos CMMRD (Shuen et al., 2019). En este ensayo, en lugar de complementar extractos nucleares de una línea celular deficiente en reparación con extractos totales de una línea transfectada con la proteína de interés, se testan directamente extractos proteicos de linfocitos inmortalizados derivados de pacientes con sospecha de CMMRD y se mide su capacidad de reparación. Después, si las células derivadas de estos pacientes son incapaces de reparar el substrato con el apareamiento erróneo, se complementan con la proteína *wild-type* MMR (generada por traducción *in vitro* o por transfección transitoria) en un nuevo ensayo de reparación para asegurar que el defecto de

reparación observado se debe realmente a la deficiencia MMR de estas células. Si la capacidad de reparación se restaura, se confirma la deficiencia MMR de las células y, en consecuencia, la condición CMMRD. Esta aproximación ha demostrado tener una sensibilidad y especificidad del 100% y se ha propuesto como herramienta diagnóstica para la preselección de este tipo de pacientes. Este punto se discutirá en más detalle en el apartado 2.1.2 de esta Discusión.

Eventualmente, nuestro ensayo *in vitro* de actividad reparadora también podría adaptarse para este uso, teniendo a su favor una validez analítica ya comprobada, requerimiento indispensable para poder determinar la validez y utilidad clínica antes de implementarlo en el diagnóstico (Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Lijmer, et al., 2003; Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Moher, et al., 2003; Burke, 2014).

Tabla 13. Principales ventajas y limitaciones de los ensayos *in vitro* de actividad reparadora con extractos celulares actualmente utilizados.

	Ensayos *in vitro* de actividad reparadora (estudio actual)	**Ensayo *in vitro* de actividad reparadora** (Takahashi, Shimodaira et al. 2007; Borras, Pineda et al. 2012 y 2013; Hinrichsen, Brieger et al. 2013)	**Ensayo *in vitro* de actividad reparadora** (Raevaara, Korhonen et al. 2005)	**Ensayo in vitro de actividad reparadora (CIMRA)** (Drost, Zonneveld et al. 2010 y 2012; Drost, Tiersma et al. 2018)
Ventajas	Protocolos estandarizados Variabilidad analítica y reproducibilidad testada El mismo extracto proteico se puede utilizar para analizar la expresión	El mismo extracto proteico se puede utilizar para analizar la expresión	El sistema de expresión en Baculovirus permite producir grandes cantidades de proteína	Cuantificación de la eficiencia de reparación mediante un método altamente sensible Resultados del ensayo CIMRA en combinación con las predicciones in silico se han calibrado para su integración en el cálculo multifactorial de probabilidad
Limitaciones	Resultados no calibrados para el cálculo multifactorial de probabilidad No se han establecido puntos de corte para la patogenicidad de las	Resultados no calibrados para el cálculo multifactorial de probabilidad No se han establecido puntos de corte para la patogenicidad de las	El mismo extracto proteico no se puede utilizar para analizar la expresión Resultados no calibrados para el cálculo multifactorial de probabilidad No se han establecido puntos de corte para la patogenicidad de las	El mismo extracto proteico no se puede utilizar para analizar la expresión

1.2.2 Interpretación de los resultados del ensayo *in vitro* de actividad reparadora

En nuestro trabajo hemos utilizado las variantes ampliamente caracterizadas p.Ile219Val (neutra) y p.Gly67Arg (patogénica) como variantes control del ensayo y, para interpretar los porcentajes de reparación obtenidos en cada variante, hemos utilizado unos puntos de corte conservadores de <35% para la deficiencia de reparación y >64% para la capacidad de reparación, como sugiere el InSIGHT tras recopilar datos funcionales de decenas de variantes analizadas mediante diferentes aproximaciones (Thompson et al., 2014).

Como se esperaba, la variante deficiente en actividad reparadora p.Thr82Ala ha presentado unos niveles de actividad similares a los del control deficiente claramente por debajo del 35% de actividad reparadora (3,2% ± 5,5), mientras que la variante neutra p.Val716Met ha presentado unos niveles similares al control de reparación por encima del 64% de actividad (79,3% ± 9,4). En cuanto a las variantes p.Ala681Thr y p.Leu622His, ambas clasificadas como patogénicas mediante el cálculo multifactorial de probabilidad y con niveles de reparación no concluyentes según lo reportado previamente en la literatura, la variante p.Ala681Thr ha presentado una actividad reparadora intermedia (54,9% ± 17,6) mientras que la variante p.Leu622His ha mostrado unos niveles de reparación por debajo del 35% y similares al control deficiente (19,5% ± 15,5).

Estudios previos al nuestro han utilizado otros criterios para interpretar los porcentajes de reparación obtenidos; sin embargo, todos ellos se basan en la comparación con variantes control. Por ejemplo, Takahashi y colaboradores decidieron utilizar el valor de reparación obtenido en la variante neutra p.Ile219Val, aproximadamente del 60% en su ensayo de actividad reparadora, como umbral para la capacidad de reparación, ya que los otras variantes neutras que estudiaron en su trabajo superaban este valor (Takahashi et al., 2007). Drost y colaboradores, en cambio, propusieron utilizar el valor medio de reparación obtenido en los controles deficientes como punto de corte para determinar si la variante conservaba la actividad reparadora, es decir, si ésta presentaba unos niveles de reparación

significativamente mayores que los controles deficientes, se podía considerar que mantenía la capacidad de reparación (Drost et al., 2012). Para establecer un punto de corte específico para nuestro ensayo, sería necesario analizar un gran número de variantes clasificadas clínicamente como patogénicas o neutras y determinar la especificidad y sensibilidad asociada a diferentes puntos de corte.

1.2.3 Utilidad del ensayo *in vitro* de actividad reparadora para determinar la patogenicidad de las variantes

Uno de los principales objetivos de esta tesis doctoral ha sido optimizar el ensayo *in vitro* de actividad reparadora y validarlo para su uso en la evaluación de la patogenicidad de las variantes MMR. A nivel analítico, el ensayo ha demostrado robustez y reproducibilidad en el análisis de la actividad reparadora de 6 variantes en *MLH1*; no obstante, su validez clínica, entendida como la capacidad para clasificar correctamente una variante como patogénica o benigna (Burke, 2014), aún no se ha evaluado. Para ello sería necesario establecer unos puntos de corte de patogenicidad y neutralidad, siguiendo la misma idea descrita en el apartado anterior (apartado 1.2.2). No obstante, cabe tener en cuenta que la patogenicidad de una variante MMR puede deberse también a defectos en la expresión de la proteína o de *splicing*, entre otras causas. Así por ejemplo, la variante p.Ala681Thr fue clasificada como patogénica gracias al cálculo multifactorial de probabilidad pero, según lo reportado en la literatura, retiene la actividad reparadora (69%-115%) (Hinrichsen et al., 2013; Hinrichsen et al., 2015; Raevaara et al., 2005; Takahashi et al., 2007). En cambio, la disminución de los niveles de expresión de la proteína del 50% sugiere que su patogenicidad es causada por defectos en la expresión (Hardt et al., 2011; Hinrichsen et al., 2013; Hinrichsen et al., 2015). En nuestro ensayo, esta variante presentó actividad reparadora intermedia (55%). Al estudiarse la expresión de la proteína mediante *Western blot*, los niveles de expresión fueron del 33%, apoyando su clasificación como variante patogénica obtenida por el cálculo multifactorial.

De forma similar, la variante p.Leu622His fue reportada como una variante con expresión reducida pero actividad reparadora competente (Borras et al., 2010; Hinrichsen et al., 2013; Kosinski et al., 2010; Takahashi et al., 2007). Es más, un trabajo previo de nuestro grupo la reportó como una mutación fundadora española con penetrancia moderada y demostró que presentaba una estabilidad de la proteína reducida al tratar las células HCT116 con cicloheximida (Borras et al., 2010). En nuestro ensayo, sin embargo, la variante ha presentado niveles defectivos de actividad reparadora (19%) y expresión (15%). Esto podría indicar que nuestro ensayo de actividad reparadora es más sensible a los defectos en expresión que otros ensayos de reparación.

Es interesante mencionar que la variante p.Val716Met se ha descrito como una variante neutra, con una ligera reducción de la expresión de la proteína pero reteniendo la actividad reparadora, en base a los ensayos realizados en muestras obtenidas de individuos con SL. No obstante, la naturaleza de esta variante es actualmente motivo de controversia, ya que se ha reportado junto a otra variante patogénica en *trans* en dos individuos con fenotipo de CMMRD (R. Gallon et al., 2019; Marcos et al., 2006) (ver apartado 2.2.3).

Recientemente, se han realizado importantes progresos en el desarrollo de estudios funcionales de alto rendimiento para variantes en *BRCA1*, que han establecido puntos de corte para clasificar las variantes como patogénicas, VUS o benignas, en función de los resultados de los ensayos funcionales (Drost et al., 2018; Findlay et al., 2018; Starita et al., 2018). Starita y colaboradores caracterizaron un ensayo de reparación homóloga en *BRCA1*, principal función de este gen, para discriminar entre las variantes patogénicas y benignas localizadas en los primeros 192 residuos de la proteína (correspondientes al dominio N-terminal, en el que también se incluye el dominio RING). Sin embargo, aunque obtuvieron un 100% de especificidad para clasificar las variantes, la sensibilidad fue del 87,5%, ya que su ensayo clasificó erróneamente aquellas variantes que presentaban defectos en otras características como el *splicing* (Starita et al., 2018). Findlay y colaboradores, en cambio, estudiaron los efectos en la supervivencia celular de células haploides de cerca de 4000

variantes puntuales en *BRCA1* y, a la hora de clasificarlas como patogénicas o benignas, obtuvieron una sensibilidad y especificidad del 96,7% y 98,2%, respectivamente. Esto fue gracias a que la supervivencia celular es una medida que permite integrar los efectos tanto a nivel de *splicing* del RNA como de expresión y función de la proteína (Findlay et al., 2018).

En el caso de los genes MMR, Drost y colaboradores calibraron el ensayo CIMRA y transformaron los resultados de actividad reparadora en probabilidades de patogenicidad, estableciendo umbrales para clasificar las variantes en las 3 categorías anteriormente mencionadas (Drost et al., 2018). Para ello, estudiaron un total de 70 variantes *missense* previamente clasificadas como patogénicas (Clase 4 y 5) o benignas (Clase 1 y 2). Aunque la especificidad para clasificar correctamente las variantes patogénicas fue del 100%, la sensibilidad del método fue del 60%. Y, para las variantes benignas, se obtuvo una especificidad del 96% y una sensibilidad del 75%. Globalmente, su ensayo CIMRA pudo clasificar correctamente el 65% de las variantes estudiadas, aunque el 32% fue clasificado como VUS y un 3% presentó clasificaciones discordantes. Al combinar los resultados funcionales con las predicciones *in silico*, consiguió incrementar el número de variantes correctamente clasificadas (87%), pero el porcentaje de discordantes se mantuvo. Una de las variantes discordantes, erróneamente clasificada como benigna a pesar de ser patogénica, fue precisamente la variante en *MLH1* p.A681T, discutida anteriormente. En el ensayo CIMRA, esta variante presentó un 73% de actividad reparadora y este dato, al incluirse en el cálculo multifactorial, tuvo suficiente fuerza como para clasificar la variante como benigna. Otras variantes, clasificadas como VUS en el ensayo CIMRA, presentaron defectos en la localización subcelular de la proteína o en la interacción con otras proteínas MMR. Todas estas discrepancias en la clasificación fueron consecuencia, precisamente, de que el ensayo CIMRA sólo mide la capacidad de reparación y la patogenicidad de una variante puede deberse a otros defectos de la proteína mutada.

La calibración y validación del ensayo CIMRA para su integración en los modelos factoriales abre la puerta a la posibilidad de integrar datos funcionales de otros ensayos,

como por ejemplo el nuestro, en el cálculo multifactorial de probabilidad. Una clasificación final correcta necesitará integrar diferentes líneas de evidencia, así como resultados de diferentes tipos de ensayos funcionales que evalúen el efecto de la variante tanto a nivel de RNA como de proteína, tal como se ha discutido en el apartado 1.1.3 de esta Discusión **(Figura 17)**.

2. Detección con alta sensibilidad de la MSI en tejido normal

2.1 El análisis *hs-MSI* como metodología para la determinación con alta sensibilidad de la MSI

En esta tesis doctoral se ha desarrollado una metodología para la detección con alta sensibilidad de la MSI en tejido normal basada en el desarrollo de un panel NGS de regiones microsatélite (González-Acosta *et al.,* 2019 - Artículo 3). La aproximación *hs-MSI* (por sus siglas del inglés "*high sensitivity-MSI*") ha demostrado una elevada precisión para detectar niveles de MSI en sangre periférica de pacientes CMMRD, significativamente mayores a los detectados en controles sanos no portadores de variantes MMR, pacientes con SL u otros síndromes con los que solapa fenotípicamente la condición CMMRD, lo que permite discriminar los pacientes CMMRD, incluso a los portadores de mutaciones bialélicas en *MSH6*, del resto de grupos con una sensibilidad y especificidad del 100%. Además, nuestra aproximación requiere de muy poca cantidad de DNA para realizar el ensayo (50 ng) y se estima que los resultados se pueden obtener en el plazo de una semana.

2.1.1 Aproximaciones basadas en el análisis de MSI con alta sensibilidad mediante NGS

Este mismo año se ha sido publicada una aproximación similar a la nuestra también basada en el desarrollo de un panel NGS de microsatélites para la detección de MSI en sangre de pacientes CMMRD (R. Gallon et al., 2019). Dicho método analiza 24 microsatélites mononucleótidos monomórficos, seleccionados a partir de datos de WGS de tumores colorrectales disponibles en el TCGA. En una primera fase, identificaron todos los microsatélites mononucleótidos de entre 7 y 12 pb (210000 microsatélites), y seleccionaron aquellos que presentaban diferencias en longitud entre controles y el grupo de tumores MSI-H (529 microsatélites). A partir de aquí, fueron aplicando diferentes filtros para reducir el número de marcadores a analizar hasta 24 marcadores. La sensibilidad de este método para

la detección de CMMRD fue del 97%, ya que hubo cierto solapamiento en el MSI score de individuos CMMRD con el de individuos control.

Ambos métodos utilizan los *molecular barcodes* para reducir los errores de secuenciación y utilizan sondas basadas en la complementariedad con los extremos de la región de DNA de interés (también llamadas *molecular inversion probes*) para capturar estas secuencias de DNA y analizarlas mediante NGS. No obstante, se diferencian en la estrategia de la selección de microsatélites, el número de marcadores analizados y otros aspectos técnicos como el propio método de captura de las regiones de interés o cómo se determina la MSI. Por otro lado, el método descrito por Gallon y colaboradores utiliza una profundidad de secuenciación menor a la utilizada en nuestro panel *hs-MSI*, lo que permite que el coste por muestra analizada mediante esta aproximación sea menor **(Figura 18)**.

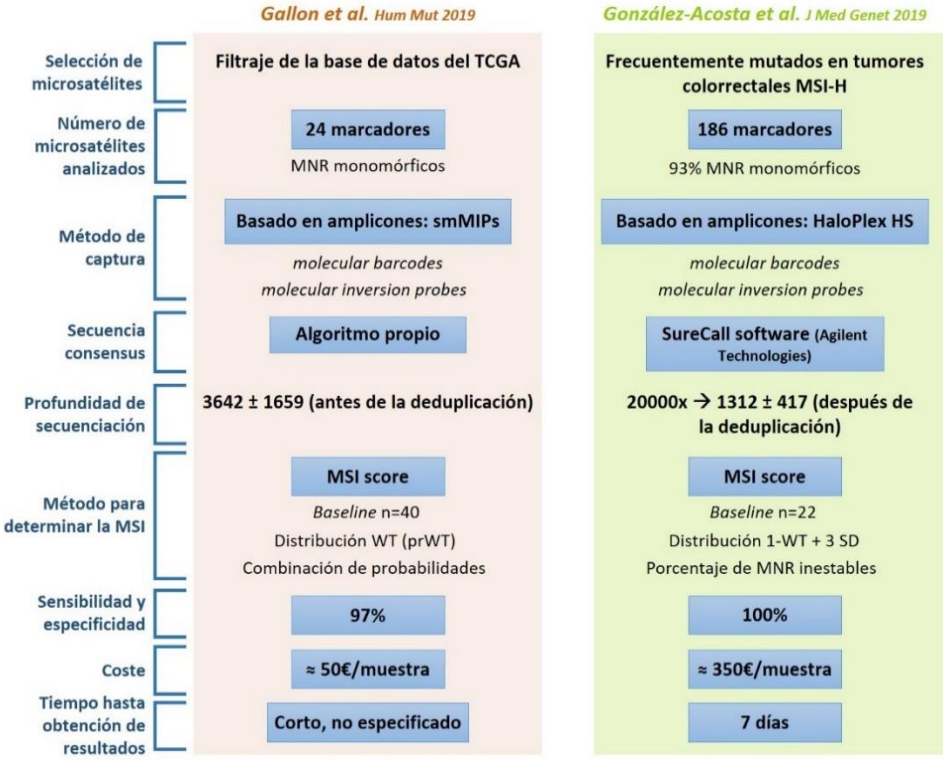


Figura 18. Comparativa de las aproximaciones NGS para la detección de MSI germinal publicadas en Gallon *et al.*, y el presente estudio.

Las puntuación de MSI obtenida por ambas metodologías en las muestras compartidas presentaron una buena correlación, lo que apunta a que los métodos basados en la detección de MSI mediante NGS serán utilizadas para la identificación de individuos CMMRD. Además, independientemente de la aproximación NGS utilizada, estos métodos presentan grandes ventajas frente los métodos utilizados hasta el momento para detectar MSI en tejido normal **(Tabla 14)**. Así, aunque la aproximación *gMSI* de Ingham y colaboradores presenta pocos requisitos técnicos y el coste por muestra es mínimo, carece de sensibilidad para identificar a portadores bialélicos de mutaciones en *MSH6* (Ingham et al., 2013). En cambio, la técnica de *evMSI* presentada por Bodo y colaboradores, en combinación con el ensayo de tolerancia a agentes metilantes, presenta una alta sensibilidad y especificidad para la identificación de CMMRD pero es técnicamente compleja y los resultados no se obtienen hasta al cabo de varios meses, lo que la hace inadecuada en el contexto del CMMRD en el cual habitualmente requiere un diagnóstico rápido (Bodo et al., 2015).

Tabla 14. Ensayos actualmente utilizados para detectar la MSI en tejido normal de individuos con CMMRD.

	Ingham et al.* *Hum Mutat 2013*	***Bodo et al.*** *Gastroenterology 2015*	***Gallon et al.*** *Hum Mutat 2019*	***González-Acosta et al.*** *J Med Genet 2019*
Técnica	gMSI	evMSI	MSI assay	hs-MSI
Descripción	PCR convencional	PCR convencional	Panel NGS personalizado	Panel NGS personalizado
Microsatélites analizados	D2S123, D17S250, D17S791	NR27, NR21, BAT26	24 MNR	186 microsatélites
Tipo de muestra	DNA de sangre periférica	DNA de linfocitos inmortalizados	DNA de sangre periférica	DNA de sangre periférica o mucosa bucal
Cantidad DNA	20 ng	20 ng	100 ng	50 ng
Sensibilidad/ especificidad	D17S791: 100% / 98,9% D2S123: 50% / 100% D17S250: 100% / 98,8%	100% / 100% (estudio caso-control)	97% / 100%	100% / 100%
Tiempo hasta el diagnóstico	2 días	45 - 120 días	Corto, no especificado	7 días
Coste relativo	+	+++	++	+++

* Sensibilidad y especificidad evaluada en portadores bialélicos de *MLH1*, *MSH2*, *PMS2*. Portadores *MSH6* excluídos por falta de sensibilidad.

2.1.2 Aportación del análisis *hs-MSI* al diagnóstico de CMMRD

Cuando se diagnostica un caso de cáncer infantil y el niño cumple con los criterios de sospecha clínica de CMMRD descritos en *Wimmer et al.* **(Tabla 8)** (Wimmer et al., 2014), se procede a realizar el estudio genético de los genes MMR. La identificación de mutaciones bialélicas patogénicas en alguno estos genes es la única manera de confirmar y validar el diagnóstico de CMMRD; sin embargo, cuando el estudio no es concluyente debido a la detección de VUS o porque no se ha detectado ninguna mutación, se han propuesto la detección de pérdida de las proteínas MMR mediante IHC y/o la detección de MSI en tejido normal como aproximaciones alternativas para confirmar con alta fiabilidad la sospecha de CMMRD. No obstante, las técnicas de *gMSI* y *evMSI*, discutidas en el apartado anterior, presentan diversas limitaciones, como la falta de sensibilidad para los casos *MSH6* o un largo tiempo hasta el diagnóstico (Bodo et al., 2015; Ingham et al., 2013). En cuanto a la IHC, ésta puede resultar no informativa cuando se trata de mutaciones que no afectan la expresión de la proteína (Okkels et al., 2012; Taeubner et al., 2018).

Por el contrario, cuando se considera estudiar un niño sano con sospecha clínica de NF1 pero sin mutación identificada en *NF1* o *SPRED1*, o existe una sospecha familiar de CMMRD (al tener, por ejemplo, un hermano afecto de este síndrome), la realización del estudio genético es motivo de controversia debido a las consecuencias que puede comportar un resultado positivo ligado a un cribado intensivo de diferentes tumores y en un contexto de conocimiento limitado de la historia natural de la enfermedad, como se ha comentado en la Introducción de estas tesis (apartado 3.2.2). En estos casos se ha propuesto realizar una preselección de los candidatos a estudio genético en dos pasos: 1) una primera pre-selección en base a las características clínicas y/o familiares; 2) seleccionar aquellos pacientes de 1) que hayan demostrado MSI mediante una metodología sensible y validada y/o pérdida de las proteínas reparadoras en el tejido normal mediante IHC (Suerink, Ripperger, et al., 2018; Wimmer et al., 2014). Sin embargo, las técnicas de *gMSI* y *evMSI* presentan las complicaciones ya mencionadas y la IHC en tejido normal, como puede ser la piel, no se recomienda en niños sanos al tratarse de un procedimiento muy invasivo.

Nuestra aproximación *hs-MSI* ha demostrado tener una sensibilidad y especificidad del 100% a la hora detectar los casos CMMRD y ofrece la posibilidad de obtener resultados en el plazo de una semana, por lo que podría aplicarse después de un análisis no concluyente de los genes MMR, o incluso realizarse en paralelo para ahorrar tiempo de diagnóstico, o como herramienta pre-test en los casos de pacientes sanos. Recientemente, también se ha propuesto como posible herramienta pre-test el ensayo *in vitro* de actividad reparadora (Shuen et al., 2019). Este ensayo presenta resultados similares a la *hs-MSI* pero supone una mayor complicación técnica y un mayor tiempo hasta la obtención de resultados que la determinación de la *hs-MSI*.

Por estas razones, proponemos un nuevo algoritmo diagnóstico para CMMRD en el que se incluya el análisis de *hs-MSI* en los casos afectos de cáncer (y que cumplen criterios de sospecha) pero sin mutaciones patogénicas identificadas, y como herramienta pre-test en los pacientes sanos con fenotipo sugestivo de CMMRD **(Figura 19)**.

Por otro lado, el resultado obtenido con nuestro panel *hs-MSI* en una muestra de mucosa bucal de un individuo con CMMRD, que presenta niveles de inestabilidad similares a los detectados en sangre, sugiere la posibilidad de realizar, en el futuro, el pre-test de forma mínimamente invasiva, que además representaría una solución para los pacientes linfopénicos o que han pasado por un trasplante alogénico de médula ósea.

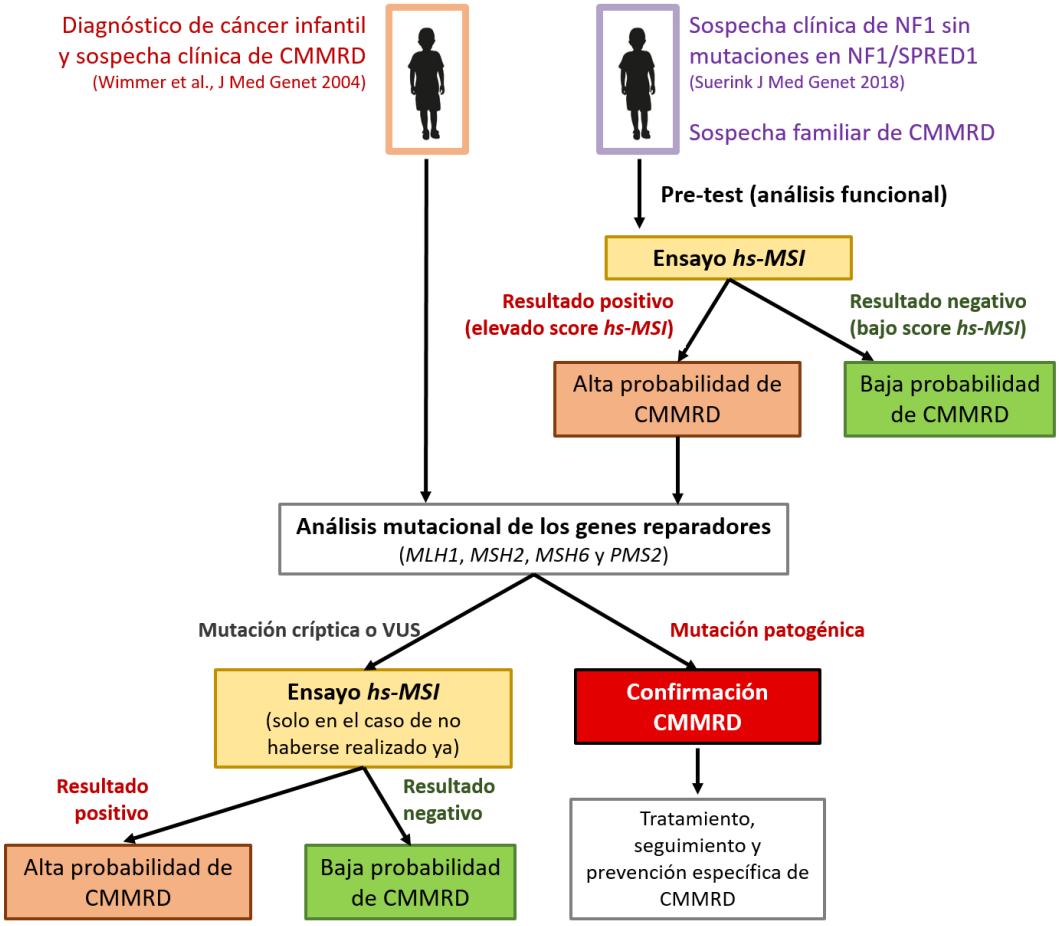


Figura 19. Propuesta de algoritmo diagnóstico para el síndrome CMMRD.

2.2 Futuras perspectivas

2.2.1 Propuestas para la mejora del rendimiento del ensayo

Uno de los principales retos a la hora de aplicar la NGS al diagnóstico clínico es mantener la rentabilidad de la técnica sin perder la calidad del proceso (Feliubadalo et al., 2013). En el caso de nuestro panel *hs-MSI*, la mitad del coste por muestra corresponde al precio de secuenciación. Esto se debe a la elevada cobertura de lecturas a la que se secuencian las muestras (20000x antes de la deduplicación). En la aproximación propuesta

por Gallon y colaboradores, en cambio, se secuencia a una cobertura mucho menor, lo que permite reducir drásticamente el precio por muestra **(Figura 18)** (R. Gallon et al., 2019).

Por otro lado, la utilización en nuestro caso de paneles NGS *custom* de casas comerciales encarece el precio por muestra. Los *kits* comerciales tienen un tamaño mínimo obligatorio que incrementa el precio si el panel diseñado es de tamaño pequeño. Así, las estrategias que diseñan las sondas de forma individual, como es el caso de las smMIPs, y siguen un protocolo casero de captura y preparación de la librería de DNA, tienen un coste menor (Carlson et al., 2015; Richard Gallon et al., 2018; R. Gallon et al., 2019; Hiatt et al., 2013; Schmitt et al., 2015; Waalkes et al., 2018).

Para mejorar el rendimiento del ensayo *hs-MSI*, se ha realizado una selección de los marcadores más informativos de nuestro panel basada en el análisis de curvas Roc específicas para cada microsatélite (AUC>0,70) y descartado aquellos que dieron positivo para la MSI en alguna de las muestras control. Siguiendo estos criterios, 32 de los 186 microsatélites serían suficientes para discriminar las muestras CMMRD de los controles sanos o de otros síndromes con solapamiento fenotípico (NF1 y el síndrome de Legius) (resultados no publicados) **(Figura 20)**. Esto abre la puerta a la posibilidad de reducir el tamaño de nuestro panel *hs-MSI* sin que se resienta su sensibilidad y especificidad. De forma similar, mediante una simulación en la que se disminuyó controladamente el número de lecturas analizadas por muestra, se estimó que una cobertura menor, de 3620x antes de deduplicar y 620x después de la deduplicación, sería suficiente para detectar el componente de inestabilidad en los casos CMMRD sin que esto redujera la precisión del método. Secuenciar a esta cobertura nos permitiría abaratar los costes de secuenciación y, en consecuencia, el coste por muestra.

Para comprobar la viabilidad de esta estrategia de optimización, será necesario el diseño de un nuevo panel *hs-MSI* aplicando los nuevos criterios de selección de marcadores y cobertura de las lecturas y validarlo, primero, en el mismo grupo de muestras que ya

hemos analizado para comprobar que la técnica mantiene su precisión y, segundo, en un nuevo grupo de muestras independientes para validar la selección de marcadores.

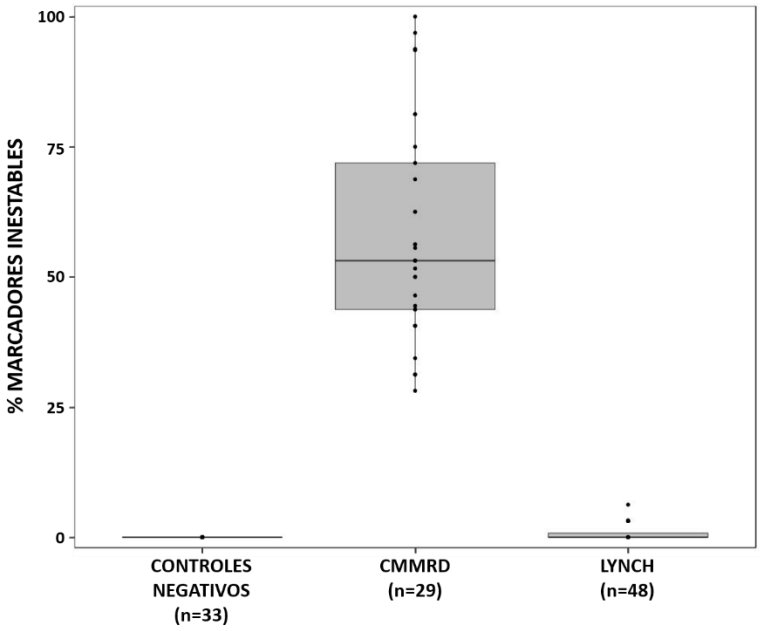


Figura 20. Simulación de los porcentajes de MSI que presentarían las muestras CMMRD y SL al analizarse con los 32 microsatélites seleccionados (AUC > 0,70) (resultados no publicados).

2.2.2 Detección de MSI en tejido normal de individuos con síndrome de Lynch

A pesar de la robustez demostrada para identificar a los individuos con CMMRD, el panel *hs-MSI* ha demostrado no tener la suficiente sensibilidad para discriminar a los pacientes con SL de los individuos control. Esto contrasta con lo reportado en varios trabajos previos a éste. Como se ha comentado en la Introducción (apartado 2.1.6), Alazzouzi y colaboradores encontraban mediante *clonal sequencing* del microsatélite BAT26 una frecuencia media de alelos inestables del 5,6% en el DNA de sangre periférica de individuos con SL. Para calcular esta inestabilidad, contaban el número de clones con alelos inestables (menos de 21 repeticiones) respecto al número total de clones (con alelos estables e inestables) (Alazzouzi et al., 2005). Del mismo modo, otros grupos observaron diferencias significativas entre la frecuencia de los alelos inestables que presentaban los individuos con SL y el resto de controles mediante la técnica de *small pool-PCR*, que analizaba tres microsatélites dinucleótidos (D2S123, D5S346 y D17S518). Para determinar la inestabilidad,

contaban también el número de alelos inestables respecto al total de alelos (M. I. Coolbaugh-Murphy et al., 2010; Hu et al., 2011). Dicha técnica presentaba un límite de detección del 3% para los alelos inestables (M. Coolbaugh-Murphy et al., 2004).

En el panel hs-MSI, el límite de detección es diferente para cada microsatélite en función de la tasa de error basal de secuenciación y utilizamos una *baseline* (generada a partir de un grupo control de individuos sanos sin mutaciones MMR) para establecerlo. El valor de *hs-MSI* representa el porcentaje de marcadores microsatélite que han dado inestabilidad en un individuo, y un microsatélite se considera inestable si la suma de las frecuencias de todos los alelos diferentes al *wild-type* (1 - frecuencia del alelo *wild-type*) en el paciente excede la media más 3 desviaciones estándar de este mismo valor en la *baseline***.** El cálculo de la inestabilidad, por tanto, es diferente entre nuestro estudio y los otros y los porcentajes de inestabilidad reportados no son comparables; además, tampoco hemos estudiado los mismos marcadores utilizados en los estudios anteriores. Tanto BAT26 como D2S123, D5S346 y D17S518 fueron incluidos en el diseño original del panel *hs-MSI*, pero, a excepción de D2S123, la poca cobertura que se obtuvo en estas regiones imposibilitó su análisis **(Tabla 15)**.

Con el objetivo de mejorar la sensibilidad de nuestro panel *hs-MSI*, se pretende mejorar el tipo de sondas con la que analizamos los microsatélites. Recientemente, se ha reportado que sondas que utilizan adaptadores formados por secuencias P5 y P7 emparejadas mediante el mismo índice y un único *molecular barcode*, llamadas como "*dual-matched index adapters*", reducen los errores producidos por la lectura cruzada de los índices de las muestras y permiten el análisis de variantes raras con una frecuencia inferior al 1% (MacConaill et al., 2018). También existen sondas cuyos adaptadores ligan la doble cadena de DNA con el mismo identificador mediante un proceso llamado *Duplex Sequencing* (o *DupSeq*). Esto permite identificar en todo momento las dos cadenas de la misma doble cadena madre y, al hacer la secuencia *consensus*, es necesario que la mutación se encuentre

en las dos cadenas para darla como verdadera. Así, la tasa de error teórica con esta tecnología bajaría al <10^{-9} (Salk et al., 2018; Schmitt et al., 2012).

Tabla 15. Comparativa de los diferentes trabajos donde se analiza la MSI en sangre periférica de individuos con SL.

	Alazzouzi et al. ***Hum Mol Genet 2005***	***Coolbaugh-Murphy et al.*** ***Hum Mutat 2010***	***Hu el al.*** ***Ann Clin Lab Sci 2011***	***González-Acosta et al.*** ***J Med Genet 2019***
Técnica	Clonal sequecing	*small-pool* PCR	*small-pool* PCR	Panel NGS personalizado
Microsatélites analizados	BAT26	D2S123, D5S346 y D17S518	D2S123, D5S346 y D17S518	186 microsatélites
Límite de detección	NA	3%	3%	Diferente para cada microsatélite
Promedio MSI en SL	5,6% (rango: 3,53% - 7,09%)	11,8% (rango: 4% - 24%)	4,6% (rango: 0%-5%)	0,85%* (rango: 0%-3,33%)
Promedio MSI en controles	0,00%	1,8% (rango: 0% - 4,3%).	2% (rango: 0,4%-3,7%)	1,1%* (rango: 0,00-3,89%)

* Los porcentajes representan el número de marcadores inestables de un total de 186 analizados. En el resto de estudios representan proporción de alelos inestables.

Aparte del tipo de sonda utilizada para preparar las muestras, existen métodos de enriquecimiento de los alelos mutantes sobre los alelos *wild-type* que también permitirían incrementar la sensibilidad para la detección de MSI. Tanto la *probe clamping primer extension-PCR* (PCPE-PCR) (Sun et al., 2006) como la E-*ice*-COLD PCR (How-Kit et al., 2018) consisten en modificaciones de una PCR convencional para enriquecer la muestra en alelos mutantes. Ambas técnicas van dirigidas al análisis de la MSI y el primer paso consiste en bloquear la amplificación del alelo *wild-type*. Para ello, se diseña una sonda complementaria a la longitud natural del microsatélite y, cuando se amplifica la región por los métodos convencionales, aquellas secuencias que contengan el alelo *wild-type* no se podrán elongar debido a la hibridación con la sonda bloqueadora. A continuación, se pueden capturar específicamente los fragmentos amplificados mediante la afinidad biotina-streptavidina, en caso de tratarse de la PCPE-PCR, y analizarse por los métodos convencionales para detectar la MSI o analizar directamente el producto de PCR si se trata de la E-*ice*-COLD PCR. Aunque ambos métodos incrementan la sensibilidad en la detección de la MSI, ninguno de ellos evita los artefactos de la PCR.

Por el contrario, la metodología llamada NaMe-PrO, del inglés "*nuclease-assisted minor-allele enrichment with probe-overlap*", permite la eliminación de los alelos *wild-type* sin pasar por la amplificación (Ladas et al., 2018). Básicamente, se diseña una sonda complementaria al alelo *wild-type* del microsatélite de interés y se hibrida con ella el DNA. A continuación, se utiliza una nucleasa específica de doble cadena para digerir todos aquellos homopolímeros que hayan hibridado perfectamente con la sonda. Los alelos inestables, en cambio, al no coincidir del todo con la sonda, crean una estructura de bucle al hibridar con ella y esto impide la digestión. Además, esta metodología permite enriquecer cientos de dianas a la vez.

Por último, cabe recordar que nuestro panel *hs-MSI* se ha testado en DNA de sangre periférica. En los últimos años se ha reportado la existencia de criptas deficientes en reparación, y por tanto MSI, en la mucosa colónica normal de los individuos con SL (Kloor et al., 2012; Pai et al., 2018; Staffa et al., 2015) y se ha estimado que éstas se podrían encontrar en el 1% de las biopsias de colon de al menos 1mm^2 (Kloor et al., 2012). Sería interesante analizar la MSI de este tipo de muestras con nuestra metodología, ya que precisamente es uno de los tejidos diana de la enfermedad, para acabar de esclarecer el componente MSI germinal de los individuos con SL.

2.2.3 Usos alternativos del panel *hs-MSI*

En los últimos años, las aproximaciones basadas en el análisis de DNA fecal han empezado a coger fuerza en el diagnóstico no invasivo del cáncer de colon (Robertson & Imperiale, 2015). La mayoría de estudios se basan en la detección de alteraciones en genes o marcadores concretos como *KRAS* o *APC* (Puig et al., 2000; Robertson & Imperiale, 2015; Sidransky et al., 1992; Traverso, Shuber, Levin, et al., 2002), pero también se ha reportado que el DNA fecal de los pacientes con tumores esporádicos inestables presenta una MSI que correlaciona con el tumor (Lim et al., 2006; Traverso, Shuber, Olsson, et al., 2002). Por otro lado, estudios recientes han demostrado la presencia de niveles bajos de MSI en el tejido endometrial normal de mujeres con SL (Nieminen et al., 2009) y en muestras no invasivas

como los aspirados endometriales, dónde los resultados de MSI correlacionaban a la perfección con la presencia o no de cáncer de endometrio (Bats et al., 2014).

El panel *hs-MSI* podría adaptarse para su uso como herramienta de seguimiento no invasiva para el cáncer colorrectal o ginecológico, enfocado sobre todo a la detección de cáncer en pacientes asintomáticos. En relación al cáncer ginecológico, nuestra aproximación podría ser útil tanto para el cáncer de endometrio como para el de cérvix u ovario, ya que las células exfoliadas que se encuentran en el aspirado endometrial pueden contener DNA tumoral de cualquiera de los 3 tipos (Kinde et al., 2013; Maritschnegg et al., 2015). En análisis preliminares realizados en nuestro laboratorio se ha detectado niveles elevados de hs-MSI en una muestra de aspirado endometrial de una paciente a la que se le había diagnosticado de cáncer de endometrio, lo que apoya esta hipótesis.

La detección de una elevada MSI en sangre periférica indica una elevada probabilidad de CMMRD y, por ende, sugiere la patogenicidad de las variantes MMR que porta el individuo. Nuestra aproximación *hs-MSI*, por tanto, también podría utilizarse como una evidencia a favor de la patogenicidad de las variantes MMR identificadas. De hecho, las guías de InSIGHT para la clasificación de variantes contemplan, actualmente, el criterio de co-ocurrencia en *trans* con una variante patogénica en un individuo clínicamente confirmado CMMRD como una evidencia de patogenicidad de la variante (Versión 2.4, https://www.insight-group.org/criteria/). Este sería el caso de las VUS en *MSH6* c.1316A>G (p.Asp439Gly) y c.2980T>A (p.Tyr994Asn), identificadas junto a otra variante patogénica en los pacientes CMMRD-07 y CMMRD-09, respectivamente (González-Acosta *et al.,* 2019 - Artículo 3). Sin embargo, es necesario ser cauteloso al hacer este tipo de asunciones porque variantes en *cis*, no identificadas aún, podrían ser las responsables del fenotipo. Es más, existen ciertas variantes deletéreas, llamadas hipomórficas, que retienen parte de la expresión y la actividad, por lo que sólo parecen asociadas a patogenicidad al encontrarse en homocigosis o junto a otra variante patogénica en *trans*. Es relevante destacar que sus portadores suelen presentar un fenotipo intermedio entre CMMRD y SL (Bougeard et al.,

2014; L. Li et al., 2015), posiblemente por la baja expresividad de la mutación en heterocigosis.

El elevado MSI score obtenido en el ensayo *hs-MSI* para la muestra DNA E, individuo portador de las variantes de *MLH1* c.62C>A (p. Ala21Glu) y c.2146G>A (p.Val716Met) y en el que se ha descartado la presencia de otras variantes (R. Gallon et al., 2019), genera controversia acerca de la clasificación de la variante c.2146G>A. Esta variante está reportada en las bases de datos como una variante neutra Clase 1 y su clasificación se debe al cálculo multifactorial de probabilidad realizado con datos de familias con sospecha de SL (probabilidad de patogenicidad <0.001); sin embargo, se ha reportado que los niveles de expresión y su actividad reparadora están ligeramente disminuidas en múltiples sistemas heterólogos (http://www.insight-database.org/classifications). Es más, en esta misma tesis se ha evaluado a nivel funcional y presentó una actividad reparadora de 79,3%±9,4 y una expresión intermedia de 60,8%±9,2 (González-Acosta *et al.,* - Artículo 2). Por otro lado, la variante ya se había reportado en un caso anterior con fenotipo CMMRD (Marcos et al., 2006). Curiosamente, el paciente DNA E presenta un valor de hs-MSI inferior al otro paciente bialélico de *MLH1* de nuestra serie (CMMRD-03) (25,6% vs 59,22%, respectivamente) y que es portador en homocigosis una variante patogénica que afecta el dominio ATPasa de la proteína. Esto corrobora lo ya observado en otro estudio donde, mediante el análisis por PCR normal del microsatélite tetranucleótico D17S1307, analizaron la MSI que presentaba el tejido normal de un individuo CMMRD homocigoto para la variante hipomórfica c.2002A>G en el gen *PMS2*, siendo ésta muy inferior a la de los portadores de variantes truncantes (L. Li et al., 2015).

Aunque sería necesario evaluar otras variantes hipomórficas y comprobar que estos bajos niveles de inestabilidad son característicos de ellas, nuestra aproximación *hs-MSI* parece tener capacidad para detectar variantes hipomórficas en *MLH1*, lo que sería particularmente útil para la evaluación de casos atípicos de SL, con una edad de debut del cáncer extremadamente joven, y así descartar la posible presencia de variantes

hipomórficas en *trans* como responsables del fenotipo del paciente (Bougeard et al., 2014). Actualmente se ha establecido una colaboración con el Dr. William Foulkes, de Canadá, para estudiar mediante nuestro análisis *hs-MSI* los porcentajes de inestabilidad que presenta la variante hipomórfica de *MSH6* c.10C>T, detectada como fundadora en la población inuit (Castellsague et al., 2015).

CONCLUSIONES

1) El análisis exhaustivo de las variantes en genes MMR mediante un algoritmo de clasificación que combina el cálculo multifactorial de probabilidad con la frecuencia poblacional, las predicciones *in silico* y los ensayos funcionales a nivel de RNA y proteína, ha permitido reclasificar a Clase 5 (patogénica) o Clase 1 (neutra) el 89% de las variantes estudiadas en esta tesis doctoral, lo que representa el 59% de las variantes identificadas en nuestra serie de individuos Lynch-like. La clasificación de estas variantes ha permitido mejorar el diagnóstico molecular, el consejo genético y el manejo de los pacientes portadores y sus familiares.

2) El ensayo de actividad reparadora presentado en esta tesis doctoral ha sido optimizado a nivel de reactivos y procedimientos, así como validado a nivel analítico, demostrando robustez y reproducibilidad. Aunque su validez clínica aún está por determinar, los protocolos estandarizados que se han establecido son un primer paso fundamental para su implementación en el diagnóstico.

3) La metodología desarrollada para la detección con alta sensibilidad de la inestabilidad de microsatélites en sangre periférica de pacientes CMMRD mejora las estrategias existentes hasta el momento para la detección de la inestabilidad basal asociada a CMMRD. Nuestro enfoque podría ser útil como herramienta de preselección para el diagnóstico de CMMRD, especialmente en los casos con fenotipo sugerente y ausencia de mutaciones patogénicas identificadas en los genes reparadores.

BIBLIOGRAFÍA

A

Aceto, G. M., De Lellis, L., Catalano, T., Veschi, S., Radice, P., Di Iorio, A., Mariani-Costantini, R., Cama, A., & Curia, M. C. (2009). Nonfluorescent denaturing HPLC-based primer-extension method for allele-specific expression: application to analysis of mismatch repair genes. *Clin Chem, 55*(9), 1711-1718. doi:10.1373/clinchem.2009.126300

Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., Kirfel, J., Tafazzoli, A., Horpaopan, S., Uhlhaas, S., Stienen, D., Friedrichs, N., Altmuller, J., Laner, A., Holzapfel, S., Peters, S., Kayser, K., Thiele, H., Holinski-Feder, E., Marra, G., Kristiansen, G., Nothen, M. M., Buttner, R., Moslein, G., Betz, R. C., Brieger, A., Lifton, R. P., & Aretz, S. (2016). Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *Am J Hum Genet, 99*(2), 337-351.

Adar, T., Rodgers, L. H., Shannon, K. M., Yoshida, M., Ma, T., Mattia, A., Lauwers, G. Y., Iafrate, A. J., & Chung, D. C. (2017). A tailored approach to BRAF and MLH1 methylation testing in a universal screening program for Lynch syndrome. *Mod Pathol, 30*(3), 440-447. doi:10.1038/modpathol.2016.211

Ahadova, A., Gallon, R., Gebert, J., Ballhausen, A., Endris, V., Kirchner, M., Stenzinger, A., Burn, J., von Knebel Doeberitz, M., Blaker, H., & Kloor, M. (2018). Three molecular pathways model colorectal carcinogenesis in Lynch syndrome. *Int J Cancer, 143*(1), 139-150. doi:10.1002/ijc.31300

Ahn, D. H., Rho, J. H., Tchah, H., & Jeon, I. S. (2016). Early onset of colorectal cancer in a 13-year-old girl with Lynch syndrome. *Korean J Pediatr, 59*(1), 40-42. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=26893603

Akiyama, T. (2000). Wnt/beta-catenin signaling. *Cytokine Growth Factor Rev, 11*(4), 273-282.

Alazzouzi, H., Domingo, E., Gonzalez, S., Blanco, I., Armengol, M., Espin, E., Plaja, A., Schwartz, S., Capella, G., & Schwartz, S., Jr. (2005). Low levels of microsatellite instability characterize MLH1 and MSH2 HNPCC carriers before tumor diagnosis. *Hum Mol Genet, 14*(2), 235-239.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjord, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jager, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., Lopez-Otin, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdes-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon,

A., Waddell, N., Yates, L. R., Australian Pancreatic Cancer Genome, I., Consortium, I. B. C., Consortium, I. M.-S., PedBrain, I., Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J., & Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature, 500*(7463), 415-421. doi:10.1038/nature12477

Altieri, F., Grillo, C., Maceroni, M., & Chichiarelli, S. (2008). DNA damage and repair: from molecular mechanisms to health implications. *Antioxid Redox Signal, 10*(5), 891-937. doi:10.1089/ars.2007.1830

Amendola, L. M., Jarvik, G. P., Leo, M. C., McLaughlin, H. M., Akkari, Y., Amaral, M. D., Berg, J. S., Biswas, S., Bowling, K. M., Conlin, L. K., Cooper, G. M., Dorschner, M. O., Dulik, M. C., Ghazani, A. A., Ghosh, R., Green, R. C., Hart, R., Horton, C., Johnston, J. J., Lebo, M. S., Milosavljevic, A., Ou, J., Pak, C. M., Patel, R. Y., Punj, S., Richards, C. S., Salama, J., Strande, N. T., Yang, Y., Plon, S. E., Biesecker, L. G., & Rehm, H. L. (2016). Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet, 98*(6), 1067-1076.

Aronson, M., Gallinger, S., Cohen, Z., Cohen, S., Dvir, R., Elhasid, R., Baris, H. N., Kariv, R., Druker, H., Chan, H., Ling, S. C., Kortan, P., Holter, S., Semotiuk, K., Malkin, D., Farah, R., Sayad, A., Heald, B., Kalady, M. F., Penney, L. S., Rideout, A. L., Rashid, M., Hasadsri, L., Pichurin, P., Riegert-Johnson, D., Campbell, B., Bakry, D., Al-Rimawi, H., Alharbi, Q. K., Alharbi, M., Shamvil, A., Tabori, U., & Durno, C. (2016). Gastrointestinal Findings in the Largest Series of Patients With Hereditary Biallelic Mismatch Repair Deficiency Syndrome: Report from the International Consortium. *Am J Gastroenterol, 111*(2), 275-284.

B

Bakry, D., Aronson, M., Durno, C., Rimawi, H., Farah, R., Alharbi, Q. K., Alharbi, M., Shamvil, A., Ben-Shachar, S., Mistry, M., Constantini, S., Dvir, R., Qaddoumi, I., Gallinger, S., Lerner-Ellis, J., Pollett, A., Stephens, D., Kelies, S., Chao, E., Malkin, D., Bouffet, E., Hawkins, C., & Tabori, U. (2014). Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. *Eur J Cancer, 50*(5), 987-996.

Baralle, D., Lucassen, A., & Buratti, E. (2009). Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep, 10*(8), 810-816. doi:10.1038/embor.2009.170

Bats, A. S., Blons, H., Narjoz, C., Le Frere-Belda, M. A., Laurent-Puig, P., & Lecuru, F. (2014). Microsatellite instability analysis in uterine cavity washings to detect endometrial cancer in Lynch syndrome. *Anticancer Res, 34*(6), 3211-3215

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M. Q., James, T., Huw Jones, T. A., Kang, G. D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R., & Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*(7218), 53-59. doi:10.1038/nature07517

Biller, L. H., Syngal, S., & Yurgelun, M. B. (2019). Recent advances in Lynch syndrome. *Fam Cancer, 18*(2), 211-219. doi:10.1007/s10689-018-00117-1

Blount, J., & Prakash, A. (2018). The changing landscape of Lynch syndrome due to PMS2 mutations. *Clin Genet, 94*(1), 61-69. doi:10.1111/cge.13205

Bodas, A., Perez-Segura, P., Maluenda, C., Caldes, T., Olivera, E., & Diaz-Rubio, E. (2008). Lynch syndrome in a 15-year-old boy. *Eur J Pediatr, 167*(10), 1213-1215.

Bodo, S., Colas, C., Buhard, O., Collura, A., Tinat, J., Lavoine, N., Guilloux, A., Chalastanis, A., Lafitte, P., Coulet, F., Buisine, M. P., Ilencikova, D., Ruiz-Ponte, C., Kinzel, M., Grandjouan, S., Brems, H., Lejeune, S., Blanche, H., Wang, Q., Caron, O., Cabaret, O., Svrcek, M., Vidaud, D., Parfait, B., Verloes, A., Knappe, U. J., Soubrier, F., Mortemousque, I., Leis, A., Auclair-Perrossier, J., Frebourg, T., Flejou, J. F., Entz-Werle, N., Leclerc, J., Malka, D., Cohen-Haguenauer, O., Goldberg, Y., Gerdes, A. M., Fedhila, F., Mathieu-Dramard, M., Hamelin, R., Wafaa, B., Gauthier-Villars, M., Bourdeaut, F., Sheridan, E., Vasen, H., Brugieres, L., Wimmer, K., Muleris, M., & Duval, A. (2015). Diagnosis of Constitutional Mismatch Repair-Deficiency Syndrome Based on Microsatellite Instability and Lymphocyte Tolerance to Methylating Agents. *Gastroenterology, 149*(4), 1017-1029 e1013.

Boland, C. R. (2005). Evolution of the nomenclature for the hereditary colorectal cancer syndromes. *Fam Cancer, 4*(3), 211-218. doi:10.1007/s10689-004-4489-x

Boland, C. R., & Lynch, H. T. (2013). The history of Lynch syndrome. *Fam Cancer, 12*(2), 145-157.

Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., & Srivastava, S. (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*(22), 5248-5257.

Borras, E., Pineda, M., Blanco, I., Jewett, E. M., Wang, F., Teule, A., Caldes, T., Urioste, M., Martinez-Bouzas, C., Brunet, J., Balmana, J., Torres, A., Ramon y Cajal, T., Sanz, J., Perez-Cabornero, L., Castellvi-Bel, S., Alonso, A., Lanas, A., Gonzalez, S., Moreno, V., Gruber, S. B., Rosenberg, N. A., Mukherjee, B., Lazaro, C., & Capella, G. (2010). MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. *Cancer Res, 70*(19), 7379-7391.

Borras, E., Pineda, M., Brieger, A., Hinrichsen, I., Gomez, C., Navarro, M., Balmana, J., Ramon y Cajal, T., Torres, A., Brunet, J., Blanco, I., Plotz, G., Lazaro, C., & Capella, G. (2012). Comprehensive functional assessment of MLH1 variants of unknown significance. *Hum Mutat, 33*(11), 1576-1588. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22736432

Borras, E., Pineda, M., Cadinanos, J., Del Valle, J., Brieger, A., Hinrichsen, I., Cabanillas, R., Navarro, M., Brunet, J., Sanjuan, X., Musulen, E., van der Klift, H., Lazaro, C., Plotz, G., Blanco, I., & Capella, G. (2013). Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. *J Med Genet, 50*(8), 552-563.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C., & Standards for Reporting of Diagnostic,

A. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem, 49*(1), 1-6.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Moher, D., Rennie, D., de Vet, H. C., Lijmer, J. G., & Standards for Reporting of Diagnostic, A. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem, 49*(1), 7-18.

Bougeard, G., Olivier-Faivre, L., Baert-Desurmont, S., Tinat, J., Martin, C., Bouvignies, E., Vasseur, S., Huet, F., Couillault, G., Vabres, P., Le Pessot, F., Chapusot, C., Malka, D., Bressac-de Paillerets, B., Tosi, M., & Frebourg, T. (2014). Diversity of the clinical presentation of the MMR gene biallelic mutations. *Fam Cancer, 13*(1), 131-135.

Brea-Fernandez, A. J., Cameselle-Teijeiro, J. M., Alenda, C., Fernandez-Rozadilla, C., Cubiella, J., Clofent, J., Rene, J. M., Anido, U., Mila, M., Balaguer, F., Castells, A., Castellvi-Bel, S., Jover, R., Carracedo, A., & Ruiz-Ponte, C. (2014). High incidence of large deletions in the PMS2 gene in Spanish Lynch syndrome families. *Clin Genet, 85*(6), 583-588.

Brieger, A., Plotz, G., Raedle, J., Weber, N., Baum, W., Caspary, W. F., Zeuzem, S., & Trojan, J. (2005). Characterization of the nuclear import of human MutLalpha. *Mol Carcinog, 43*(1), 51-58. doi:10.1002/mc.20081

Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., & et al. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature, 368*(6468), 258-261. doi:10.1038/368258a0

Buchanan, D. D., Rosty, C., Clendenning, M., Spurdle, A. B., & Win, A. K. (2014). Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). *The application of clinical genetics, 7*, 183-193. doi:10.2147/TACG.S48625

Burke, W. (2014). Genetic tests: clinical validity and clinical utility. *Curr Protoc Hum Genet, 81*, 9 15 11-18. doi:10.1002/0471142905.hg0915s81

Burn, J., Gerdes, A. M., Macrae, F., Mecklin, J. P., Moeslein, G., Olschwang, S., Eccles, D., Evans, D. G., Maher, E. R., Bertario, L., Bisgaard, M. L., Dunlop, M. G., Ho, J. W., Hodgson, S. V., Lindblom, A., Lubinski, J., Morrison, P. J., Murday, V., Ramesar, R., Side, L., Scott, R. J., Thomas, H. J., Vasen, H. F., Barker, G., Crawford, G., Elliott, F., Movahedi, M., Pylvanainen, K., Wijnen, J. T., Fodde, R., Lynch, H. T., Mathers, J. C., Bishop, D. T., & Investigators, C. (2011). Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet, 378*(9809),2081-2087. doi:10.1016/S0140-6736(11)61049-0

C

Campbell, C. S., Hombauer, H., Srivatsan, A., Bowen, N., Gries, K., Desai, A., Putnam, C. D., & Kolodner, R. D. (2014). Mlh2 is an accessory factor for DNA mismatch repair in Saccharomyces cerevisiae. *PLoS Genet, 10*(5), e1004327. doi:10.1371/journal.pgen.1004327

Carethers, J. M. (2017). Microsatellite Instability Pathway and EMAST in Colorectal Cancer. *Curr Colorectal Cancer Rep, 13*(1), 73-80. doi:10.1007/s11888-017-0352-y

Carethers, J. M., & Stoffel, E. M. (2015). Lynch syndrome and Lynch syndrome mimics: The growing complex landscape of hereditary colon cancer. *World J Gastroenterol, 21*(31), 9253-9261.

Carlson, K. D., Sudmant, P. H., Press, M. O., Eichler, E. E., Shendure, J., & Queitsch, C. (2015). MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals. *Genome Res, 25*(5), 750-761. doi:10.1101/gr.182212.114

Castellsague, E., Gonzalez, S., Guino, E., Stevens, K. N., Borras, E., Raymond, V. M., Lazaro, C., Blanco, I., Gruber, S. B., & Capella, G. (2010). Allele-specific expression of APC in adenomatous polyposis families. *Gastroenterology, 139*(2), 439-447, 447 e431. doi:10.1053/j.gastro.2010.04.047

Castellsague, E., Liu, J., Volenik, A., Giroux, S., Gagne, R., Maranda, B., Roussel-Jobin, A., Latreille, J., Laframboise, R., Palma, L., Kasprzak, L., Marcus, V. A., Breguet, M., Nolet, S., El-Haffaf, Z., Australie, K., Gologan, A., Aleynikova, O., Oros-Klein, K., Greenwood, C., Mes-Masson, A. M., Provencher, D., Tischkowitz, M., Chong, G., Rousseau, F., & Foulkes, W. D. (2015). Characterization of a novel founder MSH6 mutation causing Lynch syndrome in the French Canadian population. *Clin Genet, 87*(6), 536-542. doi:10.1111/cge.12526

Castillejo, A., Vargas, G., Castillejo, M. I., Navarro, M., Barbera, V. M., Gonzalez, S., Hernandez-Illan, E., Brunet, J., Ramon y Cajal, T., Balmana, J., Oltra, S., Iglesias, S., Velasco, A., Solanes, A., Campos, O., Sanchez Heras, A. B., Gallego, J., Carrasco, E., Gonzalez Juan, D., Segura, A., Chirivella, I., Juan, M. J., Tena, I., Lazaro, C., Blanco, I., Pineda, M., Capella, G., & Soto, J. L. (2014). Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. *Eur J Cancer, 50*(13), 2241-2250. doi:10.1016/j.ejca.2014.05.022

Cenin, D. R., Naber, S. K., Lansdorp-Vogelaar, I., Jenkins, M. A., Buchanan, D. D., Preen, D. B., Ee, H. C., & O'Leary, P. (2018). Costs and outcomes of Lynch syndrome screening in the Australian colorectal cancer population. *J Gastroenterol Hepatol, 33*(10), 1737-1744. doi:10.1111/jgh.14154

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*(5), 490-495.

Colas, C., Brugières, L., & Wimmer, K. (2018). Constitutional Mismatch Repair Deficiency. In L. Valle, S. B. Gruber, & G. Capellá (Eds.), *Hereditary Colorectal Cancer: Genetic Basis and Clinical Implications* (pp. 43-54). Cham: Springer International Publishing.

Coolbaugh-Murphy, M., Maleki, A., Ramagli, L., Frazier, M., Lichtiger, B., Monckton, D. G., Siciliano, M. J., & Brown, B. W. (2004). Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. *Genomics, 84*(2), 419-430.

Coolbaugh-Murphy, M. I., Xu, J., Ramagli, L. S., Brown, B. W., & Siciliano, M. J. (2005). Microsatellite instability (MSI) increases with age in normal somatic cells. *Mech Ageing Dev, 126*(10), 1051-1059. doi:10.1016/j.mad.2005.06.005

Coolbaugh-Murphy, M. I., Xu, J. P., Ramagli, L. S., Ramagli, B. C., Brown, B. W., Lynch, P. M., Hamilton, S. R., Frazier, M. L., & Siciliano, M. J. (2010). Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. *Hum Mutat, 31*(3), 317-324.

Couch, F. J., Rasmussen, L. J., Hofstra, R., Monteiro, A. N., Greenblatt, M. S., & de Wind, N. (2008). Assessment of functional effects of unclassified genetic variants. *Hum Mutat, 29*(11), 1314-1326.

Chang, D. K., Ricciardiello, L., Goel, A., Chang, C. L., & Boland, C. R. (2000). Steady-state regulation of the human DNA mismatch repair system. *J Biol Chem, 275*(37), 29178. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10979986

Chen, S., Wang, W., Lee, S., Nafa, K., Lee, J., Romans, K., Watson, P., Gruber, S. B., Euhus, D., Kinzler, K. W., Jass, J., Gallinger, S., Lindor, N. M., Casey, G., Ellis, N., Giardiello, F. M., Offit, K., Parmigiani, G., & Colon Cancer Family, R. (2006). Prediction of germline mutations and cancer risk in the Lynch syndrome. *Jama, 296*(12), 1479-1487. doi:10.1001/jama.296.12.1479

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One, 7*(10), e46688. doi:10.1371/journal.pone.0046688

D

Damaso, E., Castillejo, A., Arias, M. D. M., Canet-Hermida, J., Navarro, M., Del Valle, J., Campos, O., Fernandez, A., Marin, F., Turchetti, D., Garcia-Diaz, J. D., Lazaro, C., Genuardi, M., Rueda, D., Alonso, A., Soto, J. L., Hitchins, M., Pineda, M., & Capella, G. (2018). Primary constitutional MLH1 epimutations: a focal epigenetic event. *Br J Cancer, 119*(8), 978-987. doi:10.1038/s41416-018-0019-8

de Jong, L. C., Cree, S., Lattimore, V., Wiggins, G. A. R., Spurdle, A. B., kConFab, I., Miller, A., Kennedy, M. A., & Walker, L. C. (2017). Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res, 19*(1), 127. doi:10.1186/s13058-017-0919-1

de la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. *Nat Rev Cancer, 4*(10), 769-780. doi:10.1038/nrc1453

de Voer, R. M., Geurts van Kessel, A., Weren, R. D., Ligtenberg, M. J., Smeets, D., Fu, L., Vreede, L., Kamping, E. J., Verwiel, E. T., Hahn, M. M., Ariaans, M., Spruijt, L., van Essen, T., Houge, G., Schackert, H. K., Sheng, J. Q., Venselaar, H., van Ravenswaaij-Arts, C. M., van Krieken, J. H., Hoogerbrugge, N., & Kuiper, R. P. (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. *Gastroenterology, 145*(3), 544-547. doi:10.1053/j.gastro.2013.06.001

Deng, G., Bell, I., Crawley, S., Gum, J., Terdiman, J. P., Allen, B. A., Truta, B., Sleisenger, M. H., & Kim, Y. S. (2004). BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res, 10*(1 Pt 1), 191-195.

Drost, M., Koppejan, H., & de Wind, N. (2013). Inactivation of DNA mismatch repair by variants of uncertain significance in the PMS2 gene. *Hum Mutat, 34*(11), 1477-1480.

Drost, M., Tiersma, Y., Thompson, B. A., Frederiksen, J. H., Keijzers, G., Glubb, D., Kathe, S., Osinga, J., Westers, H., Pappas, L., Boucher, K. M., Molenkamp, S., Zonneveld, J. B., van Asperen, C. J., Goldgar, D. E., Wallace, S. S., Sijmons, R. H., Spurdle, A. B., Rasmussen, L. J., Greenblatt, M. S., de Wind, N., & Tavtigian, S. V. (2018). A functional assay-based procedure to classify mismatch repair gene variants in Lynch syndrome. *Genet Med*. Epub 2018 Dec 3. doi: 10.1038/s41436-018-0372-2.

Drost, M., Zonneveld, J., van Dijk, L., Morreau, H., Tops, C. M., Vasen, H. F., Wijnen, J. T., & de Wind, N. (2010). A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Hum Mutat, 31*(3), 247-253.

Drost, M., Zonneveld, J. B., van Hees, S., Rasmussen, L. J., Hofstra, R. M., & 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*(3), 488-494.

Duraturo, F., Cavallo, A., Liccardo, R., Cudia, B., De Rosa, M., Diana, G., & Izzo, P. (2013). Contribution of large genomic rearrangements in Italian Lynch syndrome patients: characterization of a novel alu-mediated deletion. *Biomed Res Int, 2013*, 219897. doi:10.1155/2013/219897

Durno, C., Boland, C. R., Cohen, S., Dominitz, J. A., Giardiello, F. M., Johnson, D. A., Kaltenbach, T., Levin, T. R., Lieberman, D., Robertson, D. J., & Rex, D. K. (2017). Recommendations on Surveillance and Management of Biallelic Mismatch Repair Deficiency (BMMRD) Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol, 112*(5), 682-690. doi:10.1038/ajg.2017.105

Durno, C. A., Sherman, P. M., Aronson, M., Malkin, D., Hawkins, C., Bakry, D., Bouffet, E., Gallinger, S., Pollett, A., Campbell, B., & Tabori, U. (2015). Phenotypic and genotypic

characterisation of biallelic mismatch repair deficiency (BMMR-D) syndrome. *Eur J Cancer, 51*(8), 977-983.

E

Elsayed, F. A., Kets, C. M., Ruano, D., van den Akker, B., Mensenkamp, A. R., Schrumpf, M., Nielsen, M., Wijnen, J. T., Tops, C. M., Ligtenberg, M. J., Vasen, H. F., Hes, F. J., Morreau, H., & van Wezel, T. (2015). Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. *Eur J Hum Genet, 23*(8), 1080-1084. doi:10.1038/ejhg.2014.242

Etzler, J., Peyrl, A., Zatkova, A., Schildhaus, H. U., Ficek, A., Merkelbach-Bruse, S., Kratz, C. P., Attarbaschi, A., Hainfellner, J. A., Yao, S., Messiaen, L., Slavc, I., & Wimmer, K. (2008). RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. *Hum Mutat, 29*(2), 299-305.

F

Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H., 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*(13), 3697-3703.

Feliubadalo, L., Lopez-Doriga, A., Castellsague, E., del Valle, J., Menendez, M., Tornero, E., Montes, E., Cuesta, R., Gomez, C., Campos, O., Pineda, M., Gonzalez, S., Moreno, V., Brunet, J., Blanco, I., Serra, E., Capella, G., & Lazaro, C. (2013). Next-generation sequencing meets genetic diagnostics: development of a comprehensive workflow for the analysis of BRCA1 and BRCA2 genes. *Eur J Hum Genet, 21*(8), 864-870. doi:10.1038/ejhg.2012.270

Feliubadaló, L., López-Fernández, A., Pineda, M., Díez, O., del Valle, J., Gutiérrez-Enríquez, S., Teulé, A., González, S., Stjepanovic, N., Salinas, M., Capellá, G., Brunet, J., Lázaro, C., Balmaña, J., & Group, o. b. o. t. C. H. C. (2019). Opportunistic testing of BRCA1, BRCA2 and mismatch repair genes improves the yield of phenotype driven hereditary cancer gene panels. *Int J Cancer, 145*(10), 2682-2691. doi:10.1002/ijc.32304

Fernandez-Rozadilla, C., Alvarez-Barona, M., Schamschula, E., Bodo, S., Lopez-Novo, A., Dacal, A., Calviño-Costas, C., Lancho, A., Amigo, J., Bello, X., Cameselle-Teijeiro, J. M., Carracedo, A., Colas, C., Muleris, M., Wimmer, K., & Ruiz-Ponte, C. (2019). Early Colorectal Cancers Provide New Evidence for a Lynch Syndrome-to-CMMRD Phenotypic Continuum. *Cancers, 11*(8), 1081. doi:10.3390/cancers11081081

Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., Janizek, J. D., Huang, X., Starita, L. M., & Shendure, J. (2018). Accurate classification of BRCA1

variants with saturation genome editing. *Nature, 562*(7726), 217-222. doi:10.1038/s41586-018-0461-z

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., & Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell, 75*(5), 1027-1038. doi:10.1016/0092-8674(93)90546-3

Flores-Rozas, H., & Kolodner, R. D. (1998). The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. *Proc Natl Acad Sci U S A, 95*(21), 12404-12409. doi:10.1073/pnas.95.21.12404

Flower, M., Lomeikaite, V., Ciosi, M., Cumming, S., Morales, F., Lo, K., Hensman Moss, D., Jones, L., Holmans, P., Monckton, D. G., Tabrizi, S. J., Investigators, T.-H., & Consortium, O. (2019). MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain, 142*(7), 1876-1886. doi:10.1093/brain/awz115

G

Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frebourg, T., & Tosi, M. (2010). Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol, 653*, 249-257. doi:10.1007/978-1-60761-759-4_15

Gallon, R., Hayes, C., Redford, L., Alhilal, G., O'Brien, O., Waltham, A., Needham, S., Arends, M., Oniscu, A., Alonso, A. M., Laguna, S. M., Sheth, H., Santibanez-Koref, M., Jackson, M. S., & Burn, J. (2018). A molecular inversion probe and sequencing-based microsatellite instability assay for high throughput cancer diagnostics and Lynch syndrome screening. *bioRxiv*, 382754. doi:10.1101/382754

Gallon, R., Muhlegger, B., Wenzel, S. S., Sheth, H., Hayes, C., Aretz, S., Dahan, K., Foulkes, W., Kratz, C. P., Ripperger, T., Azizi, A. A., Feldman, H. B., Chong, A. L., Demirsoy, U., Florkin, B., Imschweiler, T., Januszkiewicz-Lewandowska, D., Lobitz, S., Nathrath, M., Pander, H. J., Perez-Alonso, V., Perne, C., Ragab, I., Rosenbaum, T., Rueda, D., Seidel, M. G., Suerink, M., Taeubner, J., Zimmermann, S. Y., Zschocke, J., Borthwick, G. M., Burn, J., Jackson, M. S., Santibanez-Koref, M., & Wimmer, K. (2019). A sensitive and scalable microsatellite instability assay to diagnose constitutional mismatch repair deficiency by sequencing of peripheral blood leukocytes. *Hum Mutat*.

Gammie, A. E., Erdeniz, N., Beaver, J., Devlin, B., Nanji, A., & Rose, M. D. (2007). Functional characterization of pathogenic human MSH2 missense mutations in Saccharomyces cerevisiae. *Genetics, 177*(2), 707-721. doi:10.1534/genetics.107.071084

Ganster, C., Wernstedt, A., Kehrer-Sawatzki, H., Messiaen, L., Schmidt, K., Rahner, N., Heinimann, K., Fonatsch, C., Zschocke, J., & Wimmer, K. (2010). Functional PMS2 hybrid alleles containing a pseudogene-specific missense variant trace back to a single ancient intrachromosomal recombination event. *Hum Mutat, 31*(5), 552-560.

Gassman, N. R., Clodfelter, J. E., McCauley, A. K., Bonin, K., Salsbury, F. R., Jr., & Scarpinato, K. D. (2011). Cooperative nuclear localization sequences lend a novel role to the N-terminal region of MSH6. *PLoS One, 6*(3), e17907. doi:10.1371/journal.pone.0017907

Gausachs, M., Mur, P., Corral, J., Pineda, M., Gonzalez, S., Benito, L., Menendez, M., Espinas, J. A., Brunet, J., Iniesta, M. D., Gruber, S. B., Lazaro, C., Blanco, I., & Capella, G. (2012). MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. *Eur J Hum Genet, 20*(7), 762-768. doi:10.1038/ejhg.2011.277

Giannakis, M., Mu, X. J., Shukla, S. A., Qian, Z. R., Cohen, O., Nishihara, R., Bahl, S., Cao, Y., Amin-Mansour, A., Yamauchi, M., Sukawa, Y., Stewart, C., Rosenberg, M., Mima, K., Inamura, K., Nosho, K., Nowak, J. A., Lawrence, M. S., Giovannucci, E. L., Chan, A. T., Ng, K., Meyerhardt, J. A., Van Allen, E. M., Getz, G., Gabriel, S. B., Lander, E. S., Wu, C. J., Fuchs, C. S., Ogino, S., & Garraway, L. A. (2016). Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep, 17*(4), 1206.

Giardiello, F. M., Allen, J. I., Axilbund, J. E., Boland, C. R., Burke, C. A., Burt, R. W., Church, J. M., Dominitz, J. A., Johnson, D. A., Kaltenbach, T., Levin, T. R., Lieberman, D. A., Robertson, D. J., Syngal, S., & Rex, D. K. (2014). Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task Force on colorectal cancer. *Am J Gastroenterol, 109*(8), 1159-1179.

Giunti, L., Cetica, V., Ricci, U., Giglio, S., Sardi, I., Paglierani, M., Andreucci, E., Sanzo, M., Forni, M., Buccoliero, A. M., Genitori, L., & Genuardi, M. (2009). Type A microsatellite instability in pediatric gliomas as an indicator of Turcot syndrome. *Eur J Hum Genet, 17*(7), 919-927. doi:10.1038/ejhg.2008.271

Goldgar, D. E., Easton, D. F., Byrnes, G. B., Spurdle, A. B., Iversen, E. S., Greenblatt, M. S., & Group, I. U. G. V. W. (2008). Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum Mutat, 29*(11), 1265-1272. doi:10.1002/humu.20897

Goldgar, D. E., Easton, D. F., Deffenbaugh, A. M., Monteiro, A. N., Tavtigian, S. V., Couch, F. J., & 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*(4), 535-544. doi:10.1086/424388

Groothuizen, F. S., & Sixma, T. K. (2016). The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair (Amst), 38*, 14-23. doi:10.1016/j.dnarep.2015.11.012

Gruber, S. B. (2006). New developments in Lynch syndrome (hereditary nonpolyposis colorectal cancer) and mismatch repair gene testing. *Gastroenterology, 130*(2), 577-587. doi:10.1053/j.gastro.2006.01.031

Guarinos, C., Castillejo, A., Barbera, V. M., Perez-Carbonell, L., Sanchez-Heras, A. B., Segura, A., Guillen-Ponce, C., Martinez-Canto, A., Castillejo, M. I., Egoavil, C. M., Jover, R., Paya, A., Alenda, C., & Soto, J. L. (2010). EPCAM germ line deletions as causes of Lynch syndrome in Spanish patients. *J Mol Diagn, 12*(6), 765-770. doi:10.2353/jmoldx.2010.100039

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. *Mol Cell Biol, 18*(11), 6616-6623. doi:10.1128/mcb.18.11.6616

Guindalini, R. S., Win, A. K., Gulden, C., Lindor, N. M., Newcomb, P. A., Haile, R. W., Raymond, V., Stoffel, E., Hall, M., Llor, X., Ukaegbu, C. I., Solomon, I., Weitzel, J., Kalady, M., Blanco, A., Terdiman, J., Shuttlesworth, G. A., Lynch, P. M., Hampel, H., Lynch, H. T., Jenkins, M. A., Olopade, O. I., & Kupfer, S. S. (2015). Mutation spectrum and risk of colorectal cancer in African American families with Lynch syndrome. *Gastroenterology, 149*(6), 1446-1453. doi:10.1053/j.gastro.2015.07.052

H

Hampel, H. (2010). Point: justification for Lynch syndrome screening among all patients with newly diagnosed colorectal cancer. *J Natl Compr Canc Netw, 8*(5), 597-601. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20495086

Hampel, H. (2016). Genetic counseling and cascade genetic testing in Lynch syndrome. *Fam Cancer, 15*(3), 423-427. doi:10.1007/s10689-016-9893-5

Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Clendenning, M., Sotamaa, K., Prior, T., Westman, J. A., Panescu, J., Fix, D., Lockman, J., LaJeunesse, J., Comeras, I., & de la Chapelle, A. (2008). Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J Clin Oncol, 26*(35), 5783-5788. doi:10.1200/JCO.2008.17.5950

Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Nakagawa, H., Sotamaa, K., Prior, T. W., Westman, J., Panescu, J., Fix, D., Lockman, J., Comeras, I., & de la Chapelle, A. (2005). Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med, 352*(18), 1851-1860. doi:10.1056/NEJMoa043146

Hampel, H., Pearlman, R., Beightol, M., Zhao, W., Jones, D., Frankel, W. L., Goodfellow, P. J., Yilmaz, A., Miller, K., Bacher, J., Jacobson, A., Paskett, E., Shields, P. G., Goldberg, R. M., de la Chapelle, A., Shirts, B. H., Pritchard, C. C., & Ohio Colorectal Cancer Prevention Initiative Study, G. (2018). Assessment of Tumor Sequencing as a Replacement for Lynch Syndrome Screening and Current Molecular Tests for Patients With Colorectal Cancer. *JAMA Oncol, 4*(6), 806-813. doi:10.1001/jamaoncol.2018.0104

Haraldsdottir, S., Rafnar, T., Frankel, W. L., Einarsdottir, S., Sigurdsson, A., Hampel, H., Snaebjornsson, P., Masson, G., Weng, D., Arngrimsson, R., Kehr, B., Yilmaz, A., Haraldsson, S., Sulem, P., Stefansson, T., Shields, P. G., Sigurdsson, F., Bekaii-Saab, T., Moller, P. H., Steinarsdottir, M., Alexiusdottir, K., Hitchins, M., Pritchard, C. C., de la Chapelle, A., Jonasson, J. G., Goldberg, R. M., & Stefansson, K. (2017). Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat Commun, 8*, 14755. doi:10.1038/ncomms14755

Hardt, K., Heick, S. B., Betz, B., Goecke, T., Yazdanparast, H., Kuppers, R., Servan, K., Steinke, V., Rahner, N., Morak, M., Holinski-Feder, E., Engel, C., Moslein, G., Schackert, H. K., von Knebel Doeberitz, M., Pox, C., Hegemann, J. H., & Royer-Pokora, B. (2011). Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. *Fam Cancer, 10*(2), 273-284.

Harfe, B. D., & Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. *Annu Rev Genet, 34*, 359-399. doi:10.1146/annurev.genet.34.1.359

Harrington, J. M., & Kolodner, R. D. (2007). Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs. *Mol Cell Biol, 27*(18), 6546-6554. doi:10.1128/MCB.00855-07

Hause, R. J., Pritchard, C. C., Shendure, J., & Salipante, S. J. (2016). Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med, 22*(11), 1342-1350.

Heinen, C. D., & Rasmussen, L. J. (2012). Determining the functional significance of mismatch repair gene missense variants using biochemical and cellular assays. *Hered Cancer Clin Pract, 10*(1), 9.

Heinen, C. D., Wilson, T., Mazurek, A., Berardini, M., Butz, C., & Fishel, R. (2002). HNPCC mutations in hMSH2 result in reduced hMSH2-hMSH6 molecular switch functions. *Cancer Cell, 1*(5), 469-478.

Hemminki, A., Peltomaki, P., Mecklin, J. P., Jarvinen, H., Salovaara, R., Nystrom-Lahti, M., de la Chapelle, A., & Aaltonen, L. A. (1994). Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nat Genet, 8*(4), 405-410. doi:10.1038/ng1294-405

Hendriks, Y. M., Wagner, A., Morreau, H., Menko, F., Stormorken, A., Quehenberger, F., Sandkuijl, L., Moller, P., Genuardi, M., Van Houwelingen, H., Tops, C., Van Puijenbroek, M., Verkuijlen, P., Kenter, G., Van Mil, A., Meijers-Heijboer, H., Tan, G. B., Breuning, M. H., Fodde, R., Wijnen, J. T., Brocker-Vriends, A. H., & Vasen, H. (2004). Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance. *Gastroenterology, 127*(1), 17-25. doi:10.1053/j.gastro.2004.03.068

Herkert, J. C., Niessen, R. C., Olderode-Berends, M. J., Veenstra-Knol, H. E., Vos, Y. J., van der Klift, H. M., Scheenstra, R., Tops, C. M., Karrenbeld, A., Peters, F. T., Hofstra, R. M., Kleibeuker, J. H., & Sijmons, R. H. (2011). Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines. *Eur J Cancer, 47*(7), 965-982.

Hiatt, J. B., Pritchard, C. C., Salipante, S. J., O'Roak, B. J., & Shendure, J. (2013). Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res, 23*(5), 843-854. doi:10.1101/gr.147686.112

Hinrichsen, I., Brieger, A., Trojan, J., Zeuzem, S., Nilbert, M., & Plotz, G. (2013). Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis. *Clin Cancer Res, 19*(9), 2432-2441.

Hinrichsen, I., Schafer, D., Langer, D., Koger, N., Wittmann, M., Aretz, S., Steinke, V., Holzapfel, S., Trojan, J., Konig, R., Zeuzem, S., Brieger, A., & Plotz, G. (2015). Functional testing strategy for coding genetic variants of unclear significance in MLH1 in Lynch syndrome diagnosis. *Carcinogenesis, 36*(2), 202-211.

Hitchins, M. P., Lin, V. A., Buckle, A., Cheong, K., Halani, N., Ku, S., Kwok, C. T., Packham, D., Suter, C. M., Meagher, A., Stirzaker, C., Clark, S., Hawkins, N. J., & Ward, R. L. (2007). Epigenetic inactivation of a cluster of genes flanking MLH1 in microsatellite-unstable colorectal cancer. *Cancer Res, 67*(19), 9107-9116. doi:10.1158/0008-5472.CAN-07-0869

Hitchins, M. P., & Ward, R. L. (2009). Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J Med Genet, 46*(12), 793-802. doi:10.1136/jmg.2009.068122

Hoskinson, D. C., Dubuc, A. M., & Mason-Suares, H. (2017). The current state of clinical interpretation of sequence variants. *Curr Opin Genet Dev, 42*, 33-39.

Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., Lefol, C., Leone, M., Mazoyer, S., Muller, D., Remenieras, A., Revillion, F., Rouleau, E., Sokolowska, J., Vert, J. P., Lidereau, R., Soubrier, F., Sobol, H., Sevenet, N., Bressac-de Paillerets, B., Hardouin, A., Tosi, M., Sinilnikova, O. M., & Stoppa-Lyonnet, D. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat, 33*(8), 1228-1238. doi:10.1002/humu.22101

How-Kit, A., Daunay, A., Buhard, O., Meiller, C., Sahbatou, M., Collura, A., Duval, A., & Deleuze, J. F. (2018). Major improvement in the detection of microsatellite instability in colorectal cancer using HSP110 T17 E-ice-COLD-PCR. *Hum Mutat, 39*(3), 441-453. doi:10.1002/humu.23379

Howarth, D. R., Lum, S. S., Esquivel, P., Garberoglio, C. A., Senthil, M., & Solomon, N. L. (2015). Initial Results of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer and Lynch Syndrome. *Am Surg, 81*(10), 941-944.

Hu, P., Lee, C. W., Xu, J. P., Simien, C., Fan, C. L., Tam, M., Ramagli, L., Brown, B. W., Lynch, P., Frazier, M. L., Siciliano, M. J., & Coolbaugh-Murphy, M. (2011). Microsatellite instability in saliva from patients with hereditary non-polyposis colon cancer and siblings carrying germline mismatch repair gene mutations. *Ann Clin Lab Sci, 41*(4), 321-330.

I

Ingham, D., Diggle, C. P., Berry, I., Bristow, C. A., Hayward, B. E., Rahman, N., Markham, A. F., Sheridan, E. G., Bonthron, D. T., & Carr, I. M. (2013). Simple detection of germline microsatellite instability for diagnosis of constitutional mismatch repair cancer syndrome. *Hum Mutat, 34*(6), 847-852.

J

Jansen, A. M., van Wezel, T., van den Akker, B. E., Ventayol Garcia, M., Ruano, D., Tops, C. M., Wagner, A., Letteboer, T. G., Gomez-Garcia, E. B., Devilee, P., Wijnen, J. T., Hes, F. J., & Morreau, H. (2016). Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. *Eur J Hum Genet, 24*(7), 1089-1092. doi:10.1038/ejhg.2015.252

Jass, J. R. (2006). Hereditary Non-Polyposis Colorectal Cancer: the rise and fall of a confusing term. *World J Gastroenterol, 12*(31), 4943-4950. doi:10.3748/wjg.v12.i31.4943

Jiricny, J. (2013). Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol, 5*(4), a012633.

Jun, S. H., Kim, T. G., & Ban, C. (2006). DNA mismatch repair system. Classical and fresh roles. *FEBS J, 273*(8), 1609-1619. doi:10.1111/j.1742-4658.2006.05190.x

K

Kadyrov, F. A., Dzantiev, L., Constantin, N., & Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. *Cell, 126*(2), 297-308. doi:10.1016/j.cell.2006.05.039

Kadyrov, F. A., Genschel, J., Fang, Y., Penland, E., Edelmann, W., & Modrich, P. (2009). A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proc Natl Acad Sci U S A, 106*(21), 8495-8500. doi:10.1073/pnas.0903654106

Kastrinos, F., Steyerberg, E. W., Mercado, R., Balmana, J., Holter, S., Gallinger, S., Siegmund, K. D., Church, J. M., Jenkins, M. A., Lindor, N. M., Thibodeau, S. N., Burbidge, L. A.,

Wenstrup, R. J., & Syngal, S. (2011). The PREMM(1,2,6) model predicts risk of MLH1, MSH2, and MSH6 germline mutations based on cancer history. *Gastroenterology, 140*(1), 73-81. doi:10.1053/j.gastro.2010.08.021

Kastrinos, F., Uno, H., Ukaegbu, C., Alvero, C., McFarland, A., Yurgelun, M. B., Kulke, M. H., Schrag, D., Meyerhardt, J. A., Fuchs, C. S., Mayer, R. J., Ng, K., Steyerberg, E. W., & Syngal, S. (2017). Development and Validation of the PREMM5 Model for Comprehensive Risk Assessment of Lynch Syndrome. *J Clin Oncol, 35*(19), 2165-2172. doi:10.1200/JCO.2016.69.6120

Kets, C. M., Hoogerbrugge, N., van Krieken, J. H., Goossens, M., Brunner, H. G., & Ligtenberg, M. J. (2009). Compound heterozygosity for two MSH2 mutations suggests mild consequences of the initiation codon variant c.1A>G of MSH2. *Eur J Hum Genet, 17*(2), 159-164.

Kinde, I., Bettegowda, C., Wang, Y., Wu, J., Agrawal, N., Shih Ie, M., Kurman, R., Dao, F., Levine, D. A., Giuntoli, R., Roden, R., Eshleman, J. R., Carvalho, J. P., Marie, S. K., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., & Diaz, L. A., Jr. (2013). Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med, 5*(167), 167ra164. doi:10.1126/scitranslmed.3004952

Kloor, M., Huth, C., Voigt, A. Y., Benner, A., Schirmacher, P., von Knebel Doeberitz, M., & Blaker, H. (2012). Prevalence of mismatch repair-deficient crypt foci in Lynch syndrome: a pathological study. *Lancet Oncol, 13*(6), 598-606.

Knudson, A. G., Jr. (1985). Hereditary cancer, oncogenes, and antioncogenes. *Cancer Res, 45*(4), 1437-1443.

Kosinski, J., Hinrichsen, I., Bujnicki, J. M., Friedhoff, P., & Plotz, G. (2010). Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. *Hum Mutat, 31*(8), 975-982.

Kou, R., Lam, H., Duan, H., Ye, L., Jongkam, N., Chen, W., Zhang, S., & Li, S. (2016). Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations. *PLoS One, 11*(1), e0146638. doi:10.1371/journal.pone.0146638

Kovacs, M. E., Papp, J., Szentirmay, Z., Otto, S., & Olah, E. (2009). Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum Mutat, 30*(2), 197-203. doi:10.1002/humu.20942

Kunkel, T. A. (2004). DNA replication fidelity. *J Biol Chem, 279*(17), 16895-16898. doi:10.1074/jbc.R400006200

Kunkel, T. A., & Erie, D. A. (2015). Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet, 49*, 291-313.

Kurdyukov, S., & Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology, 5*(1), 3. doi:10.3390/biology5010003

Kwok, C. T., Ward, R. L., Hawkins, N. J., & Hitchins, M. P. (2010). Detection of allelic imbalance in MLH1 expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Fam Cancer, 9*(3), 345-356. doi:10.1007/s10689-009-9314-0

L

Ladas, I., Yu, F., Leong, K. W., Fitarelli-Kiehl, M., Song, C., Ashtaputre, R., Kulke, M., Mamon, H., & Makrigiorgos, G. M. (2018). Enhanced detection of microsatellite instability using pre-PCR elimination of wild-type DNA homo-polymers in tissue and liquid biopsies. *Nucleic Acids Res, 46*(12), e74. doi:10.1093/nar/gky251

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*(23), 6199-6207. doi:10.1021/bi7024639

Lavoine, N., Colas, C., Muleris, M., Bodo, S., Duval, A., Entz-Werle, N., Coulet, F., Cabaret, O., Andreiuolo, F., Charpy, C., Sebille, G., Wang, Q., Lejeune, S., Buisine, M. P., Leroux, D., Couillault, G., Leverger, G., Fricker, J. P., Guimbaud, R., Mathieu-Dramard, M., Jedraszak, G., Cohen-Hagenauer, O., Guerrini-Rousseau, L., Bourdeaut, F., Grill, J., Caron, O., Baert-Dusermont, S., Tinat, J., Bougeard, G., Frebourg, T., & Brugieres, L. (2015). Constitutional mismatch repair deficiency syndrome: clinical description in a French cohort. *J Med Genet, 52*(11), 770-778.

Lázaro, C., Feliubadaló, L., & del Valle, J. (2018). Genetic Testing in Hereditary Colorectal Cancer. In L. Valle, S. B. Gruber, & G. Capellá (Eds.), *Hereditary Colorectal Cancer: Genetic Basis and Clinical Implications* (pp. 209-232). Cham: Springer International Publishing.

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L. A., Nystrom-Lahti, M., & et al. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell, 75*(6), 1215-1225. doi:10.1016/0092-8674(93)90330-s

Lee, J. B., Cho, W. K., Park, J., Jeon, Y., Kim, D., Lee, S. H., & Fishel, R. (2014). Single-molecule views of MutS on mismatched DNA. *DNA Repair (Amst), 20*, 82-93. doi:10.1016/j.dnarep.2014.02.014

Lee, K., Krempely, K., Roberts, M. E., Anderson, M. J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Huntsman, D., Kaurah, P., Kesserwan, C., Landrith, T., Li, S., Mensenkamp, A. R., Oliveira, C., Pardo, C., Pesaran, T., Richardson, M., Slavin, T. P., Spurdle, A. B., Trapp, M., Witkowski, L., Yi, C. S., Zhang, L., Plon, S. E., Schrader, K. A., & Karam, R. (2018). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum Mutat, 39*(11), 1553-1568. doi:10.1002/humu.23650

Leenders, E., Westdorp, H., Bruggemann, R. J., Loeffen, J., Kratz, C., Burn, J., Hoogerbrugge, N., & Jongmans, M. C. J. (2018). Cancer prevention by aspirin in children with Constitutional Mismatch Repair Deficiency (CMMRD). *Eur J Hum Genet, 26*(10), 1417-1423. doi:10.1038/s41431-018-0197-0

Leong, V., Lorenowicz, J., Kozij, N., & Guarne, A. (2009). Nuclear import of human MLH1, PMS2, and MutLalpha: redundancy is the key. *Mol Carcinog, 48*(8), 742-750. doi:10.1002/mc.20514

Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res, 18*(1), 85-98.

Li, L., Hamel, N., Baker, K., McGuffin, M. J., Couillard, M., Gologan, A., Marcus, V. A., Chodirker, B., Chudley, A., Stefanovici, C., Durandy, A., Hegele, R. A., Feng, B. J., Goldgar, D. E., Zhu, J., De Rosa, M., Gruber, S. B., Wimmer, K., Young, B., Chong, G., Tischkowitz, M. D., & Foulkes, W. D. (2015). A homozygous PMS2 founder mutation with an attenuated constitutional mismatch repair deficiency phenotype. *J Med Genet, 52*(5), 348-352.

Li, Q., & Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet, 100*(2), 267-280. doi:10.1016/j.ajhg.2017.01.004

Ligtenberg, M. J., Kuiper, R. P., Chan, T. L., Goossens, M., Hebeda, K. M., Voorendt, M., Lee, T. Y., Bodmer, D., Hoenselaar, E., Hendriks-Cornelissen, S. J., Tsui, W. Y., Kong, C. K., Brunner, H. G., van Kessel, A. G., Yuen, S. T., van Krieken, J. H., Leung, S. Y., & Hoogerbrugge, N. (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*(1), 112-117. doi:10.1038/ng.283

Lim, S. B., Jeong, S. Y., Kim, I. J., Kim, D. Y., Jung, K. H., Chang, H. J., Choi, H. S., Sohn, D. K., Kang, H. C., Shin, Y., Jang, S. G., Park, J. H., & Park, J. G. (2006). Analysis of microsatellite instability in stool DNA of patients with colorectal cancer using denaturing high performance liquid chromatography. *World J Gastroenterol, 12*(41), 6689-6692. doi:10.3748/wjg.v12.i41.6689

Lindor, N. M., Guidugli, L., Wang, X., Vallee, M. P., Monteiro, A. N., Tavtigian, S., Goldgar, D. E., & Couch, F. J. (2012). A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat, 33*(1), 8-21. doi:10.1002/humu.21627

Liu, D., Keijzers, G., & Rasmussen, L. J. (2017). DNA mismatch repair and its many roles in eukaryotic cells. *Mutat Res, 773*, 174-187. doi:10.1016/j.mrrev.2017.07.001

Liu, Q., Hesson, L. B., Nunez, A. C., Packham, D., Hawkins, N. J., Ward, R. L., & Sloane, M. A. (2016). Pathogenic germline MCM9 variants are rare in Australian Lynch-like syndrome patients. *Cancer Genet, 209*(11), 497-500. doi:10.1016/j.cancergen.2016.10.001

Liu, Q., Thompson, B. A., Ward, R. L., Hesson, L. B., & Sloane, M. A. (2016). Understanding the Pathogenicity of Noncoding Mismatch Repair Gene Promoter Variants in Lynch Syndrome. *Hum Mutat, 37*(5), 417-426. doi:10.1002/humu.22971

Lynch, H. T., & Krush, A. J. (1971). Cancer family "G" revisited: 1895-1970. *Cancer, 27*(6), 1505-1511.

Lynch, H. T., Shaw, M. W., Magnuson, C. W., Larsen, A. L., & Krush, A. J. (1966). Hereditary factors in cancer. Study of two large midwestern kindreds. *Arch Intern Med, 117*(2), 206-212.

Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D., & Hitchins, M. P. (2015). Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer, 15*(3), 181-194.

M

MacConaill, L. E., Burns, R. T., Nag, A., Coleman, H. A., Slevin, M. K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M. S., Ducar, M. D., Meyerson, M., & Thorner, A. R. (2018). Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics, 19*(1), 30. doi:10.1186/s12864-017-4428-5

Maletzki, C., Huehns, M., Bauer, I., Ripperger, T., Mork, M. M., Vilar, E., Klocking, S., Zettl, H., Prall, F., & Linnebacher, M. (2017). Frameshift mutational target gene analysis identifies similarities and differences in constitutional mismatch repair-deficiency and Lynch syndrome. *Mol Carcinog, 56*(7), 1753-1764. doi:10.1002/mc.22632

Manhart, C. M., & Alani, E. (2016). Roles for mismatch repair family proteins in promoting meiotic crossing over. *DNA Repair (Amst), 38*, 84-93. doi:10.1016/j.dnarep.2015.11.024

Marcos, I., Borrego, S., Urioste, M., Garcia-Valles, C., & Antinolo, G. (2006). Mutations in the DNA mismatch repair gene MLH1 associated with early-onset colon cancer. *J Pediatr, 148*(6), 837-839.

Maritschnegg, E., Wang, Y., Pecha, N., Horvat, R., Van Nieuwenhuysen, E., Vergote, I., Heitz, F., Sehouli, J., Kinde, I., Diaz, L. A., Jr., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., Speiser, P., & Zeillinger, R. (2015). Lavage of the Uterine Cavity for Molecular Detection of Mullerian Duct Carcinomas: A Proof-of-Concept Study. *J Clin Oncol, 33*(36), 4293-4300. doi:10.1200/JCO.2015.61.3083

McCulloch, S. D., Kokoska, R. J., Chilkova, O., Welch, C. M., Johansson, E., Burgers, P. M., & Kunkel, T. A. (2004). Enzymatic switching for efficient and accurate translesion DNA replication. *Nucleic Acids Res, 32*(15), 4665-4675. doi:10.1093/nar/gkh777

McMurray, C. T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet, 11*(11), 786-799. doi:10.1038/nrg2828

Menendez, M., Castellvi-Bel, S., Pineda, M., de Cid, R., Munoz, J., Gonzalez, S., Teule, A., Balaguer, F., Ramon y Cajal, T., Rene, J. M., Blanco, I., Castells, A., & Capella, G. (2010). Founder effect of a pathogenic MSH2 mutation identified in Spanish families with Lynch syndrome. *Clin Genet, 78*(2), 186-190. doi:10.1111/j.1399-0004.2009.01346.x

Mester, J. L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K. S., Costa, H. A., Lachlan, K., Ngeow, J., Barnholtz-Sloan, J., Sesock, K., Hernandez, F., Zhang, L., Milko, L., Plon, S. E., Hegde, M., & Eng, C. (2018). Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Hum Mutat, 39*(11), 1581-1592. doi:10.1002/humu.23636

Michaeli, O., & Tabori, U. (2018). Pediatric High Grade Gliomas in the Context of Cancer Predisposition Syndromes. *J Korean Neurosurg Soc, 61*(3), 319-332. doi:10.3340/jkns.2018.0031

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M., & Mori, T. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet, 17*(3), 271-272. doi:10.1038/ng1197-271

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annu Rev Genet, 25*, 229-253. doi:10.1146/annurev.ge.25.120191.001305

Moles-Fernandez, A., Duran-Lozano, L., Montalban, G., Bonache, S., Lopez-Perolio, I., Menendez, M., Santamarina, M., Behar, R., Blanco, A., Carrasco, E., Lopez-Fernandez, A., Stjepanovic, N., Balmana, J., Capella, G., Pineda, M., Vega, A., Lazaro, C., de la Hoya, M., Diez, O., & Gutierrez-Enriquez, S. (2018). Computational Tools for Splicing Defect Prediction in Breast/Ovarian Cancer Genes: How Efficient Are They at Predicting RNA Alterations? *Front Genet, 9*, 366. doi:10.3389/fgene.2018.00366

Moller, P., Seppala, T. T., Bernstein, I., Holinski-Feder, E., Sala, P., Gareth Evans, D., Lindblom, A., Macrae, F., Blanco, I., Sijmons, R. H., Jeffries, J., Vasen, H. F. A., Burn, J., Nakken, S., Hovig, E., Rodland, E. A., Tharmaratnam, K., de Vos Tot Nederveen Cappel, W. H., Hill, J., Wijnen, J. T., Jenkins, M. A., Green, K., Lalloo, F., Sunde, L., Mints, M., Bertario, L., Pineda, M., Navarro, M., Morak, M., Renkonen-Sinisalo, L., Valentin, M. D., Frayling, I. M., Plazzer, J. P., Pylvanainen, K., Genuardi, M., Mecklin, J. P., Moeslein, G., Sampson, J. R., & Capella, G. (2018). Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. *Gut, 67*(7), 1306-1316.

Morak, M., Heidenreich, B., Keller, G., Hampel, H., Laner, A., de la Chapelle, A., & Holinski-Feder, E. (2014). Biallelic MUTYH mutations can mimic Lynch syndrome. *Eur J Hum Genet, 22*(11), 1334-1337. doi:10.1038/ejhg.2014.15

Morak, M., Schackert, H. K., Rahner, N., Betz, B., Ebert, M., Walldorf, C., Royer-Pokora, B., Schulmann, K., von Knebel-Doeberitz, M., Dietmaier, W., Keller, G., Kerker, B.,

Leitner, G., & Holinski-Feder, E. (2008). 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, 16*(7), 804-811. doi:10.1038/ejhg.2008.25

Morak, M., Schaefer, K., Steinke-Lange, V., Koehler, U., Keinath, S., Massdorf, T., Mauracher, B., Rahner, N., Bailey, J., Kling, C., Haeusser, T., Laner, A., & Holinski-Feder, E. (2019). Full-length transcript amplification and sequencing as universal method to test mRNA integrity and biallelic expression in mismatch repair genes. *European Journal of Human Genetics*. doi:10.1038/s41431-019-0472-8

Moreira, L., Balaguer, F., Lindor, N., de la Chapelle, A., Hampel, H., Aaltonen, L. A., Hopper, J. L., Le Marchand, L., Gallinger, S., Newcomb, P. A., Haile, R., Thibodeau, S. N., Gunawardena, S., Jenkins, M. A., Buchanan, D. D., Potter, J. D., Baron, J. A., Ahnen, D. J., Moreno, V., Andreu, M., Ponz de Leon, M., Rustgi, A. K., & Castells, A. (2012). Identification of Lynch syndrome among patients with colorectal cancer. *Jama, 308*(15), 1555-1565.

Moreira, L., Munoz, J., Cuatrecasas, M., Quintanilla, I., Leoz, M. L., Carballal, S., Ocana, T., Lopez-Ceron, M., Pellise, M., Castellvi-Bel, S., Jover, R., Andreu, M., Carracedo, A., Xicola, R. M., Llor, X., Boland, C. R., Goel, A., Castells, A., Balaguer, F., & Gastrointestinal Oncology Group of the Spanish Gastroenterological, A. (2015). Prevalence of somatic mutl homolog 1 promoter hypermethylation in Lynch syndrome colorectal cancer. *Cancer, 121*(9), 1395-1404. doi:10.1002/cncr.29190

N

National Society of Genetic Counselors' Definition Task, F., Resta, R., Biesecker, B. B., Bennett, R. L., Blum, S., Hahn, S. E., Strecker, M. N., & Williams, J. L. (2006). A new definition of Genetic Counseling: National Society of Genetic Counselors' Task Force report. *J Genet Couns, 15*(2), 77-83. doi:10.1007/s10897-005-9014-3

Newcomb, P. A., Baron, J., Cotterchio, M., Gallinger, S., Grove, J., Haile, R., Hall, D., Hopper, J. L., Jass, J., Le Marchand, L., Limburg, P., Lindor, N., Potter, J. D., Templeton, A. S., Thibodeau, S., Seminara, D., & Colon Cancer Family, R. (2007). Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev, 16*(11), 2331-2343. doi:10.1158/1055-9965.EPI-07-0648

Newton, K., Jorgensen, N. M., Wallace, A. J., Buchanan, D. D., Lalloo, F., McMahon, R. F., Hill, J., & Evans, D. G. (2014). Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). *J Med Genet, 51*(12), 789-796. doi:10.1136/jmedgenet-2014-102552

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., & et al. (1994). Mutations of

two PMS homologues in hereditary nonpolyposis colon cancer. *Nature, 371*(6492), 75-80. doi:10.1038/371075a0

Nieminen, T. T., Gylling, A., Abdel-Rahman, W. M., Nuorva, K., Aarnio, M., Renkonen-Sinisalo, L., Jarvinen, H. J., Mecklin, J. P., Butzow, R., & Peltomaki, P. (2009). Molecular analysis of endometrial tumorigenesis: importance of complex hyperplasia regardless of atypia. *Clin Cancer Res, 15*(18), 5772-5783. doi:10.1158/1078-0432.CCR-09-0506

Nishant, K. T., Plys, A. J., & Alani, E. (2008). A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in Saccharomyces cerevisiae. *Genetics, 179*(2), 747-755. doi:10.1534/genetics.108.086645

Niskakoski, A., Pasanen, A., Lassus, H., Renkonen-Sinisalo, L., Kaur, S., Mecklin, J. P., Butzow, R., & Peltomaki, P. (2018). Molecular changes preceding endometrial and ovarian cancer: a study of consecutive endometrial specimens from Lynch syndrome surveillance. *Mod Pathol, 31*(8), 1291-1301. doi:10.1038/s41379-018-0044-4

Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M. D., Wendl, M. C., & Ding, L. (2014). MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics, 30*(7), 1015-1016.

Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y. Y., Kobayashi, Y., Patil, N., Thusberg, J., Westbrook, M., & Topper, S. (2017). Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med, 19*(10), 1105-1117.

O

Obmolova, G., Ban, C., Hsieh, P., & Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature, 407*(6805), 703-710. doi:10.1038/35037509

Okkels, H., Lindorff-Larsen, K., Thorlasius-Ussing, O., Vyberg, M., Lindebjerg, J., Sunde, L., Bernstein, I., Klarskov, L., Holck, S., & Krarup, H. B. (2012). MSH6 mutations are frequent in hereditary nonpolyposis colorectal cancer families with normal pMSH6 expression as detected by immunohistochemistry. *Appl Immunohistochem Mol Morphol, 20*(5), 470-477. doi:10.1097/PAI.0b013e318249739b

Olkinuora, A., Nieminen, T. T., Martensson, E., Rohlin, A., Ristimaki, A., Koskenvuo, L., Lepisto, A., Swedish Extended Genetic Analysis of Colorectal Neoplasia Study, G., Gebre-Medhin, S., Nordling, M., & Peltomaki, P. (2019). Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. *Genet Med, 21*(8), 1868-1873. doi:10.1038/s41436-018-0405-x

Ollila, S., Dermadi Bebek, D., Jiricny, J., & Nystrom, M. (2008). Mechanisms of pathogenicity in human MSH2 missense mutants. *Hum Mutat, 29*(11), 1355-1363.

Ollila, S., Sarantaus, L., Kariola, R., Chan, P., Hampel, H., Holinski-Feder, E., Macrae, F., Kohonen-Corish, M., Gerdes, A. M., Peltomaki, P., Mangold, E., de la Chapelle, A., Greenblatt, M., & Nystrom, M. (2006). Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. *Gastroenterology, 131*(5), 1408-1417.

P

Pai, R. K., Dudley, B., Karloski, E., Brand, R. E., O'Callaghan, N., Rosty, C., Buchanan, D. D., Jenkins, M. A., Thibodeau, S. N., French, A. J., Lindor, N. M., & Pai, R. K. (2018). DNA mismatch repair protein deficient non-neoplastic colonic crypts: a novel indicator of Lynch syndrome. *Mod Pathol, 31*(10), 1608-1618. doi:10.1038/s41379-018-0079-6

Palomaki, G. E., McClain, M. R., Melillo, S., Hampel, H. L., & Thibodeau, S. N. (2009). EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. *Genet Med, 11*(1), 42-65. doi:10.1097/GIM.0b013e31818fa2db

Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., & et al. (1994). Mutation of a mutL homolog in hereditary colon cancer. *Science, 263*(5153), 1625-1629. doi:10.1126/science.8128251

Parsons, M. T., Tudini, E., Li, H., Hahnen, E., Wappenschmidt, B., Feliubadalo, L., Aalfs, C. M., Agata, S., Aittomaki, K., Alducci, E., Alonso-Cerezo, M. C., Arnold, N., Auber, B., Austin, R., Azzollini, J., Balmana, J., Barbieri, E., Bartram, C. R., Blanco, A., Blumcke, B., Bonache, S., Bonanni, B., Borg, A., Bortesi, B., Brunet, J., Bruzzone, C., Bucksch, K., Cagnoli, G., Caldes, T., Caliebe, A., Caligo, M. A., Calvello, M., Capone, G. L., Caputo, S. M., Carnevali, I., Carrasco, E., Caux-Moncoutier, V., Cavalli, P., Cini, G., Clarke, E. M., Concolino, P., Cops, E. J., Cortesi, L., Couch, F. J., Darder, E., de la Hoya, M., Dean, M., Debatin, I., Del Valle, J., Delnatte, C., Derive, N., Diez, O., Ditsch, N., Domchek, S. M., Dutrannoy, V., Eccles, D. M., Ehrencrona, H., Enders, U., Evans, D. G., Faust, U., Felbor, U., Feroce, I., Fine, M., Galvao, H. C. R., Gambino, G., Gehrig, A., Gensini, F., Gerdes, A. M., Germani, A., Giesecke, J., Gismondi, V., Gomez, C., Gomez Garcia, E. B., Gonzalez, S., Grau, E., Grill, S., Gross, E., Guerrieri-Gonzaga, A., Guillaud-Bataille, M., Gutierrez-Enriquez, S., Haaf, T., Hackmann, K., Hansen, T. V. O., Harris, M., Hauke, J., Heinrich, T., Hellebrand, H., Herold, K. N., Honisch, E., Horvath, J., Houdayer, C., Hubbel, V., Iglesias, S., Izquierdo, A., James, P. A., Janssen, L. A. M., Jeschke, U., Kaulfuss, S., Keupp, K., Kiechle, M., Kolbl, A., Krieger, S., Kruse, T. A., Kvist, A., Lalloo, F., Larsen, M., Lattimore, V. L., Lautrup, C., Ledig, S., Leinert, E., Lewis, A. L., Lim, J., Loeffler, M., Lopez-Fernandez, A., Lucci-Cordisco, E., Maass, N., Manoukian, S., Marabelli, M., Matricardi, L., Meindl, A., Michelli, R. D., Moghadasi, S., Moles-Fernandez, A., Montagna, M., Montalban, G., Monteiro, A. N., Montes, E., Mori, L., Moserle, L., Muller, C. R., Mundhenke, C., Naldi, N., Nathanson, K. L., Navarro, M., Nevanlinna, H., Nichols, C. B., Niederacher, D., Nielsen, H. R., Ong, K.

R., Pachter, N., Palmero, E. I., Papi, L., Pedersen, I. S., Peissel, B., Perez-Segura, P., Pfeifer, K., Pineda, M., Pohl-Rescigno, E., Poplawski, N. K., Porfirio, B., Quante, A. S., Ramser, J., Reis, R. M., Revillion, F., Rhiem, K., Riboli, B., Ritter, J., Rivera, D., Rofes, P., Rump, A., Salinas, M., Sanchez de Abajo, A. M., Schmidt, G., Schoenwiese, U., Seggewiss, J., Solanes, A., Steinemann, D., Stiller, M., Stoppa-Lyonnet, D., Sullivan, K. J., Susman, R., Sutter, C., Tavtigian, S. V., Teo, S. H., Teule, A., Thomassen, M., Tibiletti, M. G., Tognazzo, S., Toland, A. E., Tornero, E., Torngren, T., Torres-Esquius, S., Toss, A., Trainer, A. H., van Asperen, C. J., van Mackelenbergh, M. T., Varesco, L., Vargas-Parra, G., Varon, R., Vega, A., Velasco, A., Vesper, A. S., Viel, A., Vreeswijk, M. P. G., Wagner, S. A., Waha, A., Walker, L. C., Walters, R. J., Wang-Gohrke, S., Weber, B. H. F., Weichert, W., Wieland, K., Wiesmuller, L., Witzel, I., Wockel, A., Woodward, E. R., Zachariae, S., Zampiga, V., Zeder-Goss, C., Investigators, K. C., Lazaro, C., De Nicolo, A., Radice, P., Engel, C., Schmutzler, R. K., Goldgar, D. E., & Spurdle, A. B. (2019). Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Hum Mutat*. doi:10.1002/humu.23818

Pastrello, C., Pin, E., Marroni, F., Bedin, C., Fornasarig, M., Tibiletti, M. G., Oliani, C., Ponz de Leon, M., Urso, E. D., Della Puppa, L., Agostini, M., & Viel, A. (2011). Integrated analysis of unclassified variants in mismatch repair genes. *Genet Med, 13*(2), 115-124.

Peltomaki, P. (2016). Update on Lynch syndrome genomics. *Fam Cancer, 15*(3), 385-393. doi:10.1007/s10689-016-9882-8

Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P., & Wang, Y. (2015). Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics, 16*, 589. doi:10.1186/s12864-015-1806-8

Peña-Diaz, J., & Jiricny, J. (2012). Mammalian mismatch repair: error-free or error-prone? *Trends Biochem Sci, 37*(5), 206-214. doi:10.1016/j.tibs.2012.03.001

Peña-Diaz, J., & Rasmussen, L. J. (2016). Approaches to diagnose DNA mismatch repair gene defects in cancer. *DNA Repair (Amst), 38*, 147-154.

Perez-Carbonell, L., Alenda, C., Paya, A., Castillejo, A., Barbera, V. M., Guillen, C., Rojas, E., Acame, N., Gutierrez-Avino, F. J., Castells, A., Llor, X., Andreu, M., Soto, J. L., & Jover, R. (2010). Methylation analysis of MLH1 improves the selection of patients for genetic testing in Lynch syndrome. *J Mol Diagn, 12*(4), 498-504. doi:10.2353/jmoldx.2010.090212

Perez-Palma, E., Gramm, M., Nurnberg, P., May, P., & Lal, D. (2019). Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. *Nucleic Acids Res, 47*(W1), W99-W105. doi:10.1093/nar/gkz411

Pineda, M., Gonzalez, S., Lazaro, C., Blanco, I., & Capella, G. (2010). Detection of genetic alterations in hereditary colorectal cancer screening. *Mutat Res, 693*(1-2), 19-31. doi:10.1016/j.mrfmmm.2009.11.002

Pinheiro, M., Pinto, C., Peixoto, A., Veiga, I., Lopes, P., Henrique, R., Baldaia, H., Carneiro, F., Seruca, R., Tomlinson, I., Kovac, M., Heinimann, K., & Teixeira, M. R. (2015). Target gene mutational pattern in Lynch syndrome colorectal carcinomas according to tumour location and germline mutation. *Br J Cancer, 113*(4), 686-692.

Plaschke, J., Ruschoff, J., & Schackert, H. K. (2003). Genomic rearrangements of hMSH6 contribute to the genetic predisposition in suspected hereditary non-polyposis colorectal cancer syndrome. *J Med Genet, 40*(8), 597-600. doi:10.1136/jmg.40.8.597

Plazzer, J. P., Sijmons, R. H., Woods, M. O., Peltomaki, P., Thompson, B., Den Dunnen, J. T., & Macrae, F. (2013). The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. *Fam Cancer, 12*(2), 175-180. doi:10.1007/s10689-013-9616-0

Plon, S. E., Cooper, H. P., Parks, B., Dhar, S. U., Kelly, P. A., Weinberg, A. D., Staggs, S., Wang, T., & Hilsenbeck, S. (2011). Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. *Genet Med, 13*(2), 148-154. doi:10.1097/GIM.0b013e318207f564

Plotz, G., Welsch, C., Giron-Monzon, L., Friedhoff, P., Albrecht, M., Piiper, A., Biondi, R. M., Lengauer, T., Zeuzem, S., & Raedle, J. (2006). Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. *Nucleic Acids Res, 34*(22), 6574-6586.

Ponti, G., Castellsague, E., Ruini, C., Percesepe, A., & Tomasi, A. (2015). Mismatch repair genes founder mutations and cancer susceptibility in Lynch syndrome. *Clin Genet, 87*(6), 507-516. doi:10.1111/cge.12529

Puig, P., Urgell, E., Capella, G., Sancho, F. J., Pujol, J., Boadas, J., Farre, A., Lluis, F., Gonzalez-Sastre, F., & Mora, J. (2000). A highly sensitive method for K-ras mutation detection is useful in diagnosis of gastrointestinal cancer. *Int J Cancer, 85*(1), 73-77.

R

Radman, M., & Wagner, R. (1986). Mismatch repair in Escherichia coli. *Annu Rev Genet, 20*, 523-538. doi:10.1146/annurev.ge.20.120186.002515

Raevaara, T. E., Korhonen, M. K., Lohi, H., Hampel, H., Lynch, E., Lonnqvist, K. E., Holinski-Feder, E., Sutter, C., McKinnon, W., Duraisamy, S., Gerdes, A. M., Peltomaki, P., Kohonen-Ccorish, M., Mangold, E., Macrae, F., Greenblatt, M., de la Chapelle, A., & Nystrom, M. (2005). Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology, 129*(2), 537-549.

Rantanen, E., Hietala, M., Kristoffersson, U., Nippert, I., Schmidtke, J., Sequeiros, J., & Kaariainen, H. (2008). What is ideal genetic counselling? A survey of current

international guidelines. *Eur J Hum Genet, 16*(4), 445-452. doi:10.1038/sj.ejhg.5201983

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*(24), 21810-21820. doi:10.1074/jbc.M108787200

Rasmussen, L. J., Heinen, C. D., Royer-Pokora, B., Drost, M., Tavtigian, S., Hofstra, R. M., & de Wind, N. (2012). Pathological assessment of mismatch repair gene variants in Lynch syndrome: past, present, and future. *Hum Mutat, 33*(12), 1617-1625.

Reyes, G. X., Schmidt, T. T., Kolodner, R. D., & Hombauer, H. (2015). New insights into the mechanism of DNA mismatch repair. *Chromosoma, 124*(4), 443-462.

Ricciardone, M. D., Ozcelik, T., Cevher, B., Ozdag, H., Tuncer, M., Gurgey, A., Uzunalimoglu, O., Cetinkaya, H., Tanyeli, A., Erken, E., & Ozturk, M. (1999). Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. *Cancer Res, 59*(2), 290-293.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med, 17*(5), 405-424.

Ripperger, T., & Schlegelberger, B. (2016). Acute lymphoblastic leukemia and lymphoma in the context of constitutional mismatch repair deficiency syndrome. *Eur J Med Genet, 59*(3), 133-142. doi:10.1016/j.ejmg.2015.12.014

Risio, M., Reato, G., di Celle, P. F., Fizzotti, M., Rossini, F. P., & Foa, R. (1996). Microsatellite instability is associated with the histological features of the tumor in nonfamilial colorectal cancer. *Cancer Res, 56*(23), 5470-5474.

Rivera-Munoz, E. A., Milko, L. V., Harrison, S. M., Azzariti, D. R., Kurtz, C. L., Lee, K., Mester, J. L., Weaver, M. A., Currey, E., Craigen, W., Eng, C., Funke, B., Hegde, M., Hershberger, R. E., Mao, R., Steiner, R. D., Vincent, L. M., Martin, C. L., Plon, S. E., Ramos, E., Rehm, H. L., Watson, M., & Berg, J. S. (2018). ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Hum Mutat, 39*(11), 1614-1622. doi:10.1002/humu.23645

Robertson, D. J., & Imperiale, T. F. (2015). Stool Testing for Colorectal Cancer Screening. *Gastroenterology, 149*(5), 1286-1293. doi:10.1053/j.gastro.2015.05.045

Rodriguez-Bigas, M. A., Boland, C. R., Hamilton, S. R., Henson, D. E., Jass, J. R., Khan, P. M., Lynch, H., Perucho, M., Smyrk, T., Sobin, L., & Srivastava, S. (1997). A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome:

meeting highlights and Bethesda guidelines. *J Natl Cancer Inst, 89*(23), 1758-1762. doi:10.1093/jnci/89.23.1758

Rodriguez-Soler, M., Perez-Carbonell, L., Guarinos, C., Zapater, P., Castillejo, A., Barbera, V. M., Juarez, M., Bessa, X., Xicola, R. M., Clofent, J., Bujanda, L., Balaguer, F., Rene, J. M., de-Castro, L., Marin-Gabriel, J. C., Lanas, A., Cubiella, J., Nicolas-Perez, D., Brea-Fernandez, A., Castellvi-Bel, S., Alenda, C., Ruiz-Ponte, C., Carracedo, A., Castells, A., Andreu, M., Llor, X., Soto, J. L., Paya, A., & Jover, R. (2013). Risk of cancer in cases of suspected lynch syndrome without germline mutation. *Gastroenterology, 144*(5), 926-932 e921; quiz e913-924. doi:10.1053/j.gastro.2013.01.044

Rohlin, A., Rambech, E., Kvist, A., Torngren, T., Eiengard, F., Lundstam, U., Zagoras, T., Gebre-Medhin, S., Borg, A., Bjork, J., Nilbert, M., & Nordling, M. (2017). Expanding the genotype-phenotype spectrum in hereditary colorectal cancer by gene panel testing. *Fam Cancer, 16*(2), 195-203.

Rolnick, S. J., Rahm, A. K., Jackson, J. M., Nekhlyudov, L., Goddard, K. A., Field, T., McCarty, C., Nakasato, C., Roblin, D., Anderson, C. P., & Valdez, R. (2011). Barriers in identification and referral to genetic counseling for familial cancer risk: the perspective of genetic service providers. *J Genet Couns, 20*(3), 314-322. doi:10.1007/s10897-011-9351-3

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T., & Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature, 475*(7356), 348-352. doi:10.1038/nature10242

S

Salipante, S. J., Scroggins, S. M., Hampel, H. L., Turner, E. H., & Pritchard, C. C. (2014). Microsatellite instability detection by next generation sequencing. *Clin Chem, 60*(9), 1192-1199.

Salk, J. J., Schmitt, M. W., & Loeb, L. A. (2018). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet, 19*(5), 269-285.

Samorodnitsky, E., Jewell, B. M., Hagopian, R., Miya, J., Wing, M. R., Lyon, E., Damodaran, S., Bhatt, D., Reeser, J. W., Datta, J., & Roychowdhury, S. (2015). Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Hum Mutat, 36*(9), 903-914. doi:10.1002/humu.22825

Sanchez, A., Navarro, M., Ocaña, T., Pineda, M., Rodriguez-Moranta, F., Soriano, A., Cajal, T. R. y., llort, G., Yagüe, C., Picó, M. D., Jover, R., Fernandez, A. L., de Castro, E. M., Alvarez, C., Bessa, X., Rivas, L., Rodríguez-Alcalde, D., Dacal, A., Herraiz, M., Garau, C., Bujanda, L., Cid, L., Povés, C., Garzon, M., Pizarro, A., Gomez, A., Salces, I., Ponce, M., Carrillo-Palau, M., Aguirre, E., Saperas, E., Marnrique, A. V., Masferrer, D. F., Piñol, V., Carballal, S., Rivero, L., Pellise, M., Serra, M., Balmaña, J., Capellà, G., Brunet, J., Castells, A., Moreira, L., & Balaguer, F. (2017). Colorectal Cancer Incidence in Lynch Syndrome Patients: First Report of a Multicenter Nation-Wide Study. *Gastroenterology, 152*(5), S552. doi:10.1016/S0016-5085(17)32003-6

Schmitt, M. W., Fox, E. J., Prindle, M. J., Reid-Bayliss, K. S., True, L. D., Radich, J. P., & Loeb, L. A. (2015). Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat Methods, 12*(5), 423-425. doi:10.1038/nmeth.3351

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., & Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A, 109*(36), 14508-14513.

Schmutte, C., Sadoff, M. M., Shim, K. S., Acharya, S., & Fishel, R. (2001). The interaction of DNA mismatch repair proteins with human exonuclease I. *J Biol Chem, 276*(35), 33011-33018. doi:10.1074/jbc.M102670200

Shia, J., Stadler, Z. K., Weiser, M. R., Vakiani, E., Mendelsohn, R., Markowitz, A. J., Shike, M., Boland, C. R., & Klimstra, D. S. (2015). Mismatch repair deficient-crypts in non-neoplastic colonic mucosa in Lynch syndrome: insights from an illustrative case. *Fam Cancer, 14*(1), 61-68. doi:10.1007/s10689-014-9751-2

Shlien, A., Campbell, B. B., de Borja, R., Alexandrov, L. B., Merico, D., Wedge, D., Van Loo, P., Tarpey, P. S., Coupland, P., Behjati, S., Pollett, A., Lipman, T., Heidari, A., Deshmukh, S., Avitzur, N., Meier, B., Gerstung, M., Hong, Y., Merino, D. M., Ramakrishna, M., Remke, M., Arnold, R., Panigrahi, G. B., Thakkar, N. P., Hodel, K. P., Henninger, E. E., Goksenin, A. Y., Bakry, D., Charames, G. S., Druker, H., Lerner-Ellis, J., Mistry, M., Dvir, R., Grant, R., Elhasid, R., Farah, R., Taylor, G. P., Nathan, P. C., Alexander, S., Ben-Shachar, S., Ling, S. C., Gallinger, S., Constantini, S., Dirks, P., Huang, A., Scherer, S. W., Grundy, R. G., Durno, C., Aronson, M., Gartner, A., Meyn, M. S., Taylor, M. D., Pursell, Z. F., Pearson, C. E., Malkin, D., Futreal, P. A., Stratton, M. R., Bouffet, E., Hawkins, C., Campbell, P. J., & Tabori, U. (2015). Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet, 47*(3), 257-262.

Shuen, A. Y., Lanni, S., Panigrahi, G. B., Edwards, M., Yu, L., Campbell, B. B., Mandel, A., Zhang, C., Zhukova, N., Alharbi, M., Bernstein, M., Bowers, D. C., Carroll, S., Cole, K. A., Constantini, S., Crooks, B., Dvir, R., Farah, R., Hijiya, N., George, B., Laetsch, T. W., Larouche, V., Lindhorst, S., Luiten, R. C., Magimairajan, V., Mason, G., Mason, W., Mordechai, O., Mushtaq, N., Nicholas, G., Oren, M., Palma, L., Pedroza, L. A., Ramdas, J., Samuel, D., Wolfe Schneider, K., Seeley, A., Semotiuk, K., Shamvil, A., Sumerauer,

D., Toledano, H., Tomboc, P., Wierman, M., Van Damme, A., Lee, Y. Y., Zapotocky, M., Bouffet, E., Durno, C., Aronson, M., Gallinger, S., Foulkes, W. D., Malkin, D., Tabori, U., & Pearson, C. E. (2019). Functional Repair Assay for the Diagnosis of Constitutional Mismatch Repair Deficiency From Non-Neoplastic Tissue. *J Clin Oncol*, JCO1800474.

Sidransky, D., Tokino, T., Hamilton, S. R., Kinzler, K. W., Levin, B., Frost, P., & Vogelstein, B. (1992). Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science, 256*(5053), 102-105. doi:10.1126/science.1566048

Sijmons, R. H., & Hofstra, R. M. (2016). Review: Clinical aspects of hereditary DNA Mismatch repair gene mutations. *DNA Repair (Amst), 38*, 155-162.

Slean, M. M., Panigrahi, G. B., Ranum, L. P., & Pearson, C. E. (2008). Mutagenic roles of DNA "repair" proteins in antibody diversity and disease-associated trinucleotide repeat instability. *DNA Repair (Amst), 7*(7), 1135-1154. doi:10.1016/j.dnarep.2008.03.014

Soto, J. L., & Castillejo, A. (2019). Aspectos moleculares del síndrome de Lynch y de otros síndromes de predisposición a cáncer colorrectal no polipósico. In S. E. d. O. M. (SEOM) (Ed.), *Cáncer Hereditario* (pp. 189-206): GoNext Producciones S.L.

Spurdle, A. B., Couch, F. J., Hogervorst, F. B., Radice, P., & Sinilnikova, O. M. (2008). Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat, 29*(11), 1304-1313.

Srivatsan, A., Bowen, N., & Kolodner, R. D. (2014). Mispair-specific recruitment of the Mlh1-Pms1 complex identifies repair substrates of the Saccharomyces cerevisiae Msh2-Msh3 complex. *J Biol Chem, 289*(13), 9352-9364. doi:10.1074/jbc.M114.552190

Staffa, L., Echterdiek, F., Nelius, N., Benner, A., Werft, W., Lahrmann, B., Grabe, N., Schneider, M., Tariverdian, M., von Knebel Doeberitz, M., Blaker, H., & Kloor, M. (2015). Mismatch repair-deficient crypt foci in Lynch syndrome--molecular alterations and association with clinical parameters. *PLoS One, 10*(3), e0121980.

Starita, L. M., Islam, M. M., Banerjee, T., Adamovich, A. I., Gullingsrud, J., Fields, S., Shendure, J., & Parvin, J. D. (2018). A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. *Am J Hum Genet, 103*(4), 498-508. doi:10.1016/j.ajhg.2018.07.016

Stoffel, E. M., Koeppe, E., Everett, J., Ulintz, P., Kiel, M., Osborne, J., Williams, L., Hanson, K., Gruber, S. B., & Rozek, L. S. (2018). Germline Genetic Features of Young Individuals With Colorectal Cancer. *Gastroenterology, 154*(4), 897-905 e891. doi:10.1053/j.gastro.2017.11.004

Stoffel, E. M., Yurgelun, M. B., & Boland, C. R. (2018). Lynch Syndrome. In L. Valle, S. B. Gruber, & G. Capellá (Eds.), *Hereditary Colorectal Cancer: Genetic Basis and Clinical Implications* (pp. 3-19). Cham: Springer International Publishing.

Suerink, M., Potjer, T. P., Versluijs, A. B., Ten Broeke, S. W., Tops, C. M., Wimmer, K., & Nielsen, M. (2018). Constitutional mismatch repair deficiency in a healthy child: On the spot diagnosis? *Clin Genet, 93*(1), 134-137. doi:10.1111/cge.13053

Suerink, M., Ripperger, T., Messiaen, L., Menko, F. H., Bourdeaut, F., Colas, C., Jongmans, M., Goldberg, Y., Nielsen, M., Muleris, M., van Kouwen, M., Slavc, I., Kratz, C., Vasen, H. F., Brugires, L., Legius, E., & Wimmer, K. (2018). Constitutional mismatch repair deficiency as a differential diagnosis of neurofibromatosis type 1: consensus guidelines for testing a child without malignancy. *J Med Genet, 56*(2), 53-62.

Sugano, K., Nakajima, T., Sekine, S., Taniguchi, H., Saito, S., Takahashi, M., Ushiama, M., Sakamoto, H., & Yoshida, T. (2016). Germline PMS2 mutation screened by mismatch repair protein immunohistochemistry of colorectal cancer in Japan. *Cancer Sci, 107*(11), 1677-1686. doi:10.1111/cas.13073

Sun, X., Liu, Y., Lutterbaugh, J., Chen, W. D., Markowitz, S. D., & Guo, B. (2006). Detection of mononucleotide repeat sequence alterations in a large background of normal DNA for screening high-frequency microsatellite instability cancers. *Clin Cancer Res, 12*(2), 454-459. doi:10.1158/1078-0432.CCR-05-0919

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*(6), 1804-1811. doi:10.1053/gast.2002.37070

Susswein, L. R., Marshall, M. L., Nusbaum, R., Vogel Postula, K. J., Weissman, S. M., Yackowski, L., Vaccari, E. M., Bissonnette, J., Booker, J. K., Cremona, M. L., Gibellini, F., Murphy, P. D., Pineda-Alvarez, D. E., Pollevick, G. D., Xu, Z., Richard, G., Bale, S., Klein, R. T., Hruska, K. S., & Chung, W. K. (2016). Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med, 18*(8), 823-832.

Suter, C. M., Martin, D. I., & Ward, R. L. (2004). Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet, 36*(5), 497-501. doi:10.1038/ng1342

T

Tabori, U., Hansford, J. R., Achatz, M. I., Kratz, C. P., Plon, S. E., Frebourg, T., & Brugieres, L. (2017). Clinical Management and Tumor Surveillance Recommendations of Inherited Mismatch Repair Deficiency in Childhood. *Clin Cancer Res, 23*(11), e32-e37.

Taeubner, J., Wimmer, K., Muleris, M., Lascols, O., Colas, C., Fauth, C., Brozou, T., Felsberg, J., Riemer, J., Gombert, M., Ginzel, S., Hoell, J. I., Borkhardt, A., & Kuhlen, M. (2018). Diagnostic challenges in a child with early onset desmoplastic medulloblastoma and homozygous variants in MSH2 and MSH6. *Eur J Hum Genet, 26*(3), 440-444. doi:10.1038/s41431-017-0071-5

Takahashi, M., Shimodaira, H., Andreutti-Zaugg, C., Iggo, R., Kolodner, R. D., & Ishioka, C. (2007). Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. *Cancer Res, 67*(10), 4595-4604.

Tavtigian, S. V., Greenblatt, M. S., Harrison, S. M., Nussbaum, R. L., Prabhu, S. A., Boucher, K. M., Biesecker, L. G., & ClinGen Sequence Variant Interpretation Working, G. (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med, 20*(9), 1054-1060. doi:10.1038/gim.2017.210

Ten Broeke, S. W., van der Klift, H. M., Tops, C. M. J., Aretz, S., Bernstein, I., Buchanan, D. D., de la Chapelle, A., Capella, G., Clendenning, M., Engel, C., Gallinger, S., Gomez Garcia, E., Figueiredo, J. C., Haile, R., Hampel, H. L., Hopper, J. L., Hoogerbrugge, N., von Knebel Doeberitz, M., Le Marchand, L., Letteboer, T. G. W., Jenkins, M. A., Lindblom, A., Lindor, N. M., Mensenkamp, A. R., Moller, P., Newcomb, P. A., van Os, T. A. M., Pearlman, R., Pineda, M., Rahner, N., Redeker, E. J. W., Olderode-Berends, M. J. W., Rosty, C., Schackert, H. K., Scott, R., Senter, L., Spruijt, L., Steinke-Lange, V., Suerink, M., Thibodeau, S., Vos, Y. J., Wagner, A., Winship, I., Hes, F. J., Vasen, H. F. A., Wijnen, J. T., Nielsen, M., & Win, A. K. (2018). Cancer Risks for PMS2-Associated Lynch Syndrome. *J Clin Oncol, 36*(29), 2961-2968. doi:10.1200/JCO.2018.78.4777

The Cancer Genome Atlas, N., Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., Wheeler, D. A., Gibbs, R. A., Lawrence, M. S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E. S., Gabriel, S., Getz, G., Ding, L., Fulton, R. S., Koboldt, D. C., Wylie, T., Walker, J., Dooling, D. J., Fulton, L., Delehaunty, K. D., Fronick, C. C., Demeter, R., Mardis, E. R., Wilson, R. K., Chu, A., Chun, H.-J. E., Mungall, A. J., Pleasance, E., Gordon Robertson, A., Stoll, D., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chuah, E., Coope, R. J. N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R. A., Lee, D., Li, H. I., Mayo, M., Moore, R. A., Schein, J. E., Slobodan, J. R., Tam, A., Thiessen, N., Varhol, R., Zeng, T., Zhao, Y., Jones, S. J. M., Marra, M. A., Bass, A. J., Ramos, A. H., Saksena, G., Cherniack, A. D., Schumacher, S. E., Tabak, B., Carter, S. L., Pho, N. H., Nguyen, H., Onofrio, R. C., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Meyerson, M., Protopopov, A., Zhang, J., Hadjipanayis, A., Lee, E., Xi, R., Yang, L., Ren, X., Zhang, H., Sathiamoorthy, N., Shukla, S., Chen, P.-C., Haseley, P., Xiao, Y., Lee, S., Seidman, J., Chin, L., Park, P. J., Kucherlapati, R., Todd Auman, J., Hoadley, K. A., Du, Y., Wilkerson, M. D., Shi, Y., Liquori, C., Meng, S., Li, L., Turman, Y. J., Topal, M. D., Tan, D., Waring, S., Buda, E., Walsh, J., Jones, C. D., Mieczkowski, P. A., Singh, D., Wu, J., Gulabani, A., Dolina, P., Bodenheimer, T., Hoyle, A. P., Simons, J. V., Soloway, M., Mose, L. E., Jefferys, S. R., Balu, S., O'Connor, B. D., Prins, J. F., Chiang, D. Y., Neil Hayes, D., Perou, C. M., Hinoue, T., Weisenberger, D. J., Maglinte, D. T., Pan, F., Berman, B. P., Van Den Berg, D. J., Shen, H., Triche Jr, T., Baylin, S. B., Laird, P. W., Getz, G., Noble, M., Voet, D., Saksena, G., Gehlenborg, N., DiCara, D., Zhang, J., Zhang, H., Wu, C.-J., Yingchun Liu, S., Shukla, S., Lawrence, M. S., Zhou, L., Sivachenko, A., Lin, P., Stojanov,

P., Jing, R., Park, R. W., Nazaire, M.-D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P. J., Chin, L., Thorsson, V., Reynolds, S. M., Bernard, B., Kreisberg, R., Lin, J., Iype, L., Bressler, R., Erkkilä, T., Gundapuneni, M., Liu, Y., Norberg, A., Robinson, T., Yang, D., Zhang, W., Shmulevich, I., de Ronde, J. J., Schultz, N., Cerami, E., Ciriello, G., Goldberg, A. P., Gross, B., Jacobsen, A., Gao, J., Kaczkowski, B., Sinha, R., Arman Aksoy, B., Antipin, Y., Reva, B., Shen, R., Taylor, B. S., Chan, T. A., Ladanyi, M., Sander, C., Akbani, R., Zhang, N., Broom, B. M., Casasent, T., Unruh, A., Wakefield, C., Hamilton, S. R., Craig Cason, R., Baggerly, K. A., Weinstein, J. N., Haussler, D., Benz, C. C., Stuart, J. M., Benz, S. C., Zachary Sanborn, J., Vaske, C. J., Zhu, J., Szeto, C., Scott, G. K., Yau, C., Ng, S., Goldstein, T., Ellrott, K., Collisson, E., Cozen, A. E., Zerbino, D., Wilks, C., Craft, B., Spellman, P., Penny, R., Shelton, T., Hatfield, M., Morris, S., Yena, P., Shelton, C., Sherman, M., Paulauskis, J., Gastier-Foster, J. M., Bowen, J., Ramirez, N. C., Black, A., Pyatt, R., Wise, L., White, P., Bertagnolli, M., Brown, J., Chan, T. A., Chu, G. C., Czerwinski, C., Denstman, F., Dhir, R., Dörner, A., Fuchs, C. S., Guillem, J. G., Iacocca, M., Juhl, H., Kaufman, A., Kohl Iii, B., Van Le, X., Mariano, M. C., Medina, E. N., Meyers, M., Nash, G. M., Paty, P. B., Petrelli, N., Rabeno, B., Richards, W. G., Solit, D., Swanson, P., Temple, L., Tepper, J. E., Thorp, R., Vakiani, E., Weiser, M. R., Willis, J. E., Witkin, G., Zeng, Z., Zinner, M. J., Zornig, C., Jensen, M. A., Sfeir, R., Kahn, A. B., Chu, A. L., Kothiyal, P., Wang, Z., Snyder, E. E., Pontius, J., Pihl, T. D., Ayala, B., Backus, M., Walton, J., Whitmore, J., Baboud, J., Berton, D. L., Nicholls, M. C., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P. A., Alonso, S., Sanbhadti, R. N., Barletta, S. P., Greene, J. M., Pot, D. A., Mills Shaw, K. R., Dillon, L. A. L., Buetow, K., Davidsen, T., Demchok, J. A., Eley, G., Ferguson, M., Fielding, P., Schaefer, C., Sheth, M., Yang, L., Guyer, M. S., Ozenberger, B. A., Palchik, J. D., Peterson, J., Sofia, H. J., & Thomson, E. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature, 487*, 330. doi:10.1038/nature11252

Thompson, B. A., Goldgar, D. E., Paterson, C., Clendenning, M., Walters, R., Arnold, S., Parsons, M. T., Michael, D. W., Gallinger, S., Haile, R. W., Hopper, J. L., Jenkins, M. A., Lemarchand, L., Lindor, N. M., Newcomb, P. A., Thibodeau, S. N., Young, J. P., Buchanan, D. D., Tavtigian, S. V., & Spurdle, A. B. (2013). 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, 34*(1), 200-209.

Thompson, B. A., Greenblatt, M. S., Vallee, M. P., Herkert, J. C., Tessereau, C., Young, E. L., Adzhubey, I. A., Li, B., Bell, R., Feng, B., Mooney, S. D., Radivojac, P., Sunyaev, S. R., Frebourg, T., Hofstra, R. M., Sijmons, R. H., Boucher, K., Thomas, A., Goldgar, D. E., Spurdle, A. B., & Tavtigian, S. V. (2013). Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat, 34*(1), 255-265. doi:10.1002/humu.22214

Thompson, B. A., Martins, A., & Spurdle, A. B. (2015). A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays. *Clin Genet, 87*(2), 100-108. doi:10.1111/cge.12450

Thompson, B. A., Spurdle, A. B., Plazzer, J. P., Greenblatt, M. S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capella, G., den Dunnen, J. T., du Sart, D., Fabre, A., Farrell, M. P., Farrington, S. M., Frayling, I. M., Frebourg, T., Goldgar, D. E., Heinen, C. D., Holinski-Feder, E., Kohonen-Corish, M., Robinson, K. L., Leung, S. Y., Martins, A., Moller, P., Morak, M., Nystrom, M., Peltomaki, P., Pineda, M., Qi, M., Ramesar, R., Rasmussen, L. J., Royer-Pokora, B., Scott, R. J., Sijmons, R., Tavtigian, S. V., Tops, C. M., Weber, T., Wijnen, J., Woods, M. O., Macrae, F., & Genuardi, M. (2014). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet, 46*(2), 107-115.

Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat, 32*(4), 358-368. doi:10.1002/humu.21445

Tournier, I., Vezain, M., Martins, A., Charbonnier, F., Baert-Desurmont, S., Olschwang, S., Wang, Q., Buisine, M. P., Soret, J., Tazi, J., Frebourg, T., & Tosi, M. (2008). A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat, 29*(12), 1412-1424. doi:10.1002/humu.20796

Traverso, G., Shuber, A., Levin, B., Johnson, C., Olsson, L., Schoetz, D. J., Jr., Hamilton, S. R., Boynton, K., Kinzler, K. W., & Vogelstein, B. (2002). Detection of APC mutations in fecal DNA from patients with colorectal tumors. *N Engl J Med, 346*(5), 311-320. doi:10.1056/NEJMoa012294

Traverso, G., Shuber, A., Olsson, L., Levin, B., Johnson, C., Hamilton, S. R., Boynton, K., Kinzler, K. W., & Vogelstein, B. (2002). Detection of proximal colorectal cancers through analysis of faecal DNA. *Lancet, 359*(9304), 403-404. doi:10.1016/S0140-6736(02)07591-8

Tricarico, R., Kasela, M., Mareni, C., Thompson, B. A., Drouet, A., Staderini, L., Gorelli, G., Crucianelli, F., Ingrosso, V., Kantelinen, J., Papi, L., De Angioletti, M., Berardi, M., Gaildrat, P., Soukarieh, O., Turchetti, D., Martins, A., Spurdle, A. B., Nystrom, M., Genuardi, M., & InSi, G. H. T. V. I. C. (2017). Assessment of the InSiGHT Interpretation Criteria for the Clinical Classification of 24 MLH1 and MSH2 Gene Variants. *Hum Mutat, 38*(1), 64-77. doi:10.1002/humu.23117

U

Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Ruschoff, J., Fishel, R., Lindor, N. M., Burgart, L. J., Hamelin, R., Hamilton, S. R., Hiatt, R. A., Jass, J., Lindblom, A., Lynch, H. T., Peltomaki, P., Ramsey, S. D., Rodriguez-Bigas, M. A., Vasen, H. F., Hawk, E. T., Barrett, J. C., Freedman, A. N., & Srivastava, S. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst, 96*(4), 261-268. doi:10.1093/jnci/djh034

V

Valle, L., Vilar, E., Tavtigian, S. V., & Stoffel, E. M. (2019). Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. *J Pathol, 247*(5), 574-588. doi:10.1002/path.5229

van der Klift, H. M., Jansen, A. M., van der Steenstraten, N., Bik, E. C., Tops, C. M., Devilee, P., & Wijnen, J. T. (2015). Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med, 3*(4), 327-345.

van der Klift, H. M., Mensenkamp, A. R., Drost, M., Bik, E. C., Vos, Y. J., Gille, H. J., Redeker, B. E., Tiersma, Y., Zonneveld, J. B., Garcia, E. G., Letteboer, T. G., Olderode-Berends, M. J., van Hest, L. P., van Os, T. A., Verhoef, S., Wagner, A., van Asperen, C. J., Ten Broeke, S. W., Hes, F. J., de Wind, N., Nielsen, M., Devilee, P., Ligtenberg, M. J., Wijnen, J. T., & Tops, C. M. (2016). Comprehensive Mutation Analysis of PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional Mismatch Repair Deficiency Syndrome. *Hum Mutat*. *37*(11):1162-1179.

van der Klift, H. M., Tops, C. M., Bik, E. C., Boogaard, M. W., Borgstein, A. M., Hansson, K. B., Ausems, M. G., Gomez Garcia, E., Green, A., Hes, F. J., Izatt, L., van Hest, L. P., Alonso, A. M., Vriends, A. H., Wagner, A., van Zelst-Stams, W. A., Vasen, H. F., Morreau, H., Devilee, P., & Wijnen, J. T. (2010). Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. *Hum Mutat, 31*(5), 578-587.

Vargas-Parra, G. M., Gonzalez-Acosta, M., Thompson, B. A., Gomez, C., Fernandez, A., Damaso, E., Pons, T., Morak, M., Del Valle, J., Iglesias, S., Velasco, A., Solanes, A., Sanjuan, X., Padilla, N., de la Cruz, X., Valencia, A., Holinski-Feder, E., Brunet, J., Feliubadalo, L., Lazaro, C., Navarro, M., Pineda, M., & Capella, G. (2017). Elucidating the molecular basis of MSH2-deficient tumors by combined germline and somatic analysis. *Int J Cancer, 141*(7), 1365-1380. doi:10.1002/ijc.30820

Vasen, H. F., Blanco, I., Aktan-Collan, K., Gopie, J. P., Alonso, A., Aretz, S., Bernstein, I., Bertario, L., Burn, J., Capella, G., Colas, C., Engel, C., Frayling, I. M., Genuardi, M., Heinimann, K., Hes, F. J., Hodgson, S. V., Karagiannis, J. A., Lalloo, F., Lindblom, A., Mecklin, J. P., Moller, P., Myrhoj, T., Nagengast, F. M., Parc, Y., Ponz de Leon, M., Renkonen-Sinisalo, L., Sampson, J. R., Stormorken, A., Sijmons, R. H., Tejpar, S., Thomas, H. J., Rahner, N., Wijnen, J. T., Jarvinen, H. J., & Moslein, G. (2013). Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut, 62*(6), 812-823.

Vasen, H. F., Ghorbanoghli, Z., Bourdeaut, F., Cabaret, O., Caron, O., Duval, A., Entz-Werle, N., Goldberg, Y., Ilencikova, D., Kratz, C. P., Lavoine, N., Loeffen, J., Menko, F. H., Muleris, M., Sebille, G., Colas, C., Burkhardt, B., Brugieres, L., & Wimmer, K. (2014). Guidelines for surveillance of individuals with constitutional mismatch repair-

deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D). *J Med Genet, 51*(5), 283-293.

Vasen, H. F., Mecklin, J. P., Khan, P. M., & Lynch, H. T. (1991). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum, 34*(5), 424-425. doi:10.1007/bf02053699

Vasen, H. F., Moslein, G., Alonso, A., Bernstein, I., Bertario, L., Blanco, I., Burn, J., Capella, G., Engel, C., Frayling, I., Friedl, W., Hes, F. J., Hodgson, S., Mecklin, J. P., Moller, P., Nagengast, F., Parc, Y., Renkonen-Sinisalo, L., Sampson, J. R., Stormorken, A., & Wijnen, J. (2007). Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J Med Genet, 44*(6), 353-362. doi:10.1136/jmg.2007.048991

Vasen, H. F., Watson, P., Mecklin, J. P., & Lynch, H. T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology, 116*(6), 1453-1456. doi:10.1016/s0016-5085(99)70510-x

Vaughn, C. P., Robles, J., Swensen, J. J., Miller, C. E., Lyon, E., Mao, R., Bayrak-Toydemir, P., & Samowitz, W. S. (2010). Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. *Hum Mutat, 31*(5), 588-593.

W

Waalkes, A., Smith, N., Penewit, K., Hempelmann, J., Konnick, E. Q., Hause, R. J., Pritchard, C. C., & Salipante, S. J. (2018). Accurate Pan-Cancer Molecular Diagnosis of Microsatellite Instability by Single-Molecule Molecular Inversion Probe Capture and High-Throughput Sequencing. *Clin Chem, 64*(6), 950-958.

Wang, Q., Lasset, C., Desseigne, F., Frappaz, D., Bergeron, C., Navarro, C., Ruano, E., & Puisieux, A. (1999). Neurofibromatosis and early onset of cancers in hMLH1-deficient children. *Cancer Res, 59*(2), 294-297.

Ward, R. L., Dobbins, T., Lindor, N. M., Rapkins, R. W., & Hitchins, M. P. (2013). Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. *Genet Med, 15*(1), 25-35. doi:10.1038/gim.2012.91

Warren, J. J., Pohlhaus, T. J., Changela, A., Iyer, R. R., Modrich, P. L., & Beese, L. S. (2007). Structure of the human MutSalpha DNA lesion recognition complex. *Mol Cell, 26*(4), 579-592.

Warthin, A. S. (1913). Heredity with reference to carcinoma: as shown by the study of the cases examined in the pathological laboratory of the university of michigan, 1895-1913. *JAMA Internal Medicine, XII*(5), 546-555.

Waterfall, J. J., & Meltzer, P. S. (2015). Avalanching mutations in biallelic mismatch repair deficiency syndrome. *Nat Genet, 47*(3), 194-196. doi:10.1038/ng.3227

Whiley, P. J., de la Hoya, M., Thomassen, M., Becker, A., Brandao, R., Pedersen, I. S., Montagna, M., Menendez, M., Quiles, F., Gutierrez-Enriquez, S., De Leeneer, K., Tenes, A., Montalban, G., Tserpelis, D., Yoshimatsu, T., Tirapo, C., Raponi, M., Caldes, T., Blanco, A., Santamarina, M., Guidugli, L., de Garibay, G. R., Wong, M., Tancredi, M., Fachal, L., Ding, Y. C., Kruse, T., Lattimore, V., Kwong, A., Chan, T. L., Colombo, M., De Vecchi, G., Caligo, M., Baralle, D., Lazaro, C., Couch, F., Radice, P., Southey, M. C., Neuhausen, S., Houdayer, C., Fackenthal, J., Hansen, T. V., Vega, A., Diez, O., Blok, R., Claes, K., Wappenschmidt, B., Walker, L., Spurdle, A. B., Brown, M. A., & consortium, E. (2014). Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin Chem, 60*(2), 341-352. doi:10.1373/clinchem.2013.210658

Wimmer, K., Beilken, A., Nustede, R., Ripperger, T., Lamottke, B., Ure, B., Steinmann, D., Reineke-Plaass, T., Lehmann, U., Zschocke, J., Valle, L., Fauth, C., & Kratz, C. P. (2017). A novel germline POLE mutation causes an early onset cancer prone syndrome mimicking constitutional mismatch repair deficiency. *Fam Cancer, 16*(1), 67-71.

Wimmer, K., Kratz, C. P., Vasen, H. F., Caron, O., Colas, C., Entz-Werle, N., Gerdes, A. M., Goldberg, Y., Ilencikova, D., Muleris, M., Duval, A., Lavoine, N., Ruiz-Ponte, C., Slavc, I., Burkhardt, B., & Brugieres, L. (2014). Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'care for CMMRD' (C4CMMRD). *J Med Genet, 51*(6), 355-365.

Wimmer, K., & Wernstedt, A. (2014). PMS2 gene mutational analysis: direct cDNA sequencing to circumvent pseudogene interference. *Methods Mol Biol, 1167*, 289-302. doi:10.1007/978-1-4939-0835-6_20

Win, A. K., Jenkins, M. A., Dowty, J. G., Antoniou, A. C., Lee, A., Giles, G. G., Buchanan, D. D., Clendenning, M., Rosty, C., Ahnen, D. J., Thibodeau, S. N., Casey, G., Gallinger, S., Le Marchand, L., Haile, R. W., Potter, J. D., Zheng, Y., Lindor, N. M., Newcomb, P. A., Hopper, J. L., & MacInnis, R. J. (2017). Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev, 26*(3), 404-412. doi:10.1158/1055-9965.EPI-16-0693

Y

Yamamoto, H., & Imai, K. (2015). Microsatellite instability: an update. *Arch Toxicol, 89*(6), 899-921. doi:10.1007/s00204-015-1474-0

Yurgelun, M. B., Allen, B., Kaldate, R. R., Bowles, K. R., Judkins, T., Kaushik, P., Roa, B. B., Wenstrup, R. J., Hartman, A. R., & Syngal, S. (2015). Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. *Gastroenterology, 149*(3), 604-613 e620.

Yurgelun, M. B., Goel, A., Hornick, J. L., Sen, A., Turgeon, D. K., Ruffin, M. T. t., Marcon, N. E., Baron, J. A., Bresalier, R. S., Syngal, S., Brenner, D. E., Boland, C. R., & Stoffel, E. M. (2012). Microsatellite instability and DNA mismatch repair protein deficiency in Lynch syndrome colorectal polyps. *Cancer Prev Res (Phila), 5*(4), 574-582. doi:10.1158/1940-6207.CAPR-11-0519

Yurgelun, M. B., & Hampel, H. (2018). Recent Advances in Lynch Syndrome: Diagnosis, Treatment, and Cancer Prevention. *Am Soc Clin Oncol Educ Book, 38*, 101-109. doi:10.1200/EDBK_208341

Yurgelun, M. B., Kulke, M. H., Fuchs, C. S., Allen, B. A., Uno, H., Hornick, J. L., Ukaegbu, C. I., Brais, L. K., McNamara, P. G., Mayer, R. J., Schrag, D., Meyerhardt, J. A., Ng, K., Kidd, J., Singh, N., Hartman, A. R., Wenstrup, R. J., & Syngal, S. (2017). Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J Clin Oncol, 35*(10), 1086-1095. doi:10.1200/JCO.2016.71.0012

Z

Zhu, L., Huang, Y., Fang, X., Liu, C., Deng, W., Zhong, C., Xu, J., Xu, D., & Yuan, Y. (2018). A Novel and Reliable Method to Detect Microsatellite Instability in Colorectal Cancer by Next-Generation Sequencing. *J Mol Diagn, 20*(2), 225-231. doi:10.1016/j.jmoldx.2017.11.007

Zuntini, R., Ferrari, S., Bonora, E., Buscherini, F., Bertonazzi, B., Grippa, M., Godino, L., Miccoli, S., & Turchetti, D. (2018). Dealing With BRCA1/2 Unclassified Variants in a Cancer Genetics Clinic: Does Cosegregation Analysis Help? *Front Genet, 9*, 378. doi:10.3389/fgene.2018.00378

ANEXO I. OTRAS PUBLICACIONES

ARTÍCULO 4

Comprehensive characterization of MLH1 p.D41H and p.N710D variants coexisting in a Lynch syndrome family with conserved MLH1 expression tumors

Marta Pineda*, **Maribel González-Acosta***, Bryony A. Thompson, Ricardo Sánchez, Carolina Gómez, Joaquín Martínez-López, José Perea, Pilar Garre, Trinidad Caldés, Yolanda Rodríguez, Stefania Landolfi, Judith Balmaña, Conxi Lázaro, Luis Robles, Gabriel Capellá‡, Daniel Rueda‡.

* Ambos autores han contribuido en igual medida a este trabajo y comparten autoría.
‡ Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Clinical Genetics, 2014. DOI: 10.1111/cge.12467

RESUMEN:

En este trabajo se describe una amplia familia que cumple con los criterios de Amsterdam I y que es portadora de dos VUS en el gen *MLH1*: c.121G> C (p.D41H) y c.2128A> G (p.N710D). Para dilucidar la importancia clínica de las VUS identificadas en *MLH1*, se utilizaron datos clínico-patológicos, el cálculo multifactorial de probabilidad y los resultados de los estudios funcionales.

Únicamente la variante c.121G>C presentó cosegregación con los tumores asociados a SL en la familia. Además, los tumores colorrectales diagnosticados presentaron inestabilidad de microsatélites pero la tinción inmunohistoquímica no reveló pérdida de expresión de las proteínas MMR. El cálculo multifactorial de probabilidad clasificó la variante c.2128A>G como no patogénica y a la variante c.121G>C como patogénica. Los ensayos funcionales revelaron tanto una actividad reparadora como una expresión disminuida para la variante c.121G>C. De acuerdo con los resultados, el residuo N710 se localiza en el dominio no conservado C-terminal de MLH1, mientras que el residuo D41 se ubica en el dominio ATPasa, altamente conservado.

Los resultados obtenidos permitirán el correcto asesoramiento genético de los portadores de las variantes c.121G>C y c.2128A>G y sus familiares, además de ejemplificar cómo la acumulación de datos y el análisis exhaustivo es indispensable para la clasificación de las variantes MMR.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2014

CLINICAL GENETICS
doi: 10.1111/cge.12467

Short Report

Detailed characterization of MLH1 p.D41H and p.N710D variants coexisting in a Lynch syndrome family with conserved MLH1 expression tumors

Pineda M, González-Acosta M, Thompson BA, Sánchez R, Gómez C, Martínez-López J, Perea J, Caldés T, Rodríguez Y, Landolfi S, Balmaña J, Lázaro C, Robles L, Capellá G, Rueda D. Detailed characterization of MLH1 p.D41H and p.N710D variants coexisting in a Lynch syndrome family with conserved MLH1 expression tumors.
Clin Genet 2014. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2014

Lynch syndrome (LS) is an autosomal dominant cancer-susceptibility disease caused by inactivating germline mutations in mismatch repair (MMR) genes. Variants of unknown significance (VUS) are often detected in mutational analysis of MMR genes. Here we describe a large family fulfilling Amsterdam I criteria carrying two rare VUS in the *MLH1* gene: c.121G > C (p.D41H) and c.2128A > G (p.N710D). Collection of clinico-pathological data, multifactorial analysis, *in silico* predictions, and functional analyses were used to elucidate the clinical significance of the identified *MLH1* VUS. Only the c.121G > C variant cosegregated with LS-associated tumors in the family. Diagnosed colorectal tumors were microsatellite unstable although immunohistochemical staining revealed no loss of MMR proteins expression. Multifactorial likelihood analysis classified c.2128A > G as a non-pathogenic variant and c.121G > C as pathogenic. *In vitro* functional tests revealed impaired MMR activity and diminished expression of c.121G > C. Accordingly, the N710 residue is located in the unconserved MLH1 C-terminal domain, whereas D41 is highly conserved and located in the ATPase domain. The obtained results will enable adequate genetic counseling of c.121G > C and c.2128A > G variant carriers and their families. Furthermore, they exemplify how cumulative data and comprehensive analyses are mandatory to refine the classification of MMR variants.

Conflict of interest

The authors have declared no conflicting interests.

M. Pineda[a,†], M. González-Acosta[a,†], B.A. Thompson[b], R. Sánchez[c], C. Gómez[a], J. Martínez-López[c], J. Perea[d], T. Caldés[e], Y. Rodríguez[f], S. Landolfi[g], J. Balmaña[g], C. Lázaro[a], L. Robles[h], G. Capellá[a,‡] and D. Rueda[c,‡]

[a]Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL, Hospitalet de Llobregat, Barcelona, Spain, [b]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Australia, [c]Laboratory of Molecular Biology, Hematology Division, [d]Department of Surgery, Doce de Octubre University Hospital, Madrid, Spain, [e]Molecular Oncology Laboratory, Hospital Clínico San Carlos, IdISSC, Madrid, Spain, [f]Department of Pathology, Doce de Octubre University Hospital, Madrid, Spain, [g]University Hospital of Vall d'Hebron, Barcelona, Spain, and [h]Familial Cancer Counselling Unit, Oncology Division, Doce de Octubre University Hospital, Madrid, Spain

[†]These authors contributed equally to this work and share first authorship.

[‡]These authors contributed equally to this work and share senior authorship.

Key words: HNPCC – Lynch syndrome – MLH1 – variants of unknown significance

Corresponding author: Dr Daniel Rueda, Laboratory of Molecular Biology, Hematology Division, Doce de

Octubre University Hospital, Av. de Andalucía s/n, 28041-Madrid, Spain.
Tel.: +34 91 779 27 88;
e-mail: druedafer.hdoc@gmail.com

Received 7 April 2014, revised and accepted for publication 22 July 2014

Lynch syndrome (LS) (OMIM #120435) is an autosomal dominant genetic condition that increases the risk of colorectal (CRC), endometrial and other characteristic tumors. It is molecularly defined by germline mutations and epimutations that inactivate one of the DNA mismatch repair (MMR) genes: *MLH1*, *MSH2*, *MSH6* and *PMS2*. Consequently, most of the tumors of the LS spectrum exhibit a microsatellite instability (MSI) mutator phenotype and loss of expression of MMR proteins (1).

As LS accounts for approximately 2% of all CRC, genetic testing of MMR genes is recommended when MMR deficiency is suspected, based on familial aggregation and/or histological or molecular evidence. The detection of a pathogenic mutation allows the diagnosis of LS and the appropriate management of patients and their families (2).

In routine diagnosis MMR DNA variants of unknown significance (VUS) are often identified, precluding LS diagnosis for carriers and their relatives (3). To facilitate classification of MMR VUS in terms of pathogenicity, quantitative and qualitative algorithms have been developed (4–8). Recently, the International Society for Gastrointestinal Hereditary Tumors (InSiGHT) developed and applied a standardized classification scheme for MMR variants, based on multiple lines of evidence including clinical and functional data (9). Variants were classified according to the five class IARC scheme as pathogenic (class 5), likely pathogenic (class 4), uncertain (class 3), likely non-pathogenic (class 2) and non-pathogenic (class 1) (3). However, an important proportion (~30%) of variants remains as class 3. Therefore, further information on clinico-pathological, familial and functional data of a given VUS is highly valuable in order to finally establish the appropriate management of carrier individuals and their families.

Here we present a large Spanish family fulfilling Amsterdam I criteria in which c.121G > C (p.D41H) and c.2128A > G (p.N710D) *MLH1* variants coexisted. We aimed at determining the pathogenicity of both *MLH1* variants using a comprehensive characterization.

Patients and methods

Patients, MMR mutational analysis and co-segregation analysis

We identified a large Spanish family fulfilling Amsterdam I criteria. Clinical and pathological data of affected individuals were validated (Table 1). The Internal Ethics Committee of the participant hospitals approved this study and the patients enrolled gave written informed consent. Mutational screening of MMR genes was performed as described in Appendix S1, Supporting Information.

The identified variants were searched in the NHLBI Exome Sequencing Project (ESP) database (http://evs.gs.washington.edu/EVS) and screened in Spanish population cohorts, as described in Appendix S1. DNA samples from relatives were screened for the two identified *MLH1* variants by Sanger sequencing.

Multifactorial likelihood analysis and bioinformatic analyses

Multifactorial likelihood analysis was conducted as described (5). Bioinformatic analyses were performed to evaluate the impact of the *MLH1* variants on transcription, protein function, protein structure and evolutionary conservation. See Appendix S1 for details.

Functional analyses of *MLH1* c.121G > C (p.D41H) variant

At the RNA level, the effect of the variant on transcript splicing and stability was evaluated using RNA extracted from cultured carrier lymphocytes. At the protein level, MLH1 p.D41H cloned in the pcDNA3.1 vector was transfected into HEK293T cells. Protein extracts were used to evaluate the *in vitro* MMR activity and MLH1/PMS2 expression. See Appendix S1 for details.

Statistical analysis

Significant differences between groups were analyzed using the non-parametric Mann–Whitney *U* test for quantitative data. All reported p values are two sided, and $p < 0.05$ was considered significant. All calculations were performed using SPSS 19.0 (IBM, Armonk, NY).

Results

We identified a family fulfilling Amsterdam I criteria. Initially only two individuals were accessible (III:8 and IV:3) (Fig. 1). Patient III:8 was diagnosed with CRC and bladder carcinoma at age 35 and 50, respectively. Patient IV:3 was affected by CRC at age 30. Both colorectal tumors were MSI but retained the expression of MMR proteins (Table 1, Fig. S1). In *MLH1* gene, c.121G > C (p.D41H) was identified in both patients, and c.2128A > G (p.N710D) only in patient III:8. No germline mutations were identified in *MSH2* and *MSH6* genes. While *MLH1* c.121G > C had been reported in a patient affected by CRC at age 32 and suggestive family history (10), c.2128A > G variant had not been previously identified. In order to elucidate their

Table 1. Clinical characteristics of cancer affected family members and *MLH1* gene variant carriers

								IHC					MLH1 genetic test	
Family member	Gender	Age	Tumor type	Age of onset	Tumour location	Tumour stage	MSI	MLH1	MSH2	MSH6	PMS2	MLH1 methyl	c.121G>C (p.D41H)	c.2128A>G (p.N710D)
II:1	F	39†	CRC	34									*	
II:2	F	49 †	CRC	47									*	
II:3	F	†	CRC	48									*	
II:4	F	†	CRC	52										
II:5	M	†	CRC	43									*	
III:1	F	40 †	OC	38		IB-IIB							*	
			CRC	39	L	B2								
III:3	F	62	CRC	61		pT3N0M0							+	–
III:5	M	53	CRC	31	SpF		+	+	+	+	+		+	
III:6	M	49	–										+	
III:7	M	52	–										+	
III:8	M	56	CRC	35	Tr	pT2N2/G3	+	+	+	+	+	–	+	+
			BLC	50		pT1/G2								
III:9	F	54	–										–	+
III:10	F	62	–										–	+
III:13	M	47	CRC	43	R	pT4N0M0/G2	+	+	+	+	+	–	+	–
III:14	F	48	EC	45			–	+	+	+	+		+	
IV:1	F	36	CRC	36									+	–
IV:3	M	35	CRC	30	Rc		+	+	+	+	+		+	–
IV:6	F	36	–										+	–

Gender: M, male; F, female; †age at death; Tumor type: CRC, colorectal cancer; EC, endometrial cancer; OC, ovarian cancer; BLC, bladder cancer; Tumour Location: R, right colon; L, left colon; Rc, rectum; SpF, splenic flexure; Tr, transverse; Tumour stage: TNM/Dukes; MSI, microsatellite instability: +, instable; –, stable; IHC, immunohistochemical analysis of MMR proteins in tumor tissue: + conserved expression; MLH1 promoter methylation: – , non-methylated; MLH1 genetic test: +, carrier; –, non-carrier; *, obligated carrier.

pathogenicity, a comprehensive study based on collection of clinico-pathological data and functional analyses was undertaken.

Clinico-pathological data revealed a large family tree with 13 members diagnosed with LS-related neoplasms (Fig. 1). The four available CRC tumors showed normal staining of MMR proteins and MSI phenotype. One endometrial cancer diagnosed at age 45 was microsatellite stable (Table 1, Fig. S1E). None of the analyzed tumors showed loss of the wild-type allele (Fig. S2). Co-segregation analysis revealed that 7 of 10 carriers of *MLH1* c.121G > C were affected and five deceased individuals were considered obligated carriers, resulting in odds of 3253:1 in favor of causality for the variant (Fig. 1 and Table S1). Variant c.2128A > G was located in *trans* with c.121G > C variant in patient III-8 and did not cosegregate with the disease (Fig. 1).

Multifactorial likelihood analysis using the collected clinico-pathological data determined that c.121G > C and c.2128A > G have a posterior probability of pathogenicity of >0.999 and 1.924E-06, and consequently would have been classified as pathogenic and non-pathogenic, respectively, following recent Insight recommendations (9) (Table S1).

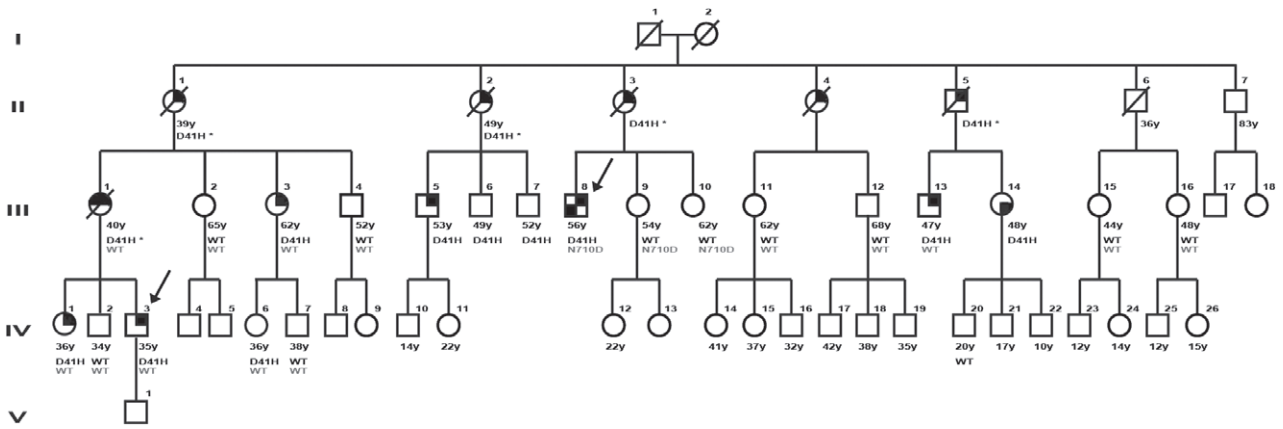


Fig. 1. Pedigree of the family under study. Current age (or age at death) and the result of the carrier status of the variants *MLH1* p.D41H (in black) or p.N710D (in gray) are indicated below the individual's symbol. Obligate carriers are indicated by an asterisk. Arrows indicate probands. Tumor types are represented as black sectors inside an individual's symbol: top right, colorectal cancer; top left, ovarian cancer; bottom right, endometrial cancer; bottom left, bladder cancer.

To further strengthen the evidence supporting variant classification, functional evaluation of *MLH1* c.121G > C and c.2128A > G were performed following our reported algorithm (6). None of the variants were described in the NHLBI ESP Database or identified in Spanish cohorts of control individuals and CRC cases. *In silico* tools did not predict any impact on splicing for both variants, while predictions on exonic splicing enhancers were inconclusive (Table S2). At the protein level, p.D41 is located at the highly conserved motif I of the MLH1 ATP binding domain (11) (Fig. 2A and Fig. S3A). p.D41H variant was stated as deleterious by four prediction programs (Table S2). Bioinformatic structure analysis predicted an increased ATP-MLH1 stability (pseudo-ΔΔG of 2.01 kcal/mol). As suggested in RNA pol II-NTP model (12), the 3.5 Å distance between histidine nitrogen NE2(H) and ATP molecule third phosphate oxygen O2G, might allow the formation of a hydrogen bond or a salt bridge in a pH dependent manner, that could hinder the ATPase cycle (Fig. 2A). On the other hand, p.N710D, predicted as neutral by *in silico* analysis (Table S2), is placed in the evolutionarily unconserved C-terminal region (Fig. S3B) precluding structural predictions.

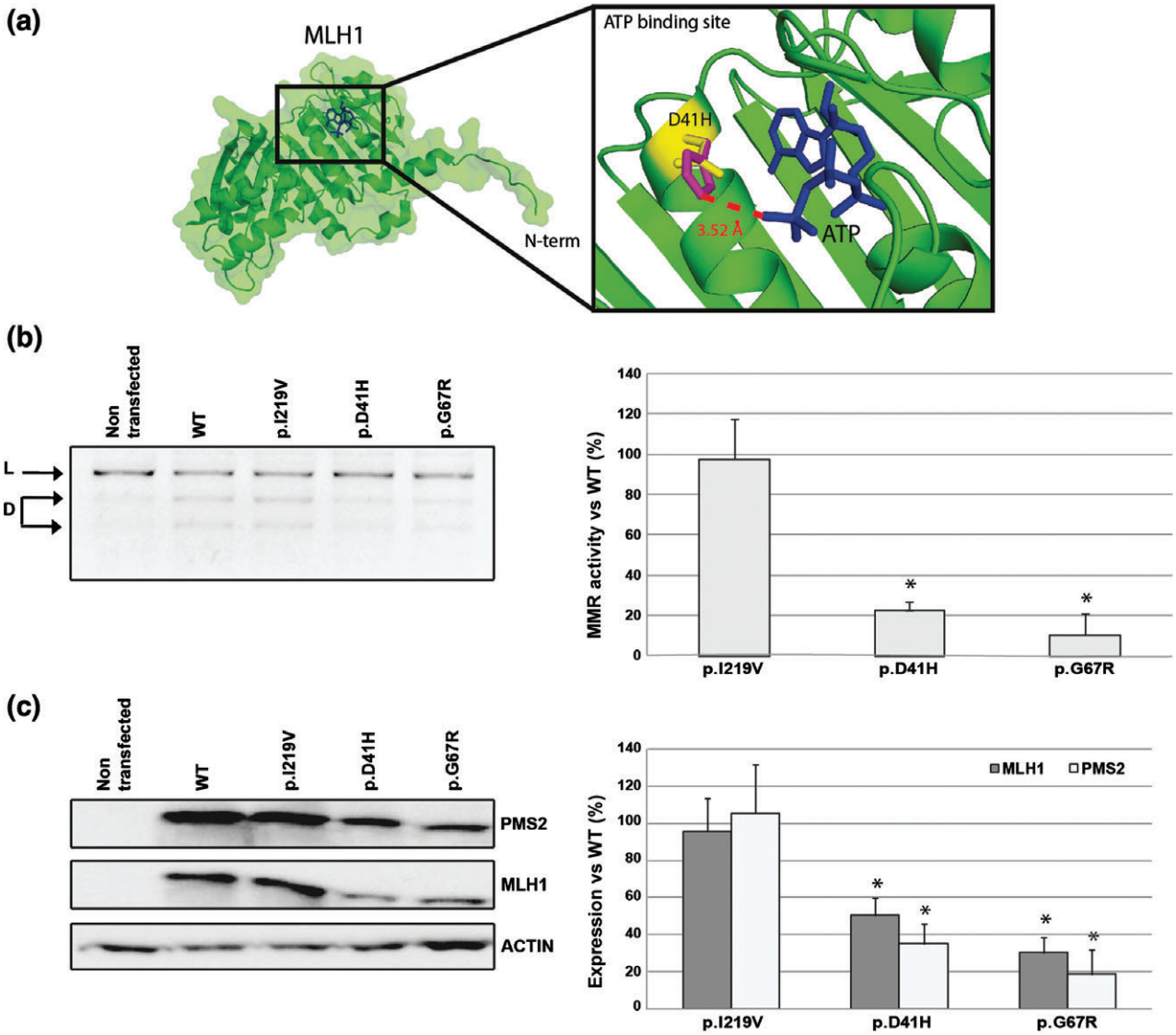


Fig. 2. Localization of D41H in the MLH1 structure, mismatch repair activity and protein expression levels. (**a**) Surface representation of the N-terminal domain (3-335, PDB 3NA3) of wild-type MLH1 superimposed with the p.D41H mutant. Zoom square, side chain of D41 (yellow) and H41 (magenta) are shown as sticks. ATP molecule is depicted as blue sticks. (**b**) Left panel: representative agarose gel showing digestion products of MMR assay. D, double-digested DNA; L, linear DNA. Right panel: quantification of repair levels of MLH1 variants in direct comparison to MLH1 wild-type. Data was obtained from three independent experiments. (**c**) Left panel: representative Western-blot analysis of MLH1 and PMS2. Right panel: relative expression of MLH1 or PMS2 is shown in dark and light gray columns, respectively. Data was obtained from three independent experiments. Statistically significant differences with the wild-type group are indicated (*$p < 0.05$). The pathogenic control c.199G > A (p.G67R) was practically inactive in MMR assays and showed a strong reduction of MLH1 and PMS2 expression levels, whereas the non-pathogenic control c.655A > C (p.I219V) was proficient in MMR activity and expression (8, 21), confirming the reliability of the technique.

Then, we focused on the functional analysis of c.121G > C (p.D41H). At the RNA level the variant did not affect transcript processing or stability (Fig. S4). At the protein level, p.D41H significantly impaired MMR activity *in vitro* (22.75% ± 5.77 of the wild-type level) (Fig. 2B). Moreover, transfection into HEK293T cells resulted in diminished MLH1 and PMS2 expression (50.39% ± 9.41 and 35.33% ± 10.54, respectively) (Fig. 2C).

Discussion

Here we report the comprehensive characterization of *MLH1* c.121G > C (p.D41H) and c.2128A > G (p.N710D) variants identified in a large Spanish family. Cumulative evidence obtained from clinicopathological, co-segregation and functional data allowed their conclusive classification (Table S3). *MLH1* c.2128A>G has been classified as not pathogenic by multifactorial analysis. It was found in *trans* with the c.121G > C mutation in III:8, who is currently aged 56 and exhibiting no evidence of Constitutional MMR-deficiency syndrome (OMIM #276300). Conversely, *MLH1* c.121G > C has a posterior probability of pathogenicity of 0.999 based on multifactorial likelihood analysis, in agreement with the obtained functional results showing impaired MMR activity.

MLH1 c.121G > C was recently classified as probably pathogenic using multifactorial likelihood analysis (InSiGHT classification criteria, version 1.9) (9), based on the existence of one MSI tumor in a carrier (10), in the absence of functional or co-segregation evidence (Table S1). Our comprehensive analysis allowed the reclassification of this variant from class 4 to class 5, providing evidence for the utility of the IARC recommendation to test relatives on a research basis for variants that are probably, but not yet definitely, pathogenic (3). Reclassification is also relevant because, although in classes 4 and 5 cases complete high-risk surveillance of carriers and testing in first-degree relatives is recommended (3), surveillance recommendations to class 4 variant non-carriers is controversial, because of the 1–5% residual likelihood that the variant may not be pathogenic. The reclassification of c.121G > C as pathogenic implies non-carriers can safely follow average risk surveillance recommendations.

The proficient transcription of *MLH1* c.121G > C variant correlates with the results obtained using *ex vivo* splicing reporter minigenes (13). At the protein level, impaired MMR activity was showed for p.D41H, supporting the validity of the multifactorial likelihood results for this variant. The observed functional defect is in agreement with structural analyses and is concordant with the analysis of another variant affecting the same residue, p.D41G, showing reduced MMR activity and expression (8, 14).

Conserved expression of MLH1 and PMS2 proteins by immunohistochemistry (IHC) and retention of wild-type allele has been observed in the analyzed tumors from p.D41H carriers [this study and (10)]. These results together with the increased ATP-MLH1 stability predicted for the variant protein suggest that it may act as a dominant negative mutant. Interestingly, other pathogenic missense variants located in the motif I of the ATP binding domain have shown the same MSI/IHC pattern (15–17). Moreover, one of these mutations (p.N38H) has been described as a dominant negative in *Bacillus subtilis* (18). Highlighting the complexity of the functional impact of these variants a diminished p.D41H protein expression was observed in transfection experiments as previously reported for p.N38H (18).

MMR proteins IHC testing is commonly used as a screening tool for identification of LS and as orientation for germline mutational analysis. It has showed equivalent informative value as MSI testing in predicting germline MMR mutations [reviewed in (19)]. However, as mentioned above, false-normal staining patterns have been reported (6, 15–17, 20) especially associated with catalytically inactive but stable *MLH1* missense mutations. Thus, as recommended by most LS clinical and diagnosis guides, MSI testing should be performed to detect MMR-deficiency in a context of clear clinical suspicion of LS but normal MMR proteins expression (2, 19).

Significant efforts have been made to reach a standardized classification scheme in the interpretation of pathogenicity evidence in MMR variants. The work presented here is an example of how novel and complementary data accumulated over time are critical to reclassify a MMR variant. Refining the genetic interpretation will enable an adequate genetic counseling of families with LS.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Acknowledgments

We are indebted to the patients and the members of the Hereditary Cancer Genetic Counseling Units. We acknowledge Anna Fernández for technical support and Adriana López-Doriga for statistical analysis support. This work was funded by the Spanish Ministry of Economy and Competitiveness (grant SAF2012-33636 and grant AES PI10/0683); the Scientific Foundation Asociación Española Contra el Cáncer; the Government of Catalonia (grant 2009SGR290), Fundación Mutua Madrileña (grant AP114252013) and RTICC MINECO Network RD12/0036/0031, RD12/0036/006 and RD12/0036/0008. M. G. is funded by the Fundación Mutua Madrileña. B. A. T. is supported by a Cancer Council of Queensland PhD Scholarship.

References

1. Boland CR, Lynch HT. The history of Lynch syndrome. Fam Cancer 2013: 12: 145–157.
2. Vasen HF, Blanco I, Aktan-Collan K et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 2013: 62: 812–823.
3. Plon SE, Eccles DM, Easton D et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 2008: 29: 1282–1291.
4. Pastrello C, Pin E, Marroni F et al. Integrated analysis of unclassified variants in mismatch repair genes. Genet Med 2011: 13: 115–124.

5. Thompson BA, Goldgar DE, Paterson C et al. 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 2013: 34: 200–209.
6. Borras E, Pineda M, Brieger A et al. Comprehensive functional assessment of MLH1 variants of unknown significance. Hum Mutat 2012: 33: 1576–1588.
7. Kansikas M, Kariola R, Nystrom M. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. Hum Mutat 2011: 32: 107–115.
8. Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G. Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis. Clin Cancer Res 2013: 19: 2432–2441.
9. Thompson B, Spurdle AB, Plazzer J-P et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 2014: 46: 107–115.
10. Bonnet D, Selves J, Toulas C et al. Simplified identification of Lynch syndrome: a prospective, multicenter study. Dig Liver Dis 2013: 44: 515–522.
11. Raschle M, Dufner P, Marra G, Jiricny J. 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: 21810–21820.
12. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell 2006: 127: 941–954.
13. 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: 1412–1424.
14. Ellison AR, Lofing J, Bitter GA. Human MutL homolog (MLH1) function in DNA mismatch repair: a prospective screen for missense mutations in the ATPase domain. Nucleic Acids Res 2004: 32: 5321–5338.
15. Hardt K, Heick SB, Betz B et al. Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. Fam Cancer 2011: 10: 273–284.
16. van Puijenbroek M, Middeldorp A, Tops CM et al. Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. Fam Cancer 2008: 7: 319–330.
17. van Riel E, Ausems MG, Hogervorst FB et al. A novel pathogenic MLH1 missense mutation, c.112A > C, p.Asn38His, in six families with Lynch syndrome. Hered Cancer Clin Pract 2010: 8: 7.
18. Bolz NJ, Lenhart JS, Weindorf SC, Simmons LA. Residues in the N-terminal domain of MutL required for mismatch repair in Bacillus subtilis. J Bacteriol 2012: 194: 5361–5367.
19. Shia J. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part I. The utility of immunohistochemistry. J Mol Diagn 2008: 10: 293–300.
20. Mangold E, Pagenstecher C, Friedl W et al. Tumours from MSH2 mutation carriers show loss of MSH2 expression but many tumours from MLH1 mutation carriers exhibit weak positive MLH1 staining. J Pathol 2005: 207: 385–395.
21. Raevaara TE, Korhonen MK, Lohi H et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology 2005: 129: 537–549.

SUPPLEMENTARY METHODS

MMR genes mutational analysis

Point mutation analysis of *MLH1*, *MSH*2 and *MSH6* genes was performed by High Resolution Melting (Roche) (primers and conditions available upon request). Samples with abnormal patterns were sequenced with BigDye Terminator Sequencing kit (Life Technologies). Genomic rearrangements in MMR genes were analyzed by multiplex ligation dependent probe amplification using SALSA-MLH1/MSH2 P003B1 and MSH6 P072 kits (MRC-Holland). Annotation of *MLH1* variants was done following the HGVS recommendations (RefSeqs NG_007109.2, NM_000249.3 and NP_000240.1).

Immunohistochemistry for DNA mismatch repair proteins

Immunohistochemistry was performed on 3-µm section slides from formalin-fixed, paraffin-embedded tissue, incubating with primary monoclonal antibodies against MLH1 (clone G168-15; Dilution 1:20; Menarini), MSH2 (clone FE11; Dilution 1:50; Menarini), MSH6 (clone BC/44 Dilution 1:70; Menarini) and PMS2 (clone A16-4, Dilution 1:200; BD Phamagen). Diaminobenzidine was used as chromogen. Normal positive DNA MMR protein expression was defined as nuclear staining within tumor cells, using adjacent normal non-neoplastic tissue on the same slide as positive internal control. Negative protein expression was defined as complete absence of nuclear staining within tumor cells. Results were confirmed on different sample block or slide from the same cancer.

Microsatellite instability analysis

Microsatellite instability (MSI) status was studied using MSI Analysis System v1.2 kit (Promega) following manufacturer recommendations. Briefly, the commercial kit uses five quasi monomorphic mononucleotide markers (BAT25, BAT26, NR21, NR24, and MONO27) for MSI determination and two polymorphic markers (PentaC and PentaD) for PBL / tumour sample matching. Fluorescent PCR products were resolved by capillary electrophoresis using an 3100 Genetic Analyzer (Applied Biosystems). Tumour samples with two or more of the five microsatellite markers instable were considered MSI positive.

Variant frequency in Spanish healthy controls and sporadic CRC cases

Screening of the *MLH1*c.121G>C variant was performed by conformation-sensitive capillary electrophoresis in a cohort of 304 controls and 324 CRC cases (1). Screening of the*MLH1*c.2128A>G variant was performed using High Resolution Melting analysis in 92

controls and 92 CRC cases. Samples with atypical profiles were PCR amplified and sequenced using BigDye Terminator Sequencing kit (Applied Biosystems).

Multifactorial likelihood analysis

Multifactorial likelihood analysis was based on estimated prior probabilities of pathogenicity and likelihood ratios for segregation and tumor characteristics (MSI phenotype and recruitment location) (2). Risk associated with each identified *MLH1* variant (c.121G>C and c.2128A>G) has been analyzed separately, under the assumption that only one variant was causal. Variants were classified according to the 5 class IARC quantitative scheme (3), based on the posterior probability.

Bioinformatic analyses

DNA sequences containing the identified *MLH1* variants were analyzed using several bioinformatic tools addressed to evaluate its impact at the RNA and protein level, as previously reported (1,4). Evolutionary conservation of variants was evaluated using a multialignment of MLH1 sequences of evolutionary divergent species on Align-GVGD (http://agvgd.iarc.fr/index.php). The structural effect of the MLH1 variants was evaluated *in silico* by means of the Site Directed Mutator (SDM) Server. PyMOL Molecular Graphics System v1.5.0.4 (Schrödinger, LLC) was used to visualize structures and to create Figure 2A. Input files were PDB file 3NA3 for MLH1 N-terminal domain and PDB file 3RBN for the C-terminal domain.

Lymphocyte culture, mRNA splicing analysis and allele specific expression analysis

Human lymphocytes from a c.121G>C (p.D41H) variant carrier (III-14) were cultured in the absence or presence of puromycin (Sigma). Total RNA was extracted from cultured lymphocytes and cDNA was synthesized as described (1). Amplification of *MLH1* exons 1-5 coding regionwas performed using primers 5'- TATCCAGCGGCCAGCTAA-3' and 5'-AGGGGCTTTCAGTTTTCCAT-3' (conditions available upon request). Sequences of carrier transcripts were compared with transcripts from three control lymphocyte cultures.
Allele specific expression (ASE) was analysed by SNuPE(1). 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 quadruplicate.

Plasmids and Site-Directed Mutagenesis

pcDNA3.1_MLH1 and pSG5_PMS2 plasmids, kindly provided by Dr. Kolodner and Dr. Nyström-Lahti, were used in MMR assays and expression analyses. The MLH1 missense variant p.D41H

(c.121G>C) was constructed by site-directed mutagenesis using the QuikChange Site Directed Mutagenesis Kit (Agilent Technologies, La Jolla, CA) using the following primers: 5'-GATTGAGAACTGTTTACATGCAAAATCCACAAG-3' and 5'-CTTGTGGATTTTGCATGTAAACAGTTCTCAATC-3',according to manufacturer's instructions. Sequencing was used to verify the presence of the variant. In addition, two control plasmids were constructed: p.G67R (c.199G>A) used as a pathogenic mutation control and p.I219V (c.655A>C) used as a neutral control.

HEK293T Cells Culture and Cell Transfection

Transfection of HEK293T cells (deficient for endogenous MLH1 and PMS2) was carried out as described (5). In brief, HEK293T cells were transfected at 30–40% confluence with MLH1 and PMS2 expression plasmids (3 mg/ml, respectively) and 0.5 mg/ml of pGFP, as a transfection control, using 20 µl/ml of the cationic polymer polyethylenimine (Polysciences, Warrington, Pennsylvania, USA; stock solution 1 mg/ml). After 48h, cells were prepared for protein extraction and cytometer analysis.

MMR activity assay

MMR assays were performed as described (5). In short, the reaction was performed in 15 µl total volume with reaction buffer (25 mM Tris–HCl pH 7.5, 110 mM KCl, 5 mM MgCl2, 50 µg/ml BSA, 1.5 mM ATP, 0.1 mM each dNTP), 50 ng DNA mismatched substrate pUC19CPDC, 50 µg nuclear extract of HEK293T cells, which are deficient in mismatch repair, and 5 µg protein extract from transfected HEK293T cells. Reactions were incubated at 37ºC for 15 min and terminated with 25 µl stop-buffer (100 mM EDTA, 10% SDS, 20 mg/ml proteinase K) by an additional incubation for 10 min at 37ºC. Plasmids were extracted from the reaction mixture by phenol-chloroform extraction and purified by ethanol co-precipitation with tRNA. Subsequent digestion with AseI, EcoRV and RNAse A produced two smaller fragments besides the linearized vector when repair was successful. Restriction digests were separated ion 2% agarose gels. Bands intensity was quantified using QuantityOne Software v.4.4 (Bio-Rad). 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 by the value of a wild-type protein that had been expressed, processed and tested in parallel. Experiments were performed in triplicate.

MLH1 and PMS2 protein expression analysis

MLH1 and PMS2 expression levels in transfected HEK293T cells were examined by SDS-PAGE, followed by Western blotting analysis with anti-MLH1 (clone G168-15, BD Pharmigen) and anti-PMS2 (clone 16-4, BD Biosciencies) antibodies. Band intensities were quantified using

QuantityOne v.4.4 (BioRad). Alfa-actin expression was assessed in parallel and used as loading control. Expression of MLH1 and PMS2 was normalized to alfa-actin expression. The relative protein expression was calculated by dividing the normalized protein expression in variant-transfected cells by the expression in wildtype MLH1/PMS2-transfected cells, processed and tested in parallel. Protein expression analyses were performed in triplicate from 3 independent transfection experiments.

REFERENCES

1. Borras E, Pineda M, Brieger A et al. Comprehensive functional assessment of MLH1 variants of unknown significance. Hum Mutat 2012: 33: 1576-1588.
2. Thompson BA, Goldgar DE, Paterson C et al. 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 2013: 34: 200-209.
3. Plon SE, Eccles DM, Easton D et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 2008: 29: 1282-1291.
4. Borras E, Pineda M, Cadinanos J et al. Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. J Med Genet 2013: 50: 552-563.
5. 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: 6574-6586.

Supplementary Table S1. Multifactorial likelihood analysis. Abreviations: LR, likelihood ratio.

***MLH1* variant**	**Reference**	**Prior probability of pathogenicity**	**Ascertainment (tumor LR)**	**Total Tumor characteristics LR**	**Segregation LR**	**Odds for causality**	**Posterior probability of pathogenicity**	**Classification**
c.121G>C (p.D41H)	Bonnet et al., 2012	0,869477	population (6.96)	6,96		6,96000	0,978886938	**Probably pathogenic (Class 4)**
	This study	0,869477	clinic (8.66)	649,461896	3252,6325	2112460,87044	0,999999929	**Pathogenic (Class 5)**
c.2128A>G (p.N710D)	This study	0,003359→0,1	clinic (8.66)	8,66	0,000002	0,00001732	1,92E-06	**Neutral (Class 1)**

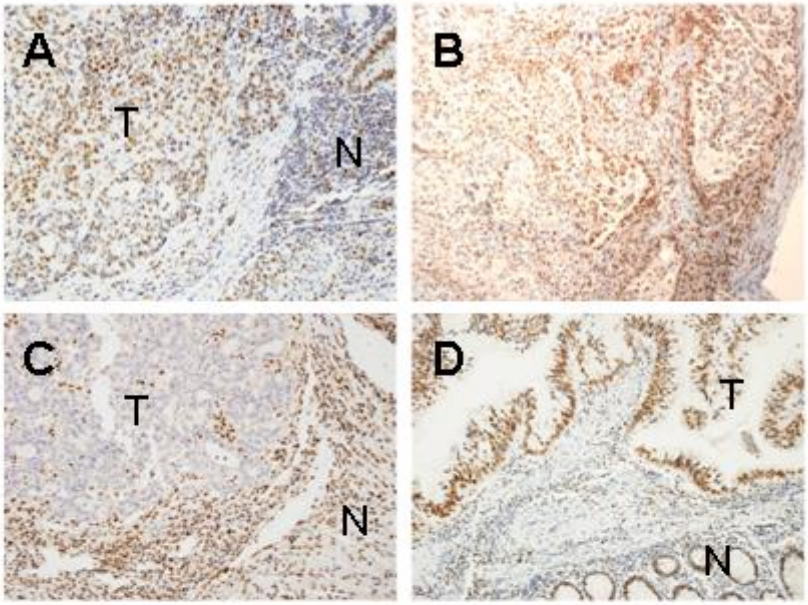
Supplementary Table S2. Bioinformatic predictions of the *MLH1* variants. Predictions are interpreted as inconclusive when the same results are not obtained by all the programs used. Abbreviations: SS, splice site; A, acceptor consensus splice site; D, donor consensus splice site; NR, consensus splice site not recognized.

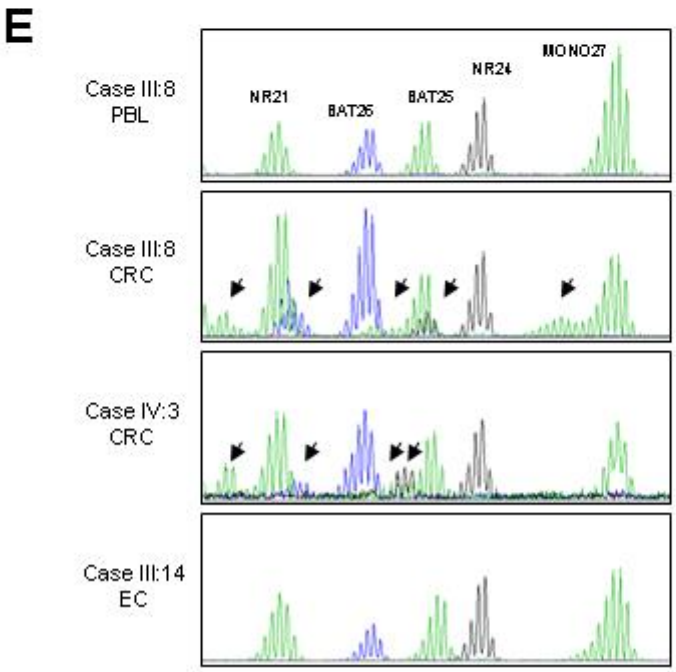
		Splicing Site Prediction									Exonic splicing enhancer prediction				Predicted impact on protein function				
		NNSplice		Spliceport		NetGene2		Softberry		Interpretation	Rescue ESE	ESE finder	PESX	Interpretation	PolyPhen-2 (score)	SIFT (score)	Condel (score)	MAP_MMR (score)	Interpretation
	SS	wild-type	variant	wild-type	variant	wild-type	variant	wild-type	variant										
c.121G>C (p.D41H)	A	0,86	0,88	-0.11	NR	0,07	0,07	4.175	4.175	**No effect**	2 sites destroyed	No change	No change	**Inconclusive**	Pr Damaging (1.00)	Damaging (2.52)	Deleterious (1.00)	Deleterious (20.32)	**Impaired**
	D	1	1	NR	NR	0	0	6.099	6.099										
c.2128A>G (p.N710D)	A	0.54	0.54	-0,35	-0,14	NR	NR	7.5	7.5	**No effect**	No change	3 sites created	–	**Inconclusive**	Benign (0.00)	Tolerated (0.64)	Neutral (0.001)	Neutral (2.45)	**Neutral**
	D	–	–	–	–	–	–	–	–										

Supplementary Table S3.Summary of results. Abbreviations: ND: not described; NP, not performed; SS, splice site; ESE, exonic splicing enhancer; wt, wild-type. (*) number of variant alleles / number of alleles in Spanish controls series.

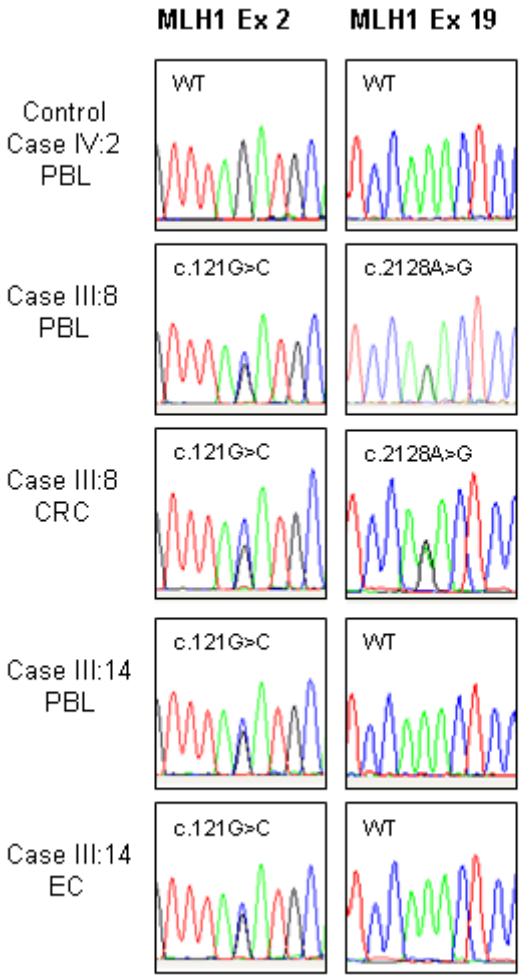
MLH1 variants	Predicted protein	Control frequency (our cohort* /ESP database)	Interpretation of mRNA predictions (SS/ESE)	mRNA analysis	cDNA stability analysis	Interpretation of protein predictions	MMR assay (% of the wt level)	MLH1 expression (% of the wt level)	PMS2 expression (% of the wt level)	Posterior probability of phatogenicity	Final classification
c.121G>C	p.D41H	(0/608) / ND	No effect / Inconclusive	r.121G>C	Non allelic imbalance	Impaired	22.75 ± 5.77	50.39 ± 9.41	35.33 ± 10.54	0,999999929	**Pathogenic (Class 5)**
c.2128A>G	p.N710D	(0/184) / ND	No effect / Inconclusive	NP	NP	Neutral	NP	NP	NP	1,92444E-06	**Neutral (Class 1)**

Supplementary Fig. S1. Immunohistochemical MLH1 staining and microsatellite instability analysis. **A)** Positive MLH1 nuclear staining was detected in colon carcinoma cells (T) and in adjacent normal colon mucosa (N) of individual III:8. **B)** Positive normal immunohistochemical MLH1 staining was also detected in case III:14 endometrial carcinoma cells and in adjacent endometrial stromal cells. **C)** Negative MLH1 protein staining in a control unrelated BRAF p.V600E mutated tumor. Colon carcinoma cells have negative nuclear staining (T) and in contrast to positive nuclear staining in normal mucosa (N). **D)** Normal MLH1 staining in a control unrelated microsatellite stable (MSS) colon carcinoma tumour (T) and adjacent normal colon mucosa (N) (x200 magnification). **E)** Electropherograms of microsatellite analysis of mononucleotide markers BAT26, BAT25, NR21, NR24 and MONO27. Arrows indicate microsatellite instability. Abreviations: PBL, peripheral blood lymphocytes; CRC, colorectal cancer; EC, endometrial cancer.





Supplementary Fig. S2. Sequence chromatogram of *MLH1* exon 2 and exon 19. Figure shows sequence chromatogram on positions covering *MLH1* c.121G>C and c.2128A>G in DNA extracted from PBL of individual IV:2, and from PBL and FFPE tumors of individuals III:8 and III:14. Abreviations: PBL, peripheral blood lymphocytes; CRC, colorectal cancer; EC, endometrial cancer.



Supplementary Fig. S3. Conservation of the aminoacid sequence by multialignment of the MLH1 protein sequence of different species. A) Sequence alignment of N-terminal MLH1. Asterisk (*) shows position of variant p.D41H. **B)** Sequence alignment of C-terminal MLH1. Asterisk (*) shows position of variant p.N710D. Acidic residues are shown on black background; basic residues are shown on white background; non polar aliphatic residues are shown on grey background; polar residues are shown with black font on a grey background; aromatic residues are shown on light greybackground. Abreviations: Homo sapiens (Hsap), Macacamulatta (monkey) (Mmul), Musmusculus (mouse) (Mmus), Canis lupus familiaris (dog) (Cfam), Monodelphisdomestica (gray short-tailed opossum - marsupialia) (Mdom), Gallus gallus (chicken - aves) (Ggal), Xenopuslaevis (frog - amphibia) (Xlae), Daniorerio (zebrafish - teleostei) (Drer). Cionaintestinalis (sea quirt - urochordata) (Cint), Branchiostomafloridae (lancelet - cephalochordate) (Bflo), Strongylocentrotuspurpuratus (purple sea urchin - echinodermata) (Spur), Trichoplaxadhearens (placozoa) (Tadh).Alignments are taken from the Align-GVGD website.

A

αA helix β1

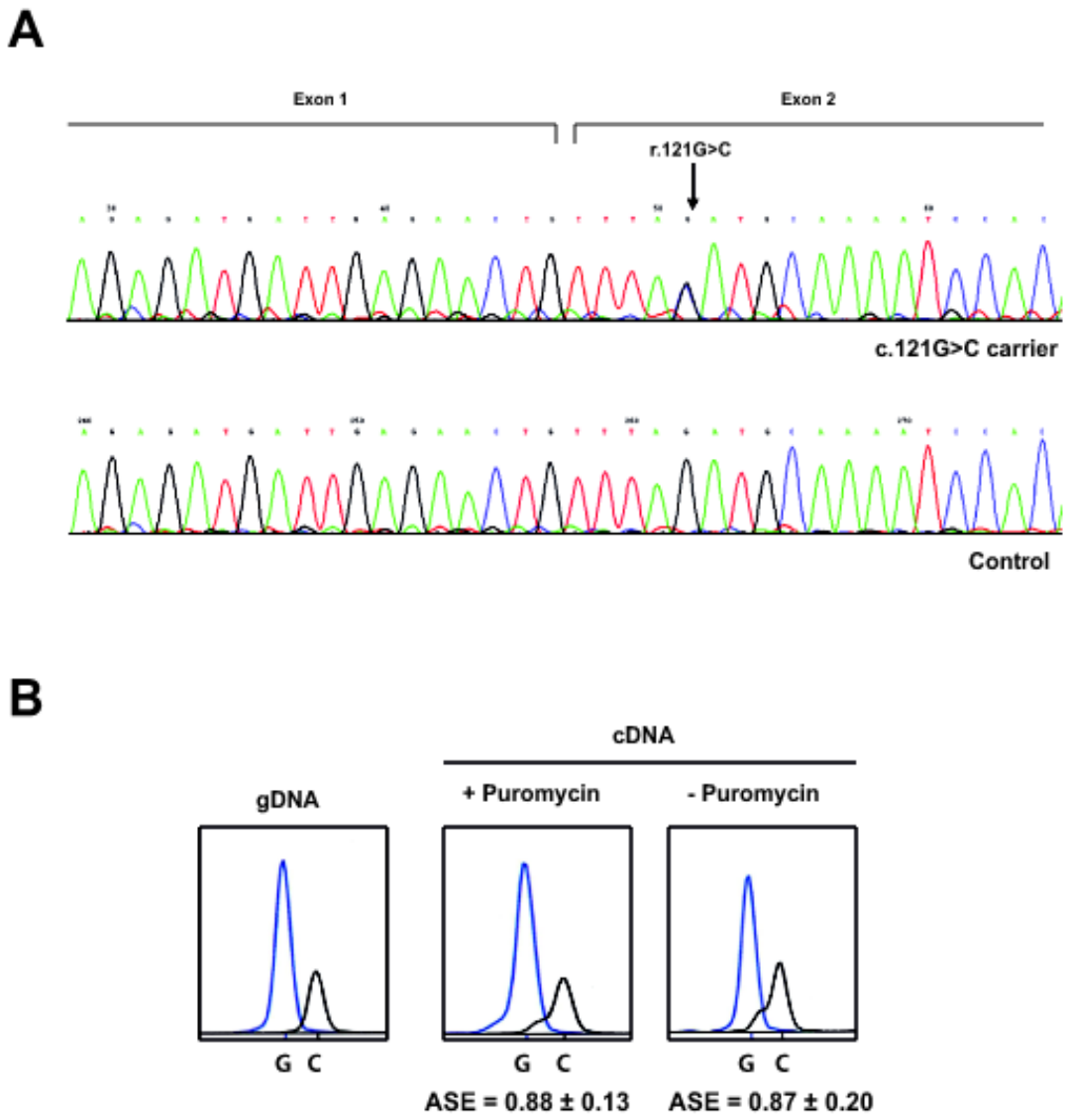
													*							
Hsap	30	N	A	I	K	E	M	I	E	N	C	L	D	A	K	S	T	S	I	Q	V
Mmul	30	N	A	I	K	E	M	I	E	N	C	L	D	A	K	S	T	S	I	Q	V
Mmus	30	N	A	I	K	E	M	I	E	N	C	L	D	A	K	S	T	N	I	Q	V
Cfam	30	N	A	I	K	E	M	I	E	N	C	L	D	A	K	S	T	S	I	Q	V
Mdom	30	N	A	I	K	E	M	I	E	N	C	L	D	A	R	A	S	A	I	Q	V
Ggal	30	N	A	I	K	E	M	I	E	N	C	L	D	A	K	S	T	S	I	Q	V
Xlae	27	N	A	I	K	E	M	I	E	N	C	L	D	A	K	S	T	S	I	Q	V
Drer	27	N	A	I	K	E	M	M	E	N	C	L	D	A	K	S	T	N	I	Q	I
Cint	31	N	A	V	K	E	M	I	E	N	C	L	D	A	G	S	T	T	I	T	V
Bflo	28	N	A	V	K	E	M	L	E	N	C	L	D	A	K	S	S	S	I	Q	V
Spur	28	N	A	L	K	E	M	I	E	N	C	L	D	A	K	S	T	S	I	Q	V
Tadh	27	N	A	I	K	E	M	I	E	N	S	L	D	A	K	A	T	L	I	Q	V

B

αE helix

													*								
Hsap	700	Q	Q	S	E	V	P	G	S	I	P	N	S	W	K	W	T	V	E	H	I
Mmul	700	Q	Q	S	E	V	P	G	S	I	P	N	S	W	K	W	T	V	E	H	I
Mmus	704	Q	Q	S	D	M	P	G	S	T	S	K	P	W	K	W	T	V	E	H	I
Cfam	701	Q	Q	S	E	V	C	G	S	S	A	N	P	W	K	W	T	V	E	H	I
Mdom	679	D	Q	K	E	E	C	E	S	S	P	V	S	W	K	W	T	V	E	H	I
Ggal	701	S	Q	N	E	D	S	D	S	G	P	P	P	W	K	W	T	V	E	H	V
Xlae	666	D	N	K	S	L	-	-	T	G	S	S	S	W	R	W	T	T	E	H	I
Drer	674	Q	D	-	-	-	-	-	-	A	E	M	S	W	Q	W	K	V	E	H	V
Cint	711	P	V	V	V	E	T	D	S	E	W	S	P	W	K	Q	M	V	E	H	V
Bflo	658	T	S	E	D	-	-	-	-	-	S	K	S	W	K	W	T	V	E	H	A
Spur	689	D	A	A	S	G	A	D	M	P	S	Y	N	W	K	W	T	I	E	F	V
Tadh	671	N	Q	V	F	Q	-	V	D	K	R	K	M	W	K	W	K	V	E	H	L

Supplementary Fig.S4. Analysis of *MLH1* c.121G>C at mRNA level. A) Direct sequencing of theRT-PCR products in lymphocytes cultured in the absence of puromycin, from an*MLH1* c.121G>C carrierand a control individual. Sequencing analysis showed the same pattern in the presence of puromycin (data not shown). **B)** Representative results of the SNuPE analysis at *MLH1* c.121G>C in gDNA and cDNA (from lymphocytes cultured in the absence or presence of puromycin) derived from a variant carrier. ASE (mean±SD) was calculated by dividing the proportion of variant/wild-type allele in cDNA by the proportion of variant/wild-type allele in gDNA. Experiments were performed in quadruplicate.



ARTÍCULO 5

Elucidating the molecular basis of MSH2-deficient tumors by combined germline and somatic analysis

Gardenia Vargas, **Maribel González-Acosta**, Bryony A. Thompson, Carolina Gómez, Anna Fernández, Estela Dámaso, Tirso Pons, Monika Morak, Jesús del Valle, Silvia Iglesias, Àngela Velasco, Ares Solanes, Xavier Sanjuan, Natàlia Padilla, Xavier de la Cruz, Alfonso Valencia, Elke Holinki-Feder, Joan Brunet, Lídia Feliubadaló, Conxi Lázaro, Matilde Navarro, Marta Pineda* and Gabriel Capellá*.

* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

International Journal of Cancer, 2017. DOI: 10.1002/ijc.30820

RESUMEN:

La hipótesis del presente trabajo es que el análisis exhaustivo, tanto a nivel germinal como somático, de las alteraciones genéticas en los genes MMR y otros genes de predisposición a cáncer colorrectal (CCR) puede ser útil para dilucidar las bases moleculares de los casos con sospecha de síndrome de Lynch o pacientes Lynch-like (SLL). Por lo tanto, se pretende estudiar la eficacia del algoritmo diagnóstico de síndrome de Lynch mediante el análisis exhaustivo de genes MMR y la implementación de un panel de secuenciación de nueva generación (NGS) para el análisis de mutaciones germinales y somáticas en genes asociados a CRC.

En el estudio se incluyeron 58 casos con tumores con pérdida de MSH2/MSH6. Se identificaron 27 variantes patogénicas y 8 probablemente patogénicas en *MSH2* y *EPCAM*. La secuenciación de las regiones promotoras identificó dos variantes en el promotor *MSH6*. El estudio del RNA identificó transcritos aberrantes en 4 de las 7 variantes evaluadas en *MSH2*. El estudio mediante un panel personalizado de NGS del DNA de sangre periférica de pacientes SLL identificó una variante patogénica previamente no identificada, y diversas variantes predichas como patogénicas en los genes *MUTYH*, *SETD2*, *BUB1* y *FAN1*. El estudio de DNA tumoral mediante el mismo panel de NGS detectó dobles mutaciones somáticas en los genes MMR de los tumores de 2 de los 5 tumores estudiados. En los casos restantes, se hallaron mutaciones heterocigotas complejas en los genes MMR (*MSH6, PMS2, MLH3*) y/o POLD1/POLE. Además, también se evidenciaron otras mutaciones somáticas en otros genes asociados a predisposición de cáncer (*APC, AXIN2, BMPR1A, PTEN* o *BUB1B*), coexistiendo con las alteraciones previamente mencionadas.

En conclusión, en pacientes SLL, la evaluación de patogenicidad de variantes de significado desconocido en genes MMR y su estudio mediante el panel NGS es útil para la

identificación de dobles mutaciones somáticas y mutaciones germinales candidatas en genes de predisposición a CCR. Esta estrategia, además, podría ayudar a dilucidar las bases moleculares del SLL.

IJC
International Journal of Cancer

Elucidating the molecular basis of MSH2-deficient tumors by combined germline and somatic analysis

Gardenia M. Vargas-Parra[1], Maribel González-Acosta[1], Bryony A. Thompson[2,3], Carolina Gómez[1], Anna Fernández[1], Estela Dámaso[1], Tirso Pons[4], Monika Morak[5,6], Jesús del Valle[1], Silvia Iglesias[1], Àngela Velasco[7], Ares Solanes[8], Xavier Sanjuan[9], Natàlia Padilla[10], Xavier de la Cruz[10,11], Alfonso Valencia[4], Elke Holinski-Feder[5,6], Joan Brunet[7], Lídia Feliubadaló[1], Conxi Lázaro[1], Matilde Navarro[1,8], Marta Pineda[1†] and Gabriel Capellá[1†]

[1] Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, CIBERONC, Hospitalet de Llobregat, Spain
[2] Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT
[3] Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Melbourne, Australia
[4] Structural Biology and Biocomputing Program, Spanish National Cancer Research Center (CNIO), Madrid, Spain
[5] Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, Ziemssenstr. Germany MGZ-Medizinisch Genetisches Zentrum, Munich, Germany
[6] MGZ-Medizinisch Genetisches Zentrum, Munich, Germany
[7] Hereditary Cancer Program, Catalan Institute of Oncology, IdIBGI, Girona, Spain
[8] Hereditary Cancer Program, Catalan Institute of Oncology, Hospital Germans Trias i Pujol, Badalona, Spain
[9] Pathology Department, Hospital Universitari de Bellvitge-IDIBELL, Hospitalet de Llobregat, Barcelona, Spain
[10] Research Unit in Translational Bioinformatics, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain
[11] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

In a proportion of patients presenting mismatch repair (MMR)-deficient tumors, no germline MMR mutations are identified, the so-called Lynch-like syndrome (LLS). Recently, MMR-deficient tumors have been associated with germline mutations in *POLE* and *MUTYH* or double somatic MMR events. Our aim was to elucidate the molecular basis of MSH2-deficient LS-suspected cases using a comprehensive analysis of colorectal cancer (CRC)-associated genes at germline and somatic level. Fifty-eight probands harboring MSH2-deficient tumors were included. Germline mutational analysis of *MSH2* (including *EPCAM* deletions) and *MSH6* was performed. Pathogenicity of *MSH2* variants was assessed by RNA analysis and multifactorial likelihood calculations. *MSH2* cDNA and methylation of *MSH2* and *MSH6* promoters were studied. Matched blood and tumor DNA were analyzed using a customized next generation sequencing panel. Thirty-five individuals were carriers of pathogenic or probably pathogenic variants in *MSH2* and *EPCAM*. Five patients harbored 4 different *MSH2* variants of unknown significance (VUS) and one had 2 novel *MSH6* promoter VUS. Pathogenicity assessment allowed the reclassification of the 4 *MSH2* VUS and 6 probably pathogenic variants as pathogenic mutations, enabling a total of 40 LS diagnostics. Predicted pathogenic germline variants in *BUB1*, *SETD2*, *FAN1* and *MUTYH* were identified in 5 cases. Three patients had double somatic hits in *MSH2* or *MSH6*, and another 2 had somatic alterations in other MMR genes and/or proofreading polymerases. In conclusion, our comprehensive strategy combining germline and somatic mutational status of CRC-associated genes by means of a subexome panel allows the elucidation of up to 86% of MSH2-deficient suspected LS tumors.

Key words: Lynch syndrome, Lynch-like, next-generation sequencing, mismatch repair-deficiency, methylation

Abbreviations: ASE: allele-specific expression; CRC: colorectal cancer; FFPE: formalin-fixed paraffin embedded; IHC: immunohistochemistry; LLS: Lynch-like syndrome; LOH: loss of heterozygosity; LS: Lynch syndrome; MMR: mismatch repair; MS-MCA: methylation-specific melting curve analysis; MSI: microsatellite instability; NGS: next-generation sequencing; PBL: peripheral blood leukocytes; VUS: variant of unknown significance

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

[†]M.P. and G.C. contributed equally to this work and shared senior authorship

Disclosures: The authors declare no conflict of interest.

Grant sponsor: Spanish Ministry of Economy and Competitiveness; **Grant numbers:** SAF2012-33636 and SAF2015-68016-R; **Grant sponsor:** FEDER, a way to build Europe; **Grant sponsor:** Spanish Association Against Cancer; **Grant sponsor:** Government of Catalonia; **Grant number:** 2014SGR338; **Grant sponsor:** Fundación Mutua Madrileña; **Grant number:** AP114252013; **Grant sponsor:** RTICC MINECO Network; **Grant numbers:** RD12/0036/0031 and RD12/0036/0008; **Grant sponsor:** Mexican National Council for Science and Technology (CONACyT); **Grant number:** 310756

DOI: 10.1002/ijc.30820

History: Received 7 Dec 2016; Accepted 16 May 2017; Online 2 June 2017

Correspondence to: Gabriel Capellá, MD, PhD and Marta Pineda, PhD, Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Av. Gran Via de l'Hospitalet, 199-203, 08908 Hospitalet de Llobregat, Spain, Tel.: +34-93-260-7319, Fax: +34-93-260-7466, E-mail: gcapella@iconcologia.net and mpineda@iconcologia.net

What's new?

Although Lynch syndrome is known as an inherited cancer syndrome causing colorectal and endometrial tumors at a young age, more than half of the affected individuals do not carry the expected germline mutations in mismatch repair genes. Here the authors comprehensively analyzed the germline and somatic mutational status of patients with suspected Lynch syndrome. They confirm marked heterogeneity in the underlying mutations and molecularly classified up to 86% of the cases, underscoring the need for a comprehensive analysis to allow meaningful genetic counseling and follow-up.

Lynch syndrome (LS) is an inherited autosomal dominant cancer syndrome that accounts for 2–4% of all newly diagnosed colorectal and endometrial cancers.[1–3] It is caused by defective mismatch repair (MMR) activity due to germline (epi)mutations in MMR genes (*MLH1, MSH2, MSH6* and *PMS2*). The diagnostic algorithm of LS is based on the identification of microsatellite instability (MSI) and/or loss of expression of MMR proteins by immunohistochemistry (IHC) in tumors. After identification of MMR deficiency (in the absence of *MLH1* promoter methylation and/or *BRAF* p.V600E mutation), germline MMR testing is performed. However, about 55% of patients with MMR-deficient colorectal and endometrial tumors lack identified pathogenic mutations by conventional analyses, thus hampering appropriate clinical management and risk assessment in these so-called Lynch-like syndrome (LLS) patients.[4] LLS patients together with their first-degree relatives are considered to have an intermediate risk of developing CRC.[5] Recently, somatic double hits in DNA repair genes have been detected in a variable proportion (30–82%) of LLS.[6–9] While somatic *MLH1* promoter hypermethylation is common in MSI tumors,[10,11] the relative contribution of somatic methylation in other MMR gene promoters in LLS has been poorly studied.[12,13]

Limitations in the molecular analysis techniques utilized could be responsible for the lack of detection of germline MMR mutations, due to false-positive IHC/MSI results, false-negative results in MMR mutational analysis due to complex or cryptic mutations[14–18] or lack of sensitivity (i.e. in mosaic cases).[9] Moreover, up to 30% MMR variants are classified as variants of unknown significance (VUS), in which their clinical significance is not evident.[19] Moreover, germline mutations in genes other than MMR genes (biallelic *MUTYH* and *POLE*) have been reported rarely in patients with MMR-deficient tumors,[7,20–23] reinforcing the need to implement NGS gene panels (either commercially available or custom-made) in the routine setting.[24,25]

In this work, we aimed at elucidating the molecular basis underlying tumorigenesis in a cohort of 58 LS-suspected patients harboring MSH2-deficient tumors using a comprehensive strategy. Sequencing of a panel of CRC-associated genes in germline and tumor formalin-fixed paraffin-embedded (FFPE) samples was used to complement the germline MMR gene (epi)mutation testing.

Material and Methods

Patients

Mutational screening of *MSH2* was performed in a cohort of 58 probands with LS-associated tumors showing loss of MSH2 protein expression by IHC (Supporting Information, Table S1). Patients were assessed at Cancer Genetic Counseling Units at the Catalan Institute of Oncology from 1998 to 2012. Twenty patients fulfilled Amsterdam criteria, 36 revised Bethesda criteria and the remaining 2 were referred to the Genetic Counseling Unit for showing histological features suggestive of MMR-deficiency and loss of MSH2 expression. Clinical and pathological information of affected individuals was recorded. DNA samples from controls of a hospital based CRC case–control study were used to analyze the frequency of the detected *MSH2* VUS.[26] Informed consent was obtained from all individuals enrolled, and internal Ethics Committees of participant hospitals approved this study. Of note, three patients initially classified as LLS were excluded from this cohort due to the detection of biallelic *MUTYH* mutations as reported.[20,27]

Isolation of genomic DNA

Peripheral blood leukocyte (PBL) DNA was extracted using FlexiGene DNA kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For each available specimen of formalin-fixed paraffin-embedded (FFPE) tissue, 10–20 × 10-μm FFPE sections were cut from a single representative block per case, using macrodissection with a scalpel as needed to enrich for tumor cells. After deparaffinization with 480 μl of Deparaffinization Solution (Qiagen, Hilden, Germany), DNA isolation was performed using either the DNAeasy Tissue Kit or QIAmp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer's instructions.

Mismatch repair genes mutational analysis

Mutational analysis of coding regions of MSH2 and MSH6 genes. Point mutation analysis of *MSH2* (NM_000251.2, NG_007110.1) and *MSH6* (NM_000179.2; NG_007111.1) was performed by PCR amplification of exonic regions and exon–intron boundaries followed by Sanger sequencing (primers and conditions available upon request). Genomic rearrangements in MMR genes were analyzed by multiplex ligation dependent probe amplification using SALSA-MLH1/MSH2 P003-B1 and MSH6 P072 kits (MRC-Holland), which include

probes at the 3′ end of *EPCAM*. Annotation of variants was done following the HGVS recommendations.

Direct sequencing of MSH2 and MSH6 promoter regions and 3′UTR of EPCAM gene. The regions encompassing 662 bases upstream of the transcriptional start site (TSS) of *MSH2*, 915 bp upstream of the *MSH6* TSS and 429 bp of the *EPCAM* 3′UTR were amplified by PCR using Megamix-Double (Microzone Ltd., UK) and sequenced using the Big-Dye Terminator v.3.1 Sequencing Kit (Applied Biosystems, CA, USA) (Supporting Information, Table S2; conditions available upon request). Sequences were analyzed on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems).

Mutational analysis of MSH2 whole transcript. Human blood lymphocytes were incubated with and without puromycin after one week of culture with Gibco® *PB-MAX*™ medium. Subsequently total RNA was extracted from cultured lymphocytes with *TRIzol*® Reagent. One microgram of RNA was retrotranscribed using *iScript Select cDNA synthesis kit (Bio-Rad, USA)*. The whole *MSH*2 transcript (*2.8Kb)* was amplified by Long Range-PCR (primers and conditions kindly provided by E. Holinski-Feder and M. Morak). Products were run in an electrophoresis gel and purified with Exonuclease 1 plus Shrimp Alkaline Phosphatase (ExoSAP). Finally, 5 primer-pairs were used to analyze the whole coding region by Sanger sequencing.

Pathogenicity assessment of *MSH2* variants

Variant frequency and co-segregation analysis. Global population frequency of the identified *MSH2* variants was retrieved from the Exome Aggregation Consortium (ExAC; http://exac.broadinstitute.org/) and NHLBI Exome Sequencing Project (ESP; http://evs.gs.washington.edu/EVS) databases and Spanish population frequency was screened by Sanger sequencing in a cohort of 246 healthy controls.[26] *MSH2* variants were also screened in DNA samples from family relatives by Sanger sequencing.

In silico prediction of the functional impact. DNA sequences containing the identified *MSH2* variants were analyzed using several bioinformatic tools to evaluate their impact at the RNA and protein level, as previously reported.[28,29] PROMO computational tool was used to analyze the predicted impact of promoter variants. Protein stability predictions were obtained by applying PoPMuSic (http://dezyme.com/), CUPSAT (http://cupsat.tu-bs.de), ERIS (http://troll.med.unc.edu/eris/), I-Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) and FoldX 4 (http://foldxsuite.crg.eu). For the structural analysis, the structure of the DNA lesion recognition complex (PDB code: 2O8B) was used, which includes human MSH2, MSH6 and a DNA substrate. PyMOL Molecular Graphics System v1.5.0.4 (Schrödinger, LLC) was used to visualize structures. The disease-related variants in the close vicinity of the MSH2 variants identified in this study were calculated using Structure-PPi.[30] The 3 D clustering of missense variants is often used as a supporting evidence for the involvement of those variants in the disease or as a basis for functional hypotheses about the clustered mutations.

Multifactorial likelihood analysis. Multifactorial likelihood analysis was based on estimated prior probabilities of pathogenicity and likelihood ratios for segregation and tumor characteristics as described.[19] For variants without available prior probabilities (exonic variants altering splicing), a prior probability of 0.5 was used (no prior assumptions). Variants were classified according to the 5 class IARC scheme,[31] based on the calculated posterior probability.

mRNA splicing analysis and allele-specific expression analysis. Human lymphocytes from variant carriers were cultured and total RNA was extracted as described above. Subsequently cDNA was synthesized as described.[28] Amplification of *MSH2* coding region containing the variants was performed using specific primers (Supporting Information, Table S2 conditions available upon request). Sequences of carriers' transcripts were compared with transcripts from three control lymphocyte cultures. Allele-specific expression (ASE) was analyzed by SNuPE[28] (Supporting Information, Table S2; conditions available upon request). ASE was calculated by dividing the ratio of variant/wildtype allele in cDNA by the ratio of variant/wildtype allele in gDNA. We used $\leq$0.5 as a threshold value for ASE definition. Experiments were performed in quadruplicate.

Targeted next generation sequencing

Agilent SureDesign web-based application (Agilent Technologies, USA) was used to design DNA capture probes of 509 target regions, including the coding exons plus 10 flanking bases of 26 genes associated to CRC, and their promoter regions (comprising 650 bases upstream their TSS) (Supporting Information, Table S3). Regions containing somatic hotspot mutations in 12 actionable target genes and MSI CRC-associated loci of *SETD2*, *SETD1B* and *SETDB2* were also included[32] (Supporting Information, Table S3). Design was optimized for FFPE samples. Final design was composed of 11,012 amplicons covering 99.61% of the submitted target regions, in a total sequenceable design size of 319,653 kb.

DNA quality was tested using NanoDrop ND 1000 Spectrophotometer (Thermo Fischer Scientific), by electrophoresis in agarose gel and by Qubit Fluorometer using dsDNA BR Assay (Invitrogen, Carlsbad, CA, USA). To assess FFPE-derived DNA integrity, a PCR amplifying two *GAPDH* products was performed and the products were visualized using High Sensitivity DNA chips in a Bioanalyzer (Agilent Technologies). Capture of the target regions was performed using HaloPlex Target Enrichment kit 1–500 kb (Agilent Technologies), according to the HaloPlex Target Enrichment System-Fast Protocol Version B. Briefly, the protocol consists of four steps: (1) digestion of genomic DNA using eight different restriction

reactions; (2) hybridization of restricted fragments to probes whose ends are complementary to the target fragments, circularization of fragments and incorporation of sequencing motifs including index sequences; (3) capture of target DNA using streptavidin beads and ligation of circularized fragments; (4) PCR amplification of captured target libraries. Quality control and dilution estimates of libraries were performed using High Sensitivity DNA chips in a Bioanalyzer. Library concentrations were normalized to 0.44 nM. Pooled libraries were sequenced in a MiSeq (Illumina) with paired-end 250 bp reads plus an 8-base index read, using MiSeq Reagent Kit v3.

Agilent SureCall application was used to trim, align and call variants. Variant filtering was performed based on Phred quality $\geq$30, alternative allele ratio $\geq$0.05, read depth $\geq$38$\times$ in PBL samples and $\geq$10$\times$ in FFPE samples. Identified variants were then filtered against common single-nucleotide polymorphisms (MAF>1 according to ExAC and ESP databases). Predicted pathogenic germline rare variants and *MSH2* double somatic hits were further confirmed by Sanger sequencing using independent DNA samples.

Loss of heterozygosity analysis

Loss of heterozygosity (LOH) was assessed in FFPE tumor DNA by analyzing the alternative allele ratio of germline heterozygous *MSH2* SNPs genotyped by NGS and three microsatellite markers (*D2S2328*, *D2S288* and *D2S378*) spreading over 17 Mb around *MSH2* (Supporting Information, Table S4).

Methylation analysis of *MSH2* and *MSH6* genes

Methylation was evaluated by Methylation Specific-Melting Curve Analysis (MS-MCA), consisting of a real-time PCR followed by temperature dissociation of bisulfite-treated DNA, using the EZ DNA Methylation-Gold Kit (Zymo Research, USA). Each promoter region was preamplified using 2 μl of external primers at 2 μM, 1 μl of bisulfite-treated DNA and 5 μl of Double MegaMix solution (Microzone Ltd., UK). Hemi-nested PCRs of both promoter regions were carried out in a LightCycler 480 II (Roche, Germany) using 1 μl of a 1:10 dilution of preamplified fragments in 9 μl of Light Cycler 480 SYBR Green I (Roche) containing 0.5 μM of each internal primer. Primer sequences are listed in Supporting Information, Table S2. The amplified region of *MSH2* and *MSH6* promoters covered 13 and 18 CpGs, respectively. *In vitro* methylated DNA from CpG methylated Jurkatt Genomic DNA (New England Biolabs, MA, USA) and a CRC sample from an *EPCAM* deletion carrier were used as methylated controls in these experiments. Analytical sensitivity of the method to detect methylation was assessed using serial dilutions of methylated Jurkatt DNA and lymphocyte DNA from a healthy patient (after bisulfite sequencing corroboration of unmethylation). Analytical sensitivities of 10 and 25% were achieved in the analysis of *MSH2* and *MSH6* promoters, respectively (Supporting Information, Fig. S1).

Results

Clinical characteristics of patients with MSH2-deficient tumors

We identified 58 probands diagnosed with LS-associated tumors showing loss of MSH2 expression (Supporting Information, Table S1A). Accordingly, MSI was evident in all the informative tumors available ($n = 28$). DNA mutational analysis allowed the identification of 25 patients harboring *bona fide* germline pathogenic variants (IARC class 5) in *MSH2* and 2 in *EPCAM* and 8 harboring *MSH2* likely pathogenic variants (class 4) according to the InSiGHT classification rules[19] (Table 1). In all, 35 of 58 patients were classified as LS. The remaining 23 were categorized as Lynch-like syndrome (LLS), 5 of them being carriers of *MSH2* variants of unknown significance (VUS; class 3).

In the identified LS patients, the mean age at first LS-associated-tumor diagnosis was of 45.8 years, while it was of 49.2 years in LLS cases (Supporting Information, Table S1B). Concerning clinical criteria fulfillment, 49% of LS cases met Bethesda criteria being this proportion higher (82.6%) in LLS. Fifty-seven percent of LS cases and 25% of LLS patients ($n = 20$ and $n = 6$, respectively) presented multiple LS-associated tumors.

Pathogenicity assessment of MSH2 variants

Four *MSH2* VUS variants (c.518T > G, c.2069A > G, exon 11 duplication and exons 11–16 duplication) and 6 probably pathogenic variants (c.211G > C, c.989T > C, c.1276G > A, c.1511–1 G > A, c.2074G > C and c.[2635–3C > T;2635–5T > C]) were identified in 13 probands (Tables 1 and 2). None of them was described in ExAC and NHLBI ESP Databases nor identified in Spanish cohorts of control individuals (Table 2). *In silico* predictions are shown in Supporting Information, Table S5.

cDNA splicing evaluation was performed in carriers of 7 variants (the 4 VUS, c.211G > C, c.989T > C and c.1276 G > A), with available lymphocytes. In four of them, aberrant transcripts were identified (Supporting Information, Table S5 and Fig. S2): (i) *MSH2* c.211G > C (Case 234) results in a partial deletion of exon 1 (r.195_211del), which is predicted to generate a truncated protein (p.Tyr66Serfs*10); (ii) *MSH2* c.1276G > A (Case 258) leads to a partial deletion of exon 7 by activation of a cryptic donor site (r.1230_1277del), which is predicted to generate an in-frame deletion of 16 amino acids (p.Ile411_Gly426del) in a highly conserved MSH2 domain (Supporting Information, Fig. S3); (iii) the duplication of exon 11 (Case 264) causes its duplication in tandem (r.1662_1759dup; p.Gly587Alafs*3); and (iv) the duplication of exons 11–16 identified (Case 120) generates a longer transcript (r.1662_*23dup) containing a tandem duplication of exons 11–16 and 23 nucleotides of the 3′UTR downstream the stop codon. Therefore, three *MSH2* variants (c.211G > C, c.1276G > A and exon 11 duplication) were reclassified as pathogenic based on the generation of aberrant transcripts leading to premature stop codons or in frame-deletions disrupting functional domains[19] (Table 1). Although the

Table 1. Summary of the results obtained in the molecular information of the MSH2-deficient included cases [Color table can be viewed at wileyonlinelibrary.com]

Patient ID	Clinical criteria	MMR mutational analysis (Sanger seq and MLPA)					VUS assessment		RNA analysis	Promoter analysis (Sanger)		3'UTR (Sanger)	MMR mutational Variants identified in		Tumor analysis			Final classification^
		MSH2/ EPCAM	*MSH6*	Identified variants , cDNA change^	Identified variant (predicted protein change)	Initial MMR variant classification (September 2015)	Splicing and multifactorial analysis	MMR variant reclassification (this study)	whole *MSH2* transcript	*MSH2*	*MSH6*	*EPCAM*	MSH2/MSH6 ^	CRC associated genes	Methylation *MSH2*	Methylation *MSH6*	*MSH2* and *MSH6* hits / LOH	
242	AC	VI	NP	*MSH2* c.(?_-68)_(366+1_367-1)del (del E1-2)	p.?	Class 5												LS
233	BC	VI	NP	*MSH2* c.528_529delTG	p.(Cys176*)	Class 5												LS
249	AC	VI	NP	*MSH2* c.536dup	p.(Asp180*)	NR (Class 5 acc.the rules)												LS
253	BC	VI	NP	*MSH2* c.536dup	p.(Asp180*)	NR (Class 5 acc. the rules)												LS
250	AC	VI	NP	*MSH2* c.602dup	p.(Leu201Phefs*31)	NR (Class 5 acc. the rules)												LS
230	AC	VI	NP	*MSH2* c.689_691delinsTT	p.(Ala230Valfs*16)	NR (Class 5 acc. the rules)												LS
252	BC	VI	NP	*MSH2* c.689_691delinsTT	p.(Ala230Valfs*16)	NR (Class 5 acc. the rules)												LS
236	AC	VI	NP	*MSH2* c.735_736insTGTT	p.(Lys246Cysfs*2)	NR (Class 5 acc. the rules)												LS
231	BC	VI	NP	*MSH2* c.897T>A	p.(Tyr299*)	NR (Class 5 acc. the rules)												LS
245	AC	VI	NP	*MSH2* c.942+3A>T	p.(Val265_Gln314del)	Class 5												LS
260	BC	VI	NP	*MSH2* c.942+3A>T	p.(Val265_Gln314del)	Class 5												LS
262	BC	VI	NP	*MSH2* c.942+3A>T	p.(Val265_Gln314del)	Class 5												LS
255	BC	VI	NP	*MSH2* c.970C>T	p.(Gln324*)	Class 5												LS
257	BC	VI	NP	*MSH2* c.1035G>A	p.(Trp345*)	Class 5												LS
229	BC	VI	NP	*MSH2* c.1076+1G>A	p.?	Class 5												LS
256	AC	VI	NP	*MSH2* c.(1076+1_1077-1)_(*272_?)del (del E7-12)	p.?	Class 5												LS
263	BC	VI	NP	*MSH2* c.1165C>T	p.(Arg389*)	Class 5												LS
238	AC	VI	NP	*MSH2* c.1216C>T	p.(Arg406*)	Class 5												LS
261	AC	VI	NP	*MSH2* c.(1276+1_1277-1)_(1386+1_1387-1)del (del E8)	p.Lys427Glyfs*4	Class 5												LS
247	AC	VI	NP	*MSH2* c.1345_1348delAAGT	p.(Lys449Phefs*4)	Class 5												LS
237	AC	VI	NP	*MSH2* c.(1386+1_1387-1)_(1661+1_1662-1)del (del E9-10)	p.Val463Glnfs*7	Class 5												LS
259	AC	VI	NP	*MSH2* c.(1386+1_1387-1)_(1661+1_1662-1)del (del E9-10)	p.Val463Glnfs*7	Class 5												LS
241	AC	VI	NP	*MSH2* c.1705_1706delGA	p.(Glu569Ilefs*2)	Class 5												LS
254	BC	VI	NP	*MSH2* c.1777C>T	p.(Gln593*)	Class 5												LS
243	AC	VI	NP	*MSH2* c.2593dup	p.(Ile865Asnfs*17)	NR (Class 5 acc. the rules)												LS
246	AC	VI	NM	*EPCAM* c.858+2568_*4596del (del E8-9)	p.?	Class 5												LS
251	AC	VI	NM	*EPCAM* c.858+2568_*4596del (del E8-9)	p.?	Class 5												LS
234	BC	VI	NP	***MSH2* c.211G>C^**	p.(Gly71Arg)	**NR (Class 4 acc. the rules)^**	r.195_211del / 0.66	**Class 5^**										**LS^**
228	AC	VI	NM	***MSH2* c.989T>C^**	p.(Leu330Pro)	**Class 4^**	r.989T>C / >0.99	**Class 5^**					*MSH2* c.989T>C	NI	NP	NP	1 MSH2 / NA	**LS^**
258	BC	VI	NM	***MSH2* c.1276G>A^**	p.(Gly426Arg)	**NR (Class 4 acc. the rules)^**	r.1230_1277del48 / 0.66	**Class 5^**										**LS^**
248	BC	VI	NM	*MSH2* c.1511-1G>A	p.?	**NR (Class 4 acc. the rules)^**	NP / >0.99	**Class 5^**										**LS^**
239	BC	VI	NP	***MSH2* c.2074G>C^**	p.(Gly692Arg)	**Class 4^**	NP / >0.99	**Class 5^**										**LS^**
232	BC	VI	NP	***MSH2* c.[2635-3C>T; 2635-5T>C]^**	p.?	**Class 4^**	NP / >0.99	**Class 5^**										**LS^**
235	AC	VI	NP	***MSH2* c.[2635-3C>T; 2635-5T>C]^**	p.?	**Class 4^**	NP / >0.99	**Class 5^**										**LS^**
240	BC	VI	NP	***MSH2* c.[2635-3C>T; 2635-5T>C]^**	p.?	**Class 4^**	NP / >0.99	**Class 5^**										**LS^**
122	BC	VI	NP	***MSH2* c.518T>G^**	p.(Leu173Arg)	**Class 3^**	r.518T>G / >0.99	**Class 5^**	NA	NI	NI	NI	NP	NP	UM	UM	NP	**LS^**
117	BC	VI	NP	***MSH2* c.518T>G^**	p.(Leu173Arg)	**Class 3^**	r.518T>G / >0.99	**Class 5^**	NA	NI	NI	NI	*MSH2* c.518T>G	NI	NP	NP	NP	**LS^**
264	BC	VI	NP	***MSH2* c.(1661+1_1662-1)_ (1759+1_1760-1)dup (duplication E11)^**	p.?	**Class 3^**	r.1662_1759dup / 0.66	**Class 5^**	NP	NI	NI	NI	NP	NP	NP	NP	NP	**LS^**
120	AC	VI	NP	***MSH2* c.(1661+1_1662-1)_(*272_?)dup (duplication E11-16)**	p.?	**Class 3^**	r.1662_*23dup / >0.99	**Class 5^**	NP	NI	NI	NI	NP	NP	UM	UM	NP	**LS^**
118	BC	VI	NM	***MSH2* c.2069A>G^**	p.(Gln690Arg)	**NR (Class 3 acc. the rules)^**	r.2069A>G / >0.99	**Class 5^**	NI	NI	NI	NI	*MSH2* c.2069A>G	NI	UM	UM	NP	**LS^**
121	PC	VI	NM	*MSH2* c.965G>A	p.(Gly322Asp)	Class 1			NP	NI	NI	NI	*MSH2* c.965G>A	NI	NA	UM	NI / No LOH	LLS
119	BC	NM	VI	*MSH6* c.431G>T	p.(Ser144Ile)	Class 1			NI	NI	NI	NI	*MSH6* c.431G>T	NI	UM	UM	NP	LLS
102	BC	NM	NM	NI					NP	NI	c.[-25C>T(;)-204C>G]	NI	*MSH6* c.[-25C>T(;)-204C>G]	NI	NP	NP	NP	LLS (MMR VUS carrier)
109#	BC	NM	NM	NI#					r.[-16_211del, 195_211del]	NI	NI	NI	***MSH2* c.211G>C^**	NI	NP	NP	NP	**LS^**
101	BC	NM	NM	NI					NA	NI	NI	NI	NI	NI	NP	NP	NP	LLS
103	PC	NM	NP	NI					NI	NI	NI	NI	NI	NI	NP	NP	NP	LLS
104	BC	NM	NM	NI					NA	NI	NI	NI	NI	NI	NP	NP	NP	LLS
105	BC	NM	NM	NI					NI	NI	NI	NI	NI	*MUTYH, FAN1, SETD2*	NP	NP	NP	LLS
123	BC	NM	NP	NI					NI	NI	NI	NI	NI	NI	NP	NP	NP	LLS
107	BC	NM	NM	NI					NI	NI	NI	NI	NI	*SETD2*	NA	UM	NP	LLS
108	BC	NM	NP	NI					NP	NI	NI	NI	NI	NI	NA	UM	NI / LOH	LLS
110	BC	NM	NM	NI					NA	NI	NI	NI	NI	*SETD2*	UM	UM	NP	LLS
111	BC	NM	NM	NI					NI	NI	NI	NI	NI	*MUTYH*	UM	UM	2 MSH2 / NA	LLS
112	BC	NM	NM	NI					NI	NI	NI	NI	NI	NI	NP	NP	NP	LLS
113	BC	NM	NM	NI					NI	NI	NI	NI	NI	NI	NP	NP	NP	LLS
114	BC	NM	NM	NI					NA	NI	NI	NI	NI	NI	NP	NP	2 MSH6 / No LOH	LLS
115	AC	NM	NM	NI					NA	NI	NI	NI	NI	*BUB1*	UM	UM	NP	LLS
116	BC	NM	NM	NI					NA	NI	NI	NI	NI	NI	UM	UM	NP	LLS

Abbreviations: MLPA, multiplex ligation-dependent probe amplification; VI, variant identified; NMI, no mutation identified; NA, no amplification; NP, not performed; NP, not reported; NI, not identified; UM, unmethylated; LS, Lynch syndrome; LLS, Lynch-like syndrome; acc., according to. Bold letter and ^ indicate cases that have been reclassified in this study. # indicates the case harboring the germline MSH2 c.211G > C variant further identified by NGS analysis.

Table 2. Results of cDNA splicing and multifactorial likelihood analyses of the *MSH2* class 3 and 4 variants identified in our series

						Multifactorial likelihood analysis												
MSH2 variant	Frequency in controls (our cohort*/ESP/ExAC)	Initial classification (InSiGHTt v.1.9; September 2015)	RNA; predicted protein	Prior probability of pathogenicity	Prior used	Family ID	Individual ID	Ascertainment	Cancer (age)	MSI/IHC status	MSI CRC LR	MSI CRC LR total	Bayes	Segregation LR	Odds for causality	Posterior odds	Posterior probability of pathogenicity	Final classification (evidence)
c.211G>C	(0/188)/NR/NR	Probably pathogenic acc. the rules (class 4, NR)	r.195_211del; p.Tyr66Serfs*10	NA (splicing aberration)	0.5	**234**	III:2	Clinic	CRC (45)	MSH2/ MSH6 loss			1.9201	1.9201	1.9201	1.9201	0.6575	**Pathogenic, class 5** (splicing mutation)
c.518T>G	(0/190)/NR/NR	Unknown significance (class 3)	r.518T>G; p.Leu173Arg	0.953499658	0.9	**122**	II:2	Clinic	CRC (59)	MSH2/ MSH6 loss			0.9887	1.5823	13.7028	123.3248	0.9920	**Pathogenic, class 5** (multifactorial analysis)
						117	II:2	Clinic	CRC (44)	MSH2/MSH6 loss			0.9359					
						A1	III:1	Clinic	CRC (38)	MSH2/MSH6 loss			0.9					
						A2	II:2	Clinic	CRC (48)	MSI-H&MSH2/ MSH6 loss	8.66	8.66	1.9					
c.989T>C	(0/236)/NR/NR	Probably pathogenic (class 4)	r.989T>C; p.Leu330Pro	0.961065305	0.9	**CTE-L0015**	Liliana Varesco; LOVD entry	Population	CRC	MSI-H	6.96	6.96	1.1815	3.7114	25.8317	232,4850	0.9957	**Pathogenic, class 5** (multifactorial analysis)
						228	II:3	Clinic	OC (55)	MSI-H&MSH2NV/ MSH6 loss			3.1413					
c.1276G>A	(0/246)/NR/NR	Probably pathogenic acc. the rules (class 4, NR)	r.1230_1277del; p.Ile411_Gly426del	NA (splicing aberration)	0.5	**258**	II:4	Clinic	OC (42)	MSH2/ MSH6 loss			1.9846	1.9846	1.9846	1.9846	0.6649	**Pathogenic, class 5** (splicing mutation)
							III:8	Clinic	EC (28)	MSI-H&MSH2/ MSH6 loss								
c.1511−1G>A	NP/NR/NR	Probably pathogenic acc. the rules (class 4, NR)	r.spl?; p.?	NA	0.96	**248**	II:1	Clinic	CRC (56)	MSI-H&MSH2/ MSH6 loss	8.66	8.66	1.1734	1.1734	10.1616	243.8795	0.9959	**Pathogenic, class 5** (multifactorial analysis)
MSH2 c.(1661 + 1_1662−1)_ (1759 + 1_1760−1)dup (duplication E11)	NP/NA/NR	Unknown significance (class 3)	r.1662_1759dup; p.Gly587Alafs*3	NA (splicing aberration)	0.5	**264**	II:1	Clinic	CRC1 (29)	MSH2/ MSH6 loss			1.9139	1.9139	1.9139	1.9139	0.6568	**Pathogenic, class 5** (splicing mutation)
								Clinic	CRC2 (51)	MSH2/ MSH6 loss								
MSH2 c.(1661 + 1_1662−1)_ (*272_?)dup (duplication E11−16)	NP/NA/NR	Unknown significance (class 3)	r.1662_*23dup; p.?	NA (splicing aberration)	0.5	**120**	II:3	Clinic	CRC (54)	MSH2/ MSH6 loss	8.66	74.99	22.2877	22.2877	1671.4794	1671.4794	0.9994	**Pathogenic, class 5** (multifactorial analysis)
							II:5	Clinic	CRC (39)	MSI-H								
							II:9	Clinic	CRC (52)	MSI-H&MSH2/ MSH6 loss	8.66							

Table 2. Results of cDNA splicing and multifactorial likelihood analyses of the *MSH2* class 3 and 4 variants identified in our series (Continued)

MSH2 variant	Frequency in controls (our cohort*/ESP/ExAC)	Initial classification (InSiGHTt v.1.9; September 2015)	RNA; predicted protein	Multifactorial likelihood analysis															Final classification (evidence)
				Prior probability of pathogenicity	Prior used	Family ID	Individual ID	Ascertainment	Cancer (age)	MSI/IHC status	MSI CRC LR	MSI CRC LR total	Bayes	Segregation LR	Odds for causality	Posterior odds	Posterior probability of pathogenicity	
c.2069A>G	(0/190)/NR/NR	Unknown significance (class 3, NR)	r.2069A>G; p.Gln690Arg	0.954182992	0.9	**118**	I:2	Clinic	CRC1 (35)	MSI-H&MSH2/ MSH6 loss	8.66	8.66	1.5290	1.5290	13.2411	119.1703	0,9917	**Pathogenic, class 5** (multifactorial analysis)
								Clinic	CRC2 (52)	MSH2/ MSH6 loss								
c.2074G>C	NP/NR/NR	Probably pathogenic (class 4)	r.?; p.(Gly692Arg)	0.961843012	0.9	**B**	Isidro et al., 2000					8.66	3.5469	3.8775	33.5789	302.2101	0.9967	**Pathogenic, class 5** (multifactorial analysis)
						239	II:2	Clinic	CRC (36)	MSI-H&MSH2/ MSH6 loss	8.66		1.0932					
c.[2635−3C>T; 2635−5T>C]	NP/NR/NR	Probably pathogenic (class 4)	r.spl?; p.?	NA (intronic)	0.26	**232**	III:1	Clinic	CRC (56)	MSH2/MSH6 loss		649.4619	1.6360	18.4104	11956.8715	4201.0629	0.9998	**Pathogenic, class 5** (multifactorial analysis)
						235	III:1	Clinic	CRC1 (28)	MSI-H&MSH2/ MSH6 loss	8.66		19.6874					
							III:2	Clinic	CRC2 (21)	MSI-H&MSH2/ MSH6 loss	8.66							
							III:5	Clinic	EC (39)	MSI-H&MSH2/ MSH6 loss								
						240	III:2	Clinic	CRC (53)	MSI-H&MSH2/ MSH6 loss	8.66		0.5716					

Abbreviations: LR, likelihood ratio; NA, not available; NR, not reported, NP, not performed; CRC, colorectal cancer; EC, endometrial cancer; OC, ovarian cancer; RC, renal cancer; acc., according to.

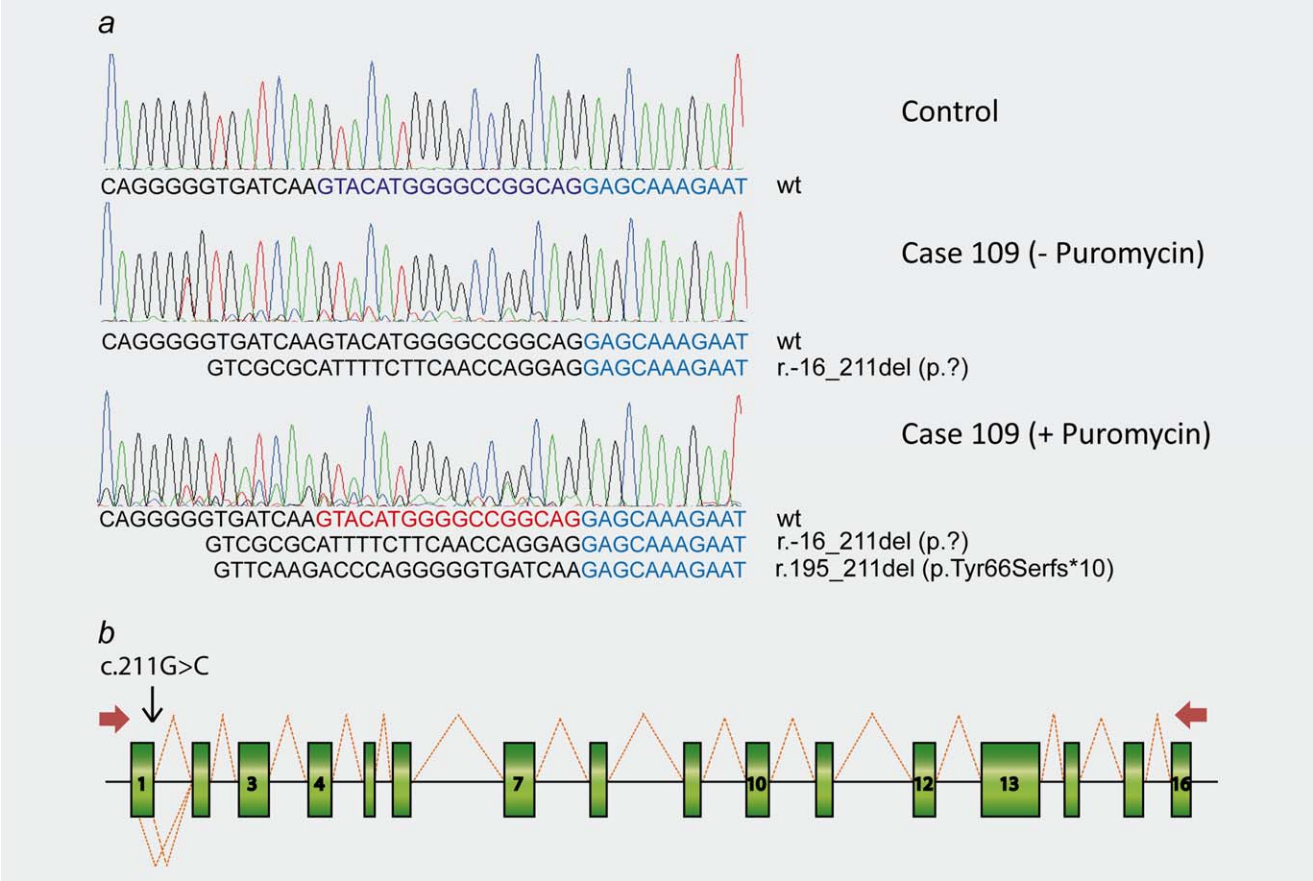


Figure 1. Results obtained in the analysis of the whole *MSH2* transcript of Case 109. (*a*) Result of the direct sequencing analysis of the RT-PCR product at the exon–intron boundary of exon 1 from a control sample and Case 109, harboring the germline *MSH2* c.211G > C variant, in presence and absence of puromycin. (*b*) A schematic representation of the normal transcript (upper dotted lines) and aberrant transcripts caused by the *MSH2* c.211G > C variant (lower dotted lines) is shown.

duplication of exons 11–16 leads to the generation of an aberrant transcript, as the duplicated region is inserted after the stop codon, its pathogenic effect at the protein level cannot be unequivocally demonstrated. The remaining *MSH2* variants (c.518T > G, c.989T > C and c.2069A > G) analyzed had no apparent effect on mRNA splicing and stability (Supporting Information, Table S5).

Clinicopathological data from all the families carrying class 3 and 4 variants were used in multifactorial calculations (Supporting Information, Fig. S4 and Table 2). As variant *MSH2* c.518 T > G was further identified in two additional families from other centers (Supporting Information, Fig. S4C), their data were also included in this analysis. Posterior probability of pathogenicity resulted >0.999 for 7 variants: c.518T > G, c.989T > C, c.1511–1G > A, c.2069A > G, c.2074G > C and c.[2635–3 C > T;2635–5T > C] and duplication of exons 11–16 (Tables 1 and 2). Therefore, multifactorial analyses allowed the classification as pathogenic mutations of the 7 variants not previously classified as pathogenic by cDNA analysis. Accordingly, the 4 missense variants (c.518T > G, c.989T > C, c.2069A > G, c.2074G > C) were *in silico* predicted as functionally damaging and destabilizing at the protein level (Supporting Information, Table S5), being involved in a network of interactions with other disease-associated variants (Supporting Information, Fig. S3).

In-depth germline analysis of LLS cases

To rule out having missed RNA-affecting mutations, we resequenced the whole *MSH2* transcript in 10 PBL samples. A splicing alteration was detected in one patient (Case 109). In absence of puromycin, a deletion of almost all the first exon was identified (r.−16_211del; p.?) (Fig. 1 and Table 1). Moreover, in presence of puromycin, an in-frame deletion of 16 bases (r.195_211del; p.Tyr66Serfs*10) was also detected. Further NGS analysis (see below) revealed a mutation in the last nucleotide of the first *MSH2* exon (c.211G > C), previously missed by Sanger sequencing due to primer design.

In the analysis of *MSH2* promoter region and the 3′UTR of *EPCAM* only known polymorphisms were detected (Table 1). Unfortunately, the low prevalence of heterozygous SNPs in these regions prevented the analysis of the presence of germline allelic imbalance (data not shown). Interestingly, 2 variants (c.−25C > T and c.−204C > G) were detected in *MSH6* promoter in Case 102. The *MSH6* variant c.−25 C > T is predicted to produce a premature out-of-frame start codon. *In silico*, variants c.−25C > T and c.−204C > G are

predicted to affect FOXP3 binding, and binding of TFII-I, STAT4, NFkappaB1, c-Ets-1, RelA and Elk-1, respectively.

Next, 17 PBL samples from LLS patients and 4 samples from reclassified *MSH2* variants carriers were analyzed with our CRC associated genes NGS custom panel (Tables 1 and 3). The *MSH2* c.211G > C variant was identified in Case 109, which is responsible for the splicing defect previously observed (Fig. 1). In LLS cases, germline missense variants predicted as pathogenic by at least 3 functional *in silico* tools were found in distinct CRC genes: one in *BUB1,* three in the H3K36 trimethyltransferase *SETD2,* 1 in *FAN1* and two in *MUTYH* (monoallelic). Of note, probably pathogenic variants in *SETD2* and *FAN1* coexisted in one of the heterozygous *MUTYH* carriers (Case 105).

In all, germline and functional characterization classified 5 additional cases as LS, 4 harboring 3 reclassified MMR VUS and the missed *MSH2* mutation (Table 1). Also, predicted pathogenic variants were identified in other genes in 5 additional cases.

Molecular analysis of LLS tumors

Next, we explored whether combined germline and somatic testing could help in elucidating the molecular basis of the remaining cases. Somatic hits in DNA repair genes were found in 5 tumors from the 4 LLS individuals tested (Table 4): double somatic hits in *MSH2* and *MSH6* (cases 111 and 114), apparent *MSH2* loss of heterozygosity (Case 108 C1/C2) and coexistence of double somatic mutations in other MMR genes and/or proof-reading polymerases (*POLD1* and *POLE*) (cases 108 C1/C2 and 121). Also, somatic mutations in other cancer genes (*APC, AXIN2, BMPR1A, PTEN* or *BUB1B*) and in CRC actionable genes coexisted with the aforementioned alterations (Tables 4 and 5). Interestingly, the two colorectal tumors from Case 108 showed completely different profiles: the MSI tumor (cancer 1) mainly harbored deletions at homopolymeric sequences, whereas the MSS tumor (cancer 2) harbored substitutions.

Somatic methylation did not account for any other case as promoter methylation in *MSH2* (0/8 tumors) or *MSH6* (0/11 tumors) was not detected (Supporting Information, Fig. S1 and Table 1). As previous studies have reported somatic methylation at the *MSH2* promoter in LS *MSH2* mutation carriers,[12] 8 additional tumor samples from LS *MSH2* mutation carriers from our LS series were studied, none of which were methylated (data not shown).

Discussion

A comprehensive germline and somatic mutational analysis allowed the molecular characterization of a high proportion of MSH2-deficient tumors in a series of LS suspected patients. The reclassification as pathogenic of 4 *MSH2* VUS and the identification of a new *MSH2* splicing mutation yielded a 71% (41/58) mutation detection rate. Furthermore, predicted pathogenic germline variants in DNA repair and genomic instability genes *BUB1*, *SETD2*, *FAN1* and *MSH6* were identified in 5 patients. Finally, the presence of double MMR or combined MMR/polymerase somatic hits in tumors from the informative LLS individuals analyzed may increase this yield up to 86% (50/58). The obtained results further evidence the great heterogeneity present in this subset of cases, as previously reported,[7,8,20–22,27] and reinforce the notion that negative germline DNA and RNA testing should be complemented with somatic analysis.

RNA analyses allowed classification of three *MSH2* variants as pathogenic mutations affecting mRNA processing. Splicing analysis in combination with multifactorial likelihood calculations offered a good performance, allowing reclassification of the 10 variants analyzed (6 class 4 variants and 4 class 3 variants) as disease causing mutations. These results highlight the benefit of applying quantitative and qualitative analyses for variant interpretation and classification. Moreover they showed the usefulness of the implementation of RNA analyses (either splicing or allelic imbalance) in the diagnostic routine, as previously also demonstrated for the identification of cryptic variants in MMR genes.[18,33] The *MSH2* variant c.211G > C, identified in two patients, illustrates the complexity of variant classification and the challenge associated with functional characterization. Splicing analysis of the whole transcript in Case 109 identified two aberrant transcripts (r.−16_211del and r.195_211del). In contrast, in Case 234, the splicing analysis performed encompassing a smaller region containing the variant (from exon 1—nucleotide c.85—to exon 4) identified only the r.195_211del transcript. The variant c.211G > C was finally classified as pathogenic based on the generation of aberrant transcripts.[19]

The germline mutational analysis of selected CRC-associated genes has yielded promising results in this set of MSH2-deficient LLS cases. Germline biallelic *MUTYH* mutation carriers were detected and excluded prior to this analysis.[20,27] The identification of a predicted pathogenic alteration in *FAN1* reinforces the notion that *FAN1* is a CRC predisposing gene.[34] To the best of our knowledge, this is the first report of a germline predicted pathogenic *BUB1*[35] variant in a patient with breast and endometrial cancers, which has been recently associated with early onset and familial CRC.[36] Moreover, 3 LLS patients diagnosed with CRC before age 50 harbored germline predicted pathogenic variants in *SETD2*, an H3K36 trimethyltransferase, which was included in our customized NGS panel for being frequently reported mutated in MSI CRC.[32] Its depletion results in MSI and elevated mutation rates *in vivo*, as H3K36me3 activity is necessary for recruiting MSH2/MSH6 to chromatin.[37] With the identification of rare and potentially pathogenic variants, *FAN1*, *BUB1* and *SETD2* are emerging as candidate genes responsible for LLS. Functional and cosegregation analysis are needed to elucidate the pathogenicity of the identified variants, and further collaborative efforts should be made to confirm their involvement in the inherited predisposition to cancer. Moreover, it must be borne in mind that both undetected germline alterations in the MMR genes —complex mutations, structural variations and

Table 3. Germline variants found with Haloplex and results from *in silico* predictions [Color table can be viewed at wileyonlinelibrary.com]

	Variant calling			Position		Coverage		*In silico* predictions				
									Protein function			
Patient ID	Gene	Transcript/ cDNA change	Predicted protein change	chr	Start	Allelic frequency	Read depth	Splicing	SIFT (score)	Mutation Taster (*p* value)	Polyphen2/ HumDiv (score)	Polyphen2/ HumVar (score)
A. PBL samples from Lynch-like syndrome patients												
121	*STK11*	NM_000455.4:c.−325A>C	p.?	19	1206588	0.282	442					
119	*AXIN2*	NM_004655.3:c.1780G>A	p.Ala594Thr	17	63533114	0.492	3404	Inconclusive	B (0.15)	B (1)	B (0.003)	B (0.003)
	FAN1	NM_014967.4:c.174G>A	p.=	15	31197040	0.489	3669	No change	-	-	-	-
	AXIN2	NM_004655.3:c.*884delT	p.?	17	63525208	0.485	3044					
	AXIN2	NM_004655.3:c.*476_ *487delTGAGCTAGGAGT	p.?	17	63525606	0.463	3684					
	BMPR1A	NM_004329.2:c.*85G>A	p.?	10	88683561	0.538	817					
102	***MSH6***	**NM_000179.2:c.−25C>T**	p.?	2	48010348	0.552	6303	Inconclusive	-	-	-	-
	MSH6	**NM_000179.2:c.−204C>G**	p.?	2	48010169	0.459	2705	Inconclusive	-	-	-	-
	SMAD4	NM_005359.5:c.*6293G>C	p.?	18	48611130	0.419	8047		-	-	-	-
101	*MSH6*	NM_000179.2:c.4002–10delT	p.?	2	48033891	0.693	913	No change	-	-	-	-
103	*CHEK2*	NM_007194.3 c.1510G>C	p.Glu504Gln	22	29085155	0.304	2101	Inconclusive	B (0.53)	B (1)	B (0.016)	B (0.005)
	EPCAM	NM_002354.2:c.831A>G	p.Ile277Met	2	47607081	0.192	2069	Inconclusive	D (0.04)	B (0.956)	PsD (0.610)	B (0.125)
	AXIN2	NM_004655.3:c.623C>T	p.Ala208Val	17	63554116	0.166	482	No change	B (0.06)	D (1)	B (0.228)	B (0.064)
	ENG	NM_000118.3:c.1844C>T	p.Ser615Leu	9	130578230	0.287	3034	No change	D (0)	D (0.745)	B (0.111)	B (0.011)
	FBXW7	NM_033632.2:c.1200C>T	p.=	4	153250860	0.136	1400	-	-	-	-	-
	POLD1	NM_001256849.1:c.−790T>C	p.?	19	50886861	0.198	4362					
104	*FAN1*	NM_014967.4:c.1129C>T	p.Arg377Trp	15	31197995	0.518	4623	No change	D (0)	D (0.993)	B (0.398)	B (0.037)
	APC	NM_000038.5:c.1959G>A	p.=	5	112173250	0.493	2504	Loss of acceptor splicing site	-	-	-	-
105	***MUTYH***	**NM_001128425.1:c.1227_1228dup**	**p.Glu410Glyfs*43**	1	45797186	0.496	3690	-	-	-	-	-
	FAN1	NM_014967.4:c.1856T>A	p.Met619Lys	15	31210411	0.558	5282	**Gain of acceptor splicing site**	**D (0)**	**D (1)**	**PsD (0.937)**	**B (0.409)**
	SETD2	**NM_014159.6:c.1204C>T**	**p.Arg402Trp**	3	47164922	0.509	6441	Inconclusive	**D (0)**	**D (0.99)**	**PrD (0.999)**	**PrD (0.923)**
123	*PMS2*	NM_00535.5:c.−493insG	p.?	7	6049143	0.453	203					
107	***SETD2***	**NM_014159.6:c.2798G>T**	**p.Gly933Val**	3	47163328	0.467	3621	**Loss of donor splicing site**	D (0.01)	B (1)	B (0.000)	B (0.000)
	ENG	NM_000118.3:c.1712G>A	p.Arg571His	9	130579457	0.483	10965	Inconclusive	D (0.02)	B (1)	B (0.225)	B (0.028)
	EPCAM	NM_002354.2:c.−280G>C	p.?	2	47596365	0.408	3278					
	MLH3	NM_001040108.1:c.*2485G>C	p.?	14	75481300	0.402	1155					
108	*CDH1*	NM_004360.3:c.2520C>T	p.=	16	68867273	0.468	5374	No change	-	-	-	-
	EPCAM	NM_002354.2:c.−485T>G	p.?	2	47596160	0.412	787					
	BUB3	NM_001007793.2:c.*173T>A	p.?	10	124924745	0.206	3229					
	ENG	NM_000118.3:c.*704delAGTT	p.?	9	130577492	0.491	4350					
110	***SETD2***	**NM_014159.6:c.2508T>G**	**p.Cys836Trp**	3	47163618	0.469	2135	No change	**D (0)**	**D (1)**	**PsD (0.833)**	**B (0.176)**
111	***MUTYH***	**NM_001128425.1:c.1187G>A**	**p.Gly396Asp**	1	45797228	0.541	2944	**Loss of donor splicing site**	**D (0)**	**D (1)**	**PrD (1.000)**	**PrD (0.999)**
	BUB3	NM_004725.3:c.*1124G>A	p.?	10	124924475	0.456	580					
	MLH3	NM_001040108.1:c.*2058G>T	p.?	14	75481727	0.413	3036					

Table 3. Germline variants found with Haloplex and results from *in silico* predictions [Color table can be viewed at wileyonlinelibrary.com] (Continued)

	Variant calling			Position		Coverage		*In silico* predictions				
									Protein function			
Patient ID	Gene	Transcript/ cDNA change	Predicted protein change	chr	Start	Allelic frequency	Read depth	Splicing	SIFT (score)	Mutation Taster (*p* value)	Polyphen2/ HumDiv (score)	Polyphen2/ HumVar (score)
112	*MLH3*	NM_001040108.1:c.2425A>G	p.Met809Val	14	75513934	0.508	1955	No change	B (0.3)	B (1)	B (0.000)	B (0.000)
	CDH1	NM_004360.3:c.2292C>T	p.=	16	68862204	0.408	1184		-	-	-	-
	BUB3	NM_004725.3:c.*371A>G	p.?	10	124923722	0.358	1641					
	PTEN	NM_000314.4:c.−632C>T	p.?	10	89623595	0.489	1225					
	ENG	NM_000118.3:c.−186G>A	p.?	9	130616820	0.515	1932					
	ENG	NM_000118.3:c.−289A>T	p.?	9	130616923	0.524	2234					
113	*POLD1*	NM_001256849.1:c.136G>A	p.Ala46Thr	19	50902244	0.467	4757	Inconclusive	B (0.22)	D (0.988)	B (0.295)	B (0.037)
	FAN1	NM_014967.4:c.603C>T	p.=	15	31197469	0.544	1515	No change			-	-
114	*FAN1*	NM_014967.4:c.434G>T	p.Arg145His	15	31197300	0.484	2112	No change	D (0.03)	B (1)	B (0.025)	B (0.007)
	SMAD4	NM_005359.5:c.*2218G>T	p.?	18	48607055	0.582	212					
	PMS1	NM_000534.4:C.−116G>C	p.?	2	190649224	0.515	2260					
115	*MLH3*	NM_001040108.1:c.1870G>C	p.Glu624Gln	14	75514489	0.376	1024	Inconclusive	B (0.05)	B (0.892)	PrD (0.990)	PsD (0.637)
	BUB1	**NM_004336.4:c.3005C>G**	**p.Thr1002Ser**	2	111397376	0.378	2652	**Loss of acceptor splicing site**	B (0.63	B (0.639)	B (0.005)	B (0.018)
116	*TP53*	NM_000546: c.−594insA	p.?	17	7591514	0.505	1692					
	MSH3	NM_002439.3:c.−457G>C	p.?	5	79950090	0.467	2088					
	TP53	NM_000546.5:c.*409C>A	p.?	17	7572518	0.51	937					
B. PBL samples from Lynch syndrome patients												
228	***MSH2***	**NM_000251.2:c.989T>C**	**p.Leu330Pro**	2	47643481	0.53	2033	-	**D (0)**	**D (1)**	**PrD (1.000)**	**PrD (1.000)**
	STK11	NM_000455.4:c.945G>A	p.=	19	1223008	0.469	4487	Inconclusive			-	-
	POLD1	NM_001256849.1:c.1138−8A>G	p.?	19	50906742	0.515	4834	Inconclusive			-	-
117	***MSH2***	**NM_000251.2:c.518T>G**	**p.Leu173Arg**	**2**	47637384	0.38	24	**Inconclusive**	**D (0)**	**D (1)**	**PrD (0.999)**	**PrD (0.992)**
	FAN1	NM_014967.4:c.1851C>T	p.=	15	31210406	0.546	2662	No change	-	-	-	-
	POLE	NM_006231.3:c.6072C>T	p.=	12	133209314	0.526	1415	No change	-	-	-	-
118	***MSH2***	**NM_000251.2:c.2069A>G**	**p.Gln690Arg**	2	47703569	0.427	1931	Inconclusive	**D (0)**	**D (1)**	**PrD (0.999)**	**PrD (0.992)**
	MLH1	NM_000249.3:c.*32_*34delCTT	p.?	3	37092170	0.501	1701					
109	***MSH2***	**NM_000251.2:c.211G>C**	**p.Gly71Arg**	2	47630541	0.432	520	**Loss of donor splicing site**	D (0.03)	D (1)	B (0.107)	B (0.076)
	PMS1	NM_000534.4:c.2186A>G	p.Asn690Ser	2	190728798	0.482	2250	Inconclusive	B (0.62)	B (1)	B (0.000)	B (0.000)
	TP53	NM_000546.5:c.*1175A>C	p.?	17	7571752	0.427	4674					
	APC	NM_000038.4:c.*1684A>G	p.?	5	112181507	0.321	594					
	ENG	NM_000118.3:c.*704delAGTT	p.?	9	130577491	0.995	6680					

Abbreviations: B, benign; D, damaging; PrD, probably damaging; PsD, possibly damaging
Brown-colored, frameshift and predicted probably pathogenic variants.

Table 4. Somatic variants found in the analysis of 26 CRC associated genes and results from *in silico* predictions [Color table can be viewed at wileyonlinelibrary.com]

	Variant calling			Position		Coverage		*In silico* predictions					
									Protein function				
Patient ID_ Tumor tested^	Gene	Transcript/cDNA change	Predicted protein change	chr	start	Allelic frequency	Read depth	Splicing	SIFT (score)	Mutation taster (p-value)	Polyphen2/ HumDiv (score)	Polyphen2/ HumVar (score)	LOH in *MSH2* locus°
A. Tumors from Lynch-like syndrome patients													
121_C1	***SETD1B***	NM_015048.1:c.22del	p.His8Thrfs*27	12	122242655	0.466	10593	No change	-	-	-	-	No apparent
	PMS2	NM_000535.5:c.325del	p.Glu109Lysfs*3	7	6043348	0.205	420	No change	-	-	-	-	
	PTEN	NM_000314.4:c.968del	p.Asn323Metfs*21	10	89720811	0.0581	172	No change	-	-	-	-	
	SETD2	NM_014159.6:c.3165T>A	p.Asp1055Glu	3	47162961	0.185	3519	No change	D (0)	D (0.992)	B (0.041)	B (0.044)	
	MSH6	NM_000179.2:c.1082G>A	p.Arg361His	2	48026204	0.207	8083	No change	B (0.21)	B (1)	PsD (0.837)	B (0.243)	
	POLD1	NM_001256849.1:c.1330C>T	p.Arg444Trp	19	50909526	0.196	6980	Inconclusive	D (0)	D (1)	PrD (1.000)	PrD (0.999)	
	MLH3	NM_001040108.1: c.1755delA	p.Glu586Asnfs*24	14	75514602	0.27	5221	No change	-	-	-	-	
	BUB3	NM_001007793.2:c.973T>C	p.Ser325Pro	10	124924564	0.0583	634	Loss of acceptor splice site(N)	B (0.07)	D (0.999)	PsD (0.782)	PsD (0.838)	
	STK11	NM_000455.4:c.−325A>C	p.?	19	1206588	0.214	3743	Inconclusive					
	AXIN2	NM_004655.3:c.*633del	p.?	17	63525459	0.241	11042	-					
	AXIN2	NM_004655.3:c.−618del	p.?	17	63558067	0.154	18797	-					
	STK11	NM_000455.4:c.−117del	p.?	19	1206796	0.236	6632	Inconclusive					
108_C2	***BUB1B***	NM_001211.5:c.1738G>T	p.Glu580*	15	40498388	0.0556	107	Inconclusive	B (1)	-	-	-	Possible
	MLH1	NM_001167618.1:c.1253G>A	p.Arg418Gln	3	37090087	0.0976	204	Inconclusive	B (0.07)	D (1)	PrD (1.000)	PrD (0.986)	
	MSH6	NM_000179.2:c.2625G>T	p.Met875Ile	2	48027747	0.0731	423	No change	B (0.17)	D (1)	B (0.001)	B (0.004)	
	BMPR1A	NM_004329.2:c.878C>T	p.Ala293Val	10	88678938	0.272	440	No change	D (0)	D (1)	PrD (1.000)	PrD (1.000)	
	POLE	NM_006231.2:c.2284C>T	p.Arg762Trp	12	133244124	0.0511	704	No change	D (0)	D (1)	PrD (1.000)	PrD (1.000)	
	TP53	NM_000546.5:c.993 + 284C>T	p.?	17	7576569	0.131	106	Inconclusive	-	-	-	-	
108_C1	***SETD1B***	NM_015048.1:c.22del	p.His8Thrfs*27	12	122242656	0.309	6428	No change	-	-	-	-	Possible
	MSH3	NM_002439.4:c.1114delAA	p.Lys383Argfs*32	5	79970914	0.158	796	Inconclusive	-	-	-	-	
	PMS2	NM_000535.5:c.1501G>A	p.Val501Met	7	6026895	0.114	6174	No change	B (0.12)	B (1)	B (0.003)	B (0.002)	
	MLH1	NM_001167618.1:c.697C>T	p.Arg233Trp	3	37070285	0.0758	131	No change	D (0.02)	D (1)	PrD (0.990)	PsD (0.513)	
	STK11	NM_000455.4:c.*787G>A	p.?	19	1228359	0.11	4842	No change					
	MSH2	NM_000251.2:c.−440delT	p.?	2	47629890	0.127	2152	Inconclusive					
	AXIN2	NM_004655.3:c.*631delT	p.?	17	63525462	0.277	2796	-					
	AXIN2	NM_004655.3:c.−330delA	p.?	17	63558069	0.167	3739	-					
	APC	NM_000038.5:c.*1884delT	p.?	5	112181707	0.087	137	-					
	STK11	NM_000455.4:c.−117del	p.?	19	1206796	0.229	667	Inconclusive					
111_C3	***SETD1B***	NM_015048.1:c.22del	p.His8Thrfs*27	12	122242656	0.83	1620	No change	-	-	-	-	Not analyzed
	MSH2	NM_000251.2:c.1600delC	p.Arg534Valfs*9	2	47693885	0.394	747	No change	-	-	-	-	
	MSH2	NM_000251.2:c.1741delA	p.Ile581Leufs*9	2	471698181	0.45	9	No change	-	-	-	-	
	MLH3	NM_001040108.1:c.1755del	p.Glu586Asnfs*24	14	75514603	0.39	136	No change	-	-	-	-	
	MSH3	NM_002439: c.1114delAA	p.Lys383Argfs*32	5	79970914	0.682	456	Inconclusive	-	-	-	-	
	BMPR1A	NM_004329.2:c.419del	p.Pro140Leufs*4	10	88659631	0.23	209	No change	-	-	-	-	

Table 4. Somatic variants found in the analysis of 26 CRC associated genes and results from *in silico* predictions [Color table can be viewed at wileyonlinelibrary.com] (Continued)

	Variant calling			Position		Coverage		*In silico* predictions					
									Protein function				
Patient ID_ Tumor tested^	Gene	Transcript/cDNA change	Predicted protein change	chr	start	Allelic frequency	Read depth	Splicing	SIFT (score)	Mutation taster (p-value)	Polyphen2/ HumDiv (score)	Polyphen2/ HumVar (score)	LOH in *MSH2* locus°
	CHEK2	NM_007194.3:c.880G>A	p.Ala294Thr	22	29099521	0.157	126	No change	B (0.32)	D (0.993)	B (0.002)	B (0.001)	
	MLH1	NM_001167618.1:c.443G>A	p.Arg148Gln	3	37067255	0.205	515	No change	B (0.22)	D (1)	PsD (0.602)	B (0.100)	
	MUTYH	NM_001128425.1:c.643G>A	p.Val187Met	1	45798293	0.346	1624	Inconclusive	D (0)	D (1)	PrD (1.000)	PrD (0.999)	
	POLE	NM_006231.3:c.2375A>G	p.Lys792Arg	12	133241981	0.47	2116	Gain of donor splice site	B (0.11)	D (1)	PrD (0.971)	PsD (0.887)	
	BUB3	NM_001007793.2:c.972−88G>A	p.?	10	124924475	0.603	67	-					
	SMAD4	NM_005359.5:c.*3760delT	p.?	18	48608588	0.331	181	-					
	AXIN2	NM_004655.3:c.*631delAA	p.?	17	63525462	0.441	1431	-					
	AXIN2	NM_004655.3:c.−619delT	p.?	17	63558069	0.129	1959	-					
	MLH3	NM_001040108.1:c.*2058G>T	p.?	14	75481727	0.551	496	-					
	MLH3	NM_001040108.1:c.−71G>A	p.?	14	75518090	0.421	1940	-					
	APC	NM_000038.5:c.*1884delT	p.?	5	112181707	0.299	147	-					
	STK11	NM_000455.4:c.−117del	p.?	19	1206796	0.479	572	Inconclusive					
114_C1	***MSH6***	NM_000179.2:c.741delA	p.Lys247Asnfs*32	2	48025856	0.104	881	No change	-	-	-	-	No apparent
	AXIN2	NM_004655.3:c.1994delG	p.Gly665Alafs*24	17	63532584	0.121	1200	No change	-	-	-	-	
	MLH1	NM_001167617.1:c.713G>A	p.Gly238Asp	3	37061923	0.103	496	Inconclusive	D (0.01)	D (1)	PsD (0.884)	PsD (0.596)	
	MSH6	NM_000179.2:c.2765G>A	p.Arg922Gln	2	48027887	0.0724	607	No change	D (0.04)	D (1)	PsD (0.680)	B (0.190)	
	AXIN2	NM_004655.3:c.*631del	p.?	17	63525462	0.21	1109	-					
	AXIN2	NM_004655.3:c.957−3558_957−3559del	p.?	17	63558069	0.129	1673	-					
	SMAD4	NM_005359.3:c.*5757del	p.?	18	48610584	0.0693	722	-					
B. Tumor from a Lynch syndrome patient (carrier of germline *MSH2*c.989T>C)													
228_C1	***MSH3***	NM_002439.4:c.1141delA	p.Lys383Argfs*32	5	79970914	0.278	3154	Inconclusive	-	-	-	-	Not analyzed
	MUTYH	NM_001128425.1:c.1484G>A	p.Arg467His	1	45796222	0.242	14879	Inconclusive	D (0.02)	B (0.901)	B (0.218)	B (0.049)	
	MSH2	NM_000251.2:c.1601delG	p.Arg534Leufs*9	2	47693885	0.265	11983	Inconclusive					
	POLE	NM_006231.3:c.2865−4delT	p.?	12	133237747	0.506	24174	No change					
	MSH3	NM_002439.4:c.238−7G>A	p.?	5	79952223	0.238	21208	Inconclusive	-	-	-	-	
	BUB3	NM_004725.3:c.*1131delT	p.?	10	124924482	0.765	2396	-					
	SMAD4	NM_005359.5:c.*5835delT	p.?	18	48610584	0.28	12938	-					
	AXIN2	NM_004655.3:c.*636delAA	p.?	17	63525458	0.404	28588	-					
	PTEN	NM_000314.4:c.*655delT	p.?	10	89725884	0.193	4724	-					
	PTEN	NM_000314.4:c.*1631delT	p.?	10	89726860	0.196	1518	-					

^The number of "C" in tumor tested corresponds to the Cancer number of Table 1. (°) See Supporting Information, Table 5.
Abbreviations: B, benign; D, damaging; PrD, probably damaging; PsD, possibly damaging.
Brown-colored, frameshift and predicted probably pathogenic variants.

Table 5. Somatic mutations in targeted exons from CRC actionable genes and results from *in silico* predictions [Color table can be viewed at wileyonlinelibrary.com]

	Variant calling			Position		Coverage		*In silico* predictions				
									Protein function			
Patient ID_ Tumor tested^	Gene	Transcript/cDNA change	Predicted protein change	chr	start	Allelic frequency	Read depth	Splicing	SIFT (score)	Mutation taster (*p* value)	Polyphen2/ HumDiv (score)	Polyphen2/ HumVar (score)
A. Tumors from Lynch-like syndrome patients												
121_C1	*APC*	NM_001127511.2:c.4121C>A	p.Ser1374*	5	112175466	0.164	2068	Inconclusive	-	-	-	-
	KRAS	NM_004985.4:c.38G>A	p.Gly13Asp	12	25398281	0.145	1164	Inconclusive	D (0)	D (1)	B (0.215)	B (0.175)
	FBXW7	NM_001013415.1:c.1391C>T	p.Ser464Leu	4	153245446	0.352	4613	No change	D (0.01)	D (1)	PrD (1.000)	PrD (0.988)
	PIK3CA	NM_006218.2:c.113G>A	p.Arg38His	3	178916726	0.227	2019	Inconclusive	D (0.03)	D (1)	PrD (1.000)	PrD (0.992)
	GNAS	NM_001077489.2:c.429A>C	p.=	20	57480479	0.214	13390	Inconclusive	-	-	-	-
108_C2	*APC*	NM_001127511.2:c.2572C>T	p.Arg858*	5	112173917	0.0903	597	Inconclusive	B (0.1)	-	-	-
	TP53	NM_000546.5:c.856G>A	p.Glu154Lys	17	7577082	0.348	1087	No change	D (0)	D (1)	PrD (0.999)	PrD (0.982)
	KRAS	NM_004985.4:c.35G>A	p.Gly12Asp	12	25398284	0.248	104	No change	D (0)	D (1)	B (0.385)	B (0.257)
108_C1	*CTNNB1*	NM_001098209.1:c.122C>T	p.Thr41Ile	3	41266125	0.114	454	No change	D (0)	D (1)	PrD (0.996)	PrD (0.955)
	FBXW7	NM_001013415.1:c.1711C>T	p.Arg571Trp	4	153244092	0.0628	477	No change	D (0)	D (1)	PrD (1.000)	PrD (1.000)
111_C3	*KRAS*	NM_004985.4:c.35G>A	p.Gly12Asp	12	25398284	0.326	42	No change	D (0)	D (1)	B (0.385)	B (0.257)
	PIK3CA	NM_006218.2:c.3145G>C	p.Gly1049Arg	3	178952090	0.401	226	Inconclusive	D (0.01)	D (1)	B (0.300)	B (0.096)
114_C1												
B. Tumor from a Lynch syndrome patient												
228_C1	*PTEN*	NM_000314:c.636delT	p.Pro213Leufs*8	10	89717610	0.19	459	Inconclusive	-	-	-	-

^ The number of "C" in tumor tested corresponds to the cancer number of Table 1.
Abbreviations: B, benign; D, damaging; PrD, probably damaging; PsD, possibly damaging.
Brown-colored, frameshift and predicted probably pathogenic variants.

variants in regulatory regions— and alterations in genes not analyzed in this study could be having a role in LLS.

The accumulation of somatic alterations in DNA repair genes can certainly mimic germline associated phenotypes. Subexome analysis at a high coverage has shown to be useful for the identification and characterization of these cases. Indeed, somatic double hits in MMR genes were evidenced in two of 5 tumors, confirming previous observations.[6–9] In the remaining three, putative loss of heterozygosity and double heterozygous MMR genes and/or proofreading polymerases were identified in accordance with a recent report.[38] The limited number of cases analyzed precludes drawing conclusions on these findings although it must be considered that pediatric tumors arising in CMMRD cases strongly associate with mutations in the exonuclease domain of proofreading polymerases.[39] In line with previous reports, our observations reinforce the notion that somatic variants in *MSH2* or *MSH6* may be a frequent event in LLS cases, while somatic promoter hypermethylation does not play a significant role.

The lack of detection of *MSH2* methylation in LLS MSH2-deficient tumors is in agreement with the low proportion of methylated tumors in MSH2-deficient LLS patients (1 of 46) reported in two previous series[12,13]. When testing for methylation, the dependability of the technique is critical. MS-MCA is a robust technique that could simultaneously analyze several CpGs.[40] The use of methylation-independent primers further increases its consistency validated by the inclusion of adequate positive and negative controls in each run. Moreover, none of the 8 available tumors from *MSH2* mutated LS cases were methylated, in contrast to a previous report.[12] It must be emphasized that they analyzed an upstream region not included in our amplicon and, in consequence, the results could not be directly compared.

Our study highlights the importance of an in-depth strategy, combining germline and somatic mutational analysis by parallel high-throughput deep sequencing and characterization of variants identified. The yield of subexome testing is directly related to the selection of genes, the sample type analyzed and the quality and depth of the analysis. With a mean coverage of 1200×, we have probably ruled out most germline mosaicisms with a 5% cutoff value in PBLs. While mean coverage was similar for PBL and FFPE DNA, it was highly variable in FFPE samples depending upon the amplicon chosen. The combined germline-somatic analysis allows for a *bona fide* identification of somatic variants. However, the better the quality of FFPE DNA, the higher the yield of the analysis. The Haloplex technology partially bypasses possible artifacts related to sample processing by using many probes of different lengths at distinct regions minimizing lack of hybridization due to DNA fragmentation.

Our study also shows some limitations. The identified somatic mutations have been detected in amplicons with good coverage (1400×) making our findings dependable, although variability may have led to the loss of other relevant findings. The lack of available tumor sample has precluded the identification of second hits within the tumor in the majority of cases. Of note, the prevalence of double somatic MMR mutations in the analyzed tumors is similar to other series.[6,9] It must be acknowledged that our custom-made subexome panel can be improved by including novel putative CRC predisposition genes. Also, the inclusion of homopolymers as target regions could help in ascertaining MSI in tumor samples, as recently reported.[41] Finally, the yield observed in MSH2-deficient tumors needs to be confirmed when loss of other MMR proteins is observed.

In all, comprehensive germline and somatic analysis has proved useful in the elucidation of the underlying molecular basis of suspected LS in MSH2-deficient cases. Subexome analysis opens the scope of the genes underlying the development of these tumors, expanding the spectrum of overlapping phenotypes in these selected cases. Further studies of larger series and more in-depth functional characterization of variants detected are mandatory to establish the true clinical validity of the proposed algorithm. Our approach further illustrates the relevance of germline and somatic testing when deciphering the genetic basis of LLS or other CRC predisposition syndromes.

Acknowledgements

We are indebted to the patients and their families. We thank all the members of the Hereditary Cancer Program at the Catalan Institute of Oncology. We thank Eduard Serra, Elisabeth Castellanos and Bernat Gel for their support with NGS panel design and data analysis. BAT is a National Health and Medical Research Council CJ Martin Early Career Fellow.

Statement of Authors Contributors

GV, MP and GC conceived and designed experiments, analyzed and interpreted data and drafted the manuscript. GV, CG and AF, ED carried out experiments. MG, BAT, TP, XS, JV, NP, XC, AV, LF and CL analyzed and interpreted data. SI, AV, AS, MN and JB contributed to patient recruitment and acquisition of clinical data. MM and EH contributed to the acquisition of molecular data. All authors were involved in revising the manuscript and give final approval of the submitted and published versions.

References

1. 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.
2. 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.
3. Moreira L, Balaguer F, Lindor N, et al. Identification of Lynch syndrome among patients with colorectal cancer. *JAMA* 2012; 308: 1555–65.
4. Buchanan DD, Rosty C, Clendenning M, et al. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). *Appl Clin Genet* 2014; 7:183–93.
5. Rodriguez-Soler M, Perez-Carbonell L, Guarinos C, et al. Risk of cancer in cases of suspected lynch syndrome without germline mutation. *Gastroenterology* 2013; 144:926–32 e1; quiz e13–4.
6. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as

a cause of microsatellite-unstable cancers. *J Pathol* 2014; 234:548–59.
7. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. *Gastroenterology* 2014; 147:1308–16.e1.
8. Mensenkamp AR, Vogelaar IP, van Zelst–Stams WAG, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. *Gastroenterology* 2014; 146:643–6.e8.
9. Sourrouille I, Coulet F, Lefevre JH, et al. Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. *Fam Cancer* 2013; 12:27–33.
10. Gausachs M, Mur P, Corral J, et al. MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. *EJHG* 2012; 20:762–8.
11. 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.
12. Nagasaka T, Rhees J, Kloor M, et al. Somatic hypermethylation of MSH2 is a frequent event in Lynch Syndrome colorectal cancers. *Cancer Res* 2010; 70:3098–108.
13. Rumilla K, Schowalter KV, Lindor NM, et al. Frequency of deletions of EPCAM (TACSTD1) in MSH2-associated Lynch syndrome cases. *J Mol Diagn* 2011; 13:93–9.
14. Chen J-M. The 10-Mb paracentric inversion of chromosome arm 2p in activating MSH2 and causing hereditary nonpolyposis colorectal cancer: re-annotation and mutational mechanisms. *Genes Chromosomes Cancer* 2008; 47:543–5.
15. Clendenning M, Buchanan DD, Walsh MD, et al. Mutation deep within an intron of MSH2 causes Lynch syndrome. *Fam Cancer* 2011; 10:297–301.
16. Rhees J, Arnold M, Boland CR. Inversion of exons 1–7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. *Fam Cancer* 2014; 13:219–25.
17. Wagner A, Hendriks Y, Meijers-Heijboer EJ, et al. Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. *J Med Genet* 2001; 38:318–22.
18. Liu Q, Hesson LB, Nunez AC, et al. A cryptic paracentric inversion of MSH2 exons 2–6 causes Lynch syndrome. *Carcinogenesis* 2016; 37:10–7.
19. Thompson BA, Spurdle AB, Plazzer J-P, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet* 2014; 46:107–15.
20. Castillejo A, Vargas G, Castillejo I, et al. Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. *Eur J Cancer* 2014; 50:2241–50.
21. Elsayed F, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. *EJHG* 2014; 1–5.
22. Morak M, Heidenreich B, Keller G, et al. Biallelic MUTYH mutations can mimic Lynch syndrome. *EJHG* 2014; 22:1334–7.
23. Palles C, Cazier J-B, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2012; 45: 136–44.
24. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet* 2013; 14:295–300.
25. Stadler ZK, Schrader K, Vijai J, et al. Cancer genomics and inherited risk. *JCO* 2014; 32:687–98.
26. 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.
27. Seguí N, Navarro M, Pineda M, et al. Exome sequencing identifies MUTYH mutations in a family with colorectal cancer and an atypical phenotype. *Gut* 2015; 64:355–6.
28. Borràs E, Pineda M, Brieger A, et al. Comprehensive functional assessment of *MLH1* variants of unknown significance. *Hum Mut* 2012; 33:1576–88.
29. Borràs E, Pineda M, Cadiñanos J, et al. Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. *JMG* 2013; 50:552–63.
30. Vazquez M, Valencia A, Pons T. Structure-PPi: a module for the annotation of cancer-related single-nucleotide variants at protein-protein interfaces. *Bioinformatics (Oxford, England)* 2015; 31: 2397–9.
31. Plon SE, Cooper HP, Parks B, et al. Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. *Genet Med* 2008; 13:148–54.
32. Choi YJ, Oh HR, Choi MR, et al. Frameshift mutation of a histone methylation-related gene SETD1B and its regional heterogeneity in gastric and colorectal cancers with high microsatellite instability. *Hum Pathol* 2014; 45:1674–81.
33. Kwok CT, Ward RL, Hawkins NJ, et al. Detection of allelic imbalance in MLH1 expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Fam Cancer* 2014; 9:345–56.
34. Seguí N, Mina LB, Lázaro C, et al. Germline mutations in FAN1 cause hereditary colorectal cancer by impairing DNA repair. *Gastroenterology* 2015; 1–4.
35. Leland S, Nagarajan P, Polyzos A, et al. Heterozygosity for a Bub1 mutation causes female-specific germ cell aneuploidy in mice. *PNAS* 2009; 106: 12776–81.
36. de Voer RM, Geurts van Kessel A, Weren RDa, et al. Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. *Gastroenterology* 2013; 145:544–7.
37. Li F, Mao G, Tong D, et al. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* 2013; 153:590–600.
38. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. *Eur J Hum Genet* 2015; 24:1089–92.
39. Shlien A, Campbell BB, de Borja R, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet* 2015; 47: 257–62.
40. Pineda M, Mur P, Iniesta MD, et al. MLH1 methylation screening is effective in identifying epimutation carriers. *EJHG* 2012; 1256–64.
41. Kloth M, Ruesseler V, Engel C, et al. Activating ERBB2/HER2 mutations indicate susceptibility to pan-HER inhibitors in Lynch and Lynch-like colorectal cancer. *Gut* 2015; 1–10.

SUPPLEMENTARY MATERIAL

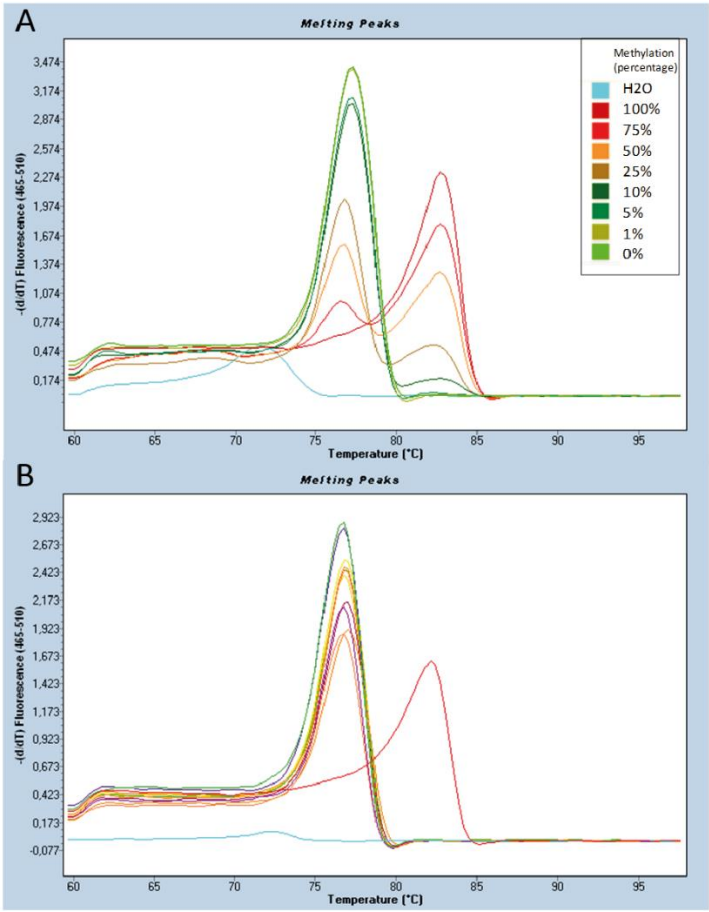


Figure S1. Methylation Specific-Melting Curve Analysis (MS-MCA) of *MSH2* promoter. **A.** Analytical sensitivity of MS-MCA for the detection of *MSH2* promoter methylation. Serial dilutions of a methylated control (CpG Methylated Jurkat Genomic DNA, New England Biolabs) were made with an unmethylated reference DNA. The methylated and unmethylated peaks have melting temperature of 82.6°C 76.8°C, respectively. The MS-MCA assay can detect up to 5% of the methylated alleles. B. Results obtained in the analysis of MSH2-deficient tumors. The methylated control is shown in red. All the samples analyzed show the presence of the unmethylated peak.

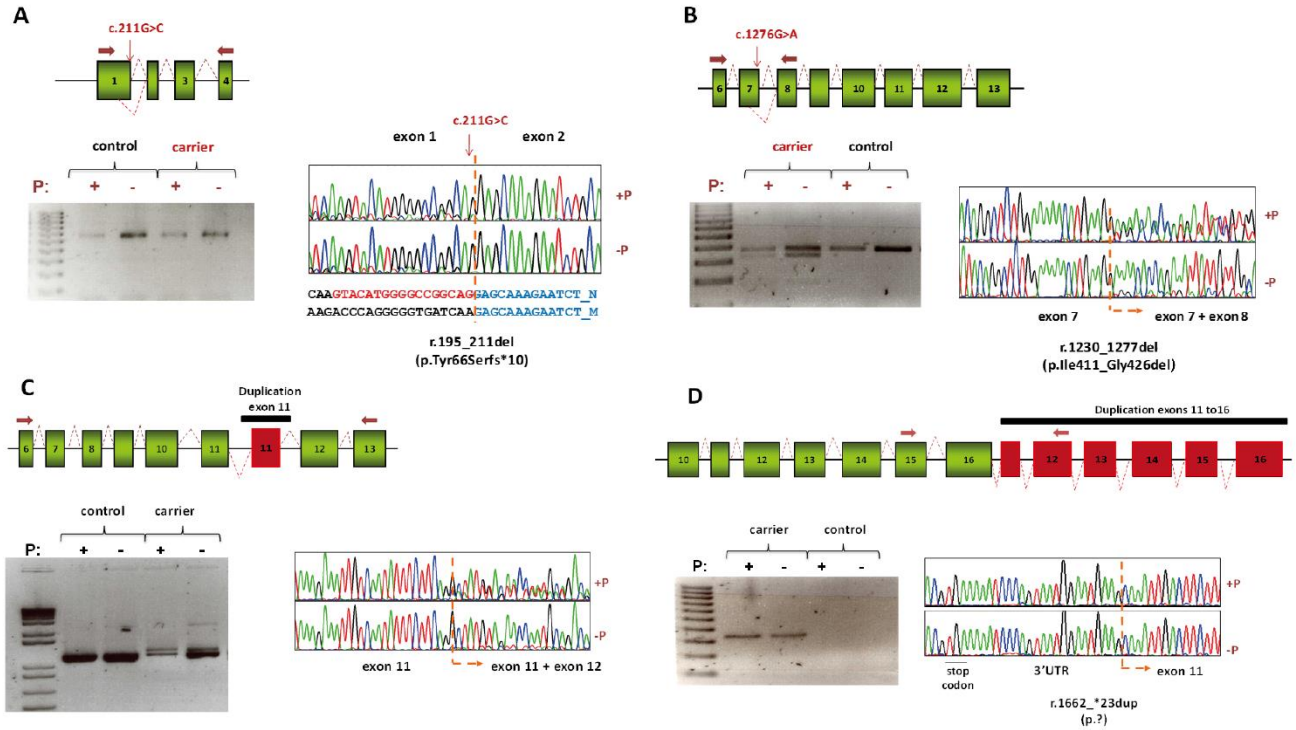


Figure S2. cDNA characterization of the *MSH2* c.211G>C (A), c.1276G>A (B), duplication of exon 11 (C), duplication of exons 11 – 16 (D). On the top, a schematic representation of normal transcripts (upper dotted lines) and aberrant transcripts (lower dotted lines) caused by the variants is shown. Green and red boxes indicate exons. On the bottom left, the gels showing RT-PCR products from controls and carriers in absence and presence of puromycin. On bottom right, direct sequencing of the RT-PCR products from variant carriers.

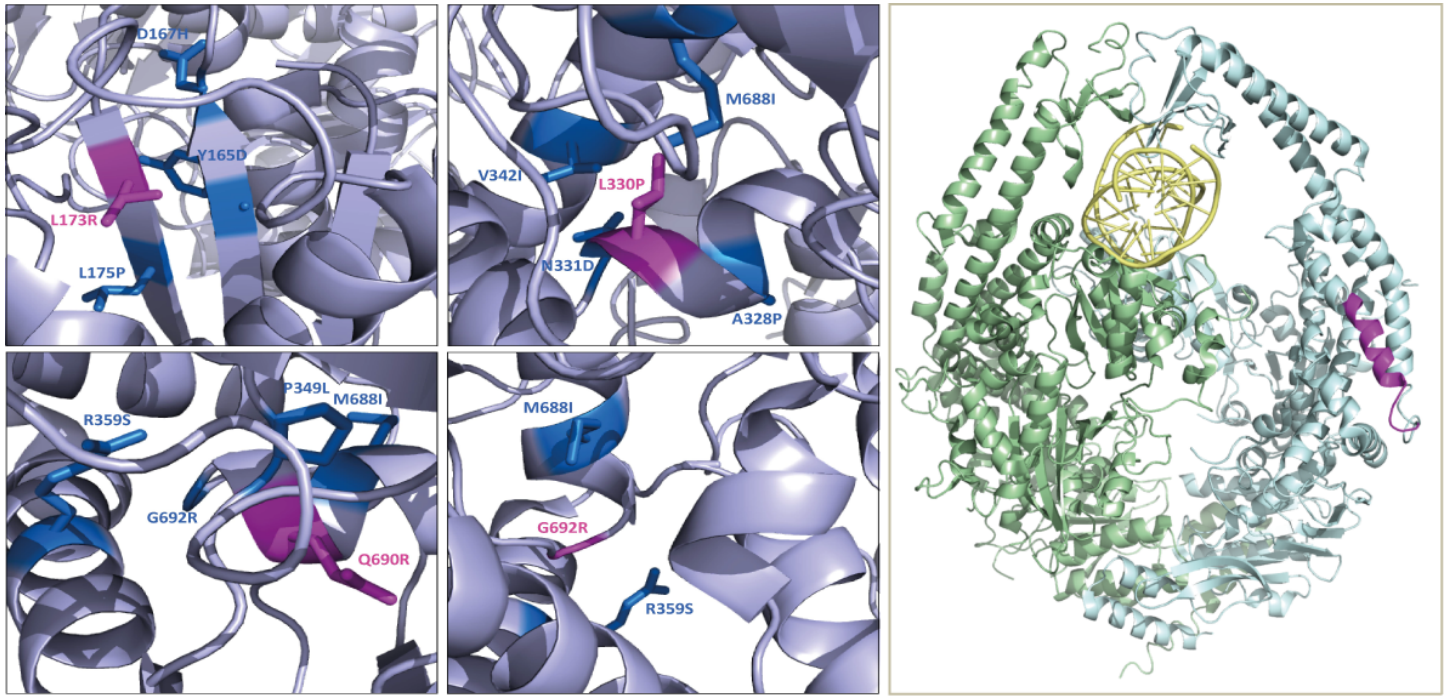


Figure S3. Protein structure location of MSH2 variants. The local environment of the four substitutions described in this work: L173R (top-left), L330P (top-right), Q690R (bottom-left) and G692R (bottom-right) is shown. In each case, the variant residue (in magenta) and their close neighbors which accommodate disease-associated mutations reported in UniProt (in blue) are shown with sticks. Among them, underlined variants correspond to likely pathogenic or pathogenic variants according to InSight classification. To the right, an overall view of the location of the deletion I411_G426del is shown. The protein complex (green and light blue) around the DNA (yellow) is represented with a ribbon in which the pink helix corresponds to the deleted sequence.

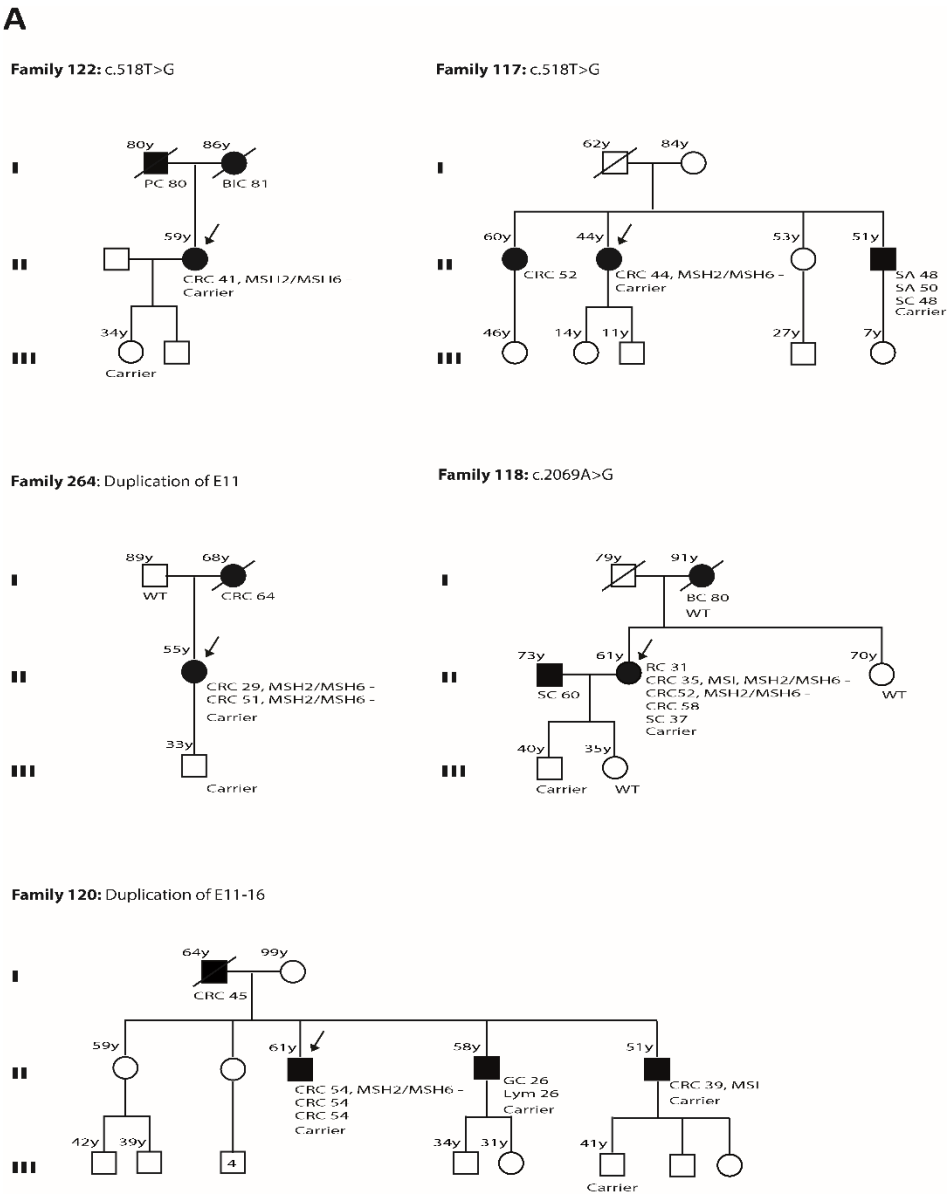


Figure S4. Family pedigrees from carriers of *MSH2* class 3 variants (A), class 4 variants (B) and further identified c.518T>G variant (C). Filled symbol, cancer; arrow, index case. Cosegregation results are indicated below individual's symbols as "carrier" or "WT". Current ages and ages at death, when available, are indicated on the top-left corner of each individual's symbol. CRC, colorectal cancer; PC, pancreas cancer; BC, breast cancer; SC, skin cancer; SA, sebaceous adenoma; BL, Bladder cancer; GC, gastric cancer; Lym, Lymphoma; UC, Uterine cancer; Me, melanome; Ap M, appendix malignant; OC, Ovarian Cancer; LiC, Liver cancer; HFN, head/face/neck cancer; PrC, prostate cancer; MSI, microsatellite instable; MSS, microsatellite stable; IHC+, conserved MMR protein expression; the pattern of expression of MSH2 and MSH6 proteins is indicated (-, loss; NV, non-valuable).

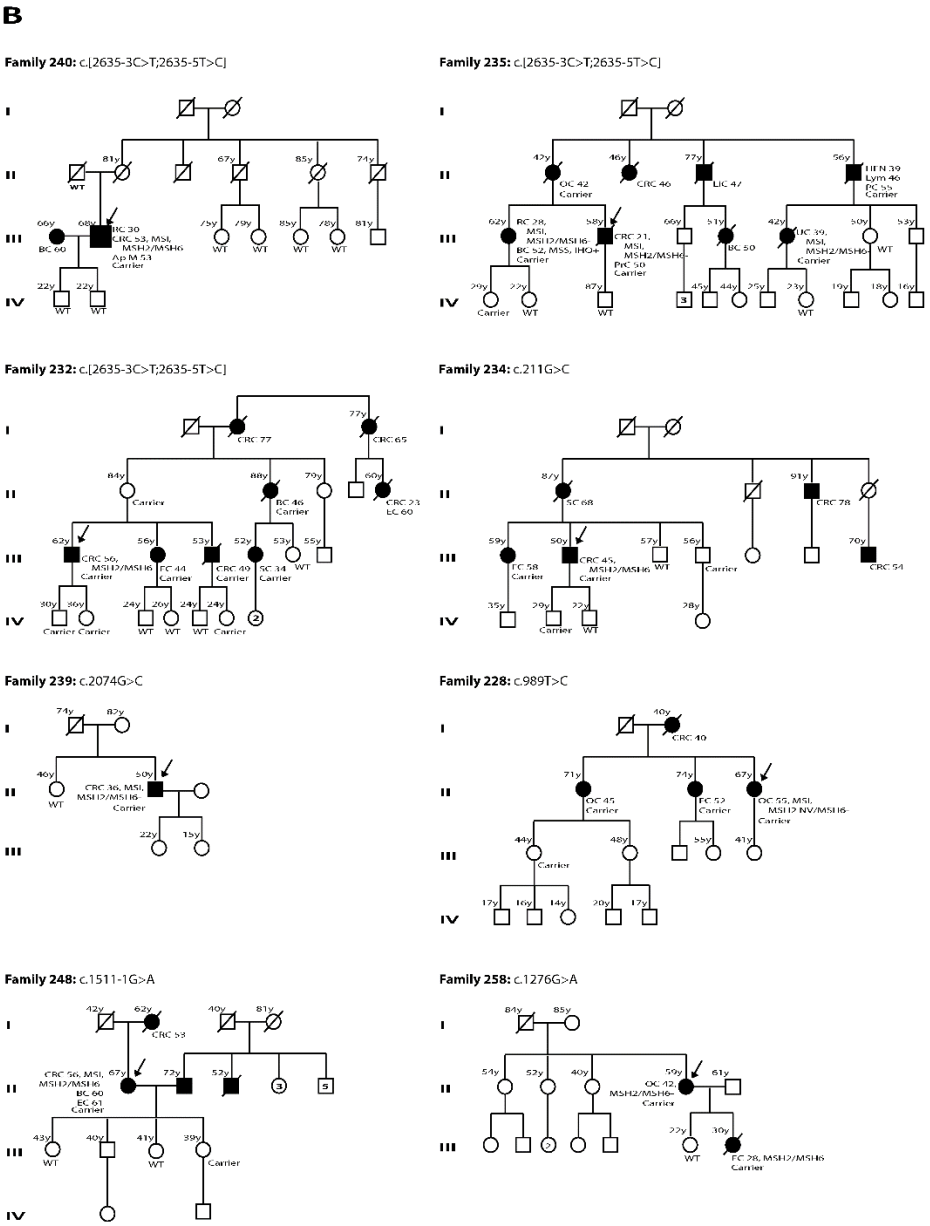


Figure S4. Family pedigrees from carriers of *MSH2* class 3 variants (A), class 4 variants (B) and further identified c.518T>G variant (C). **(Cont.)**

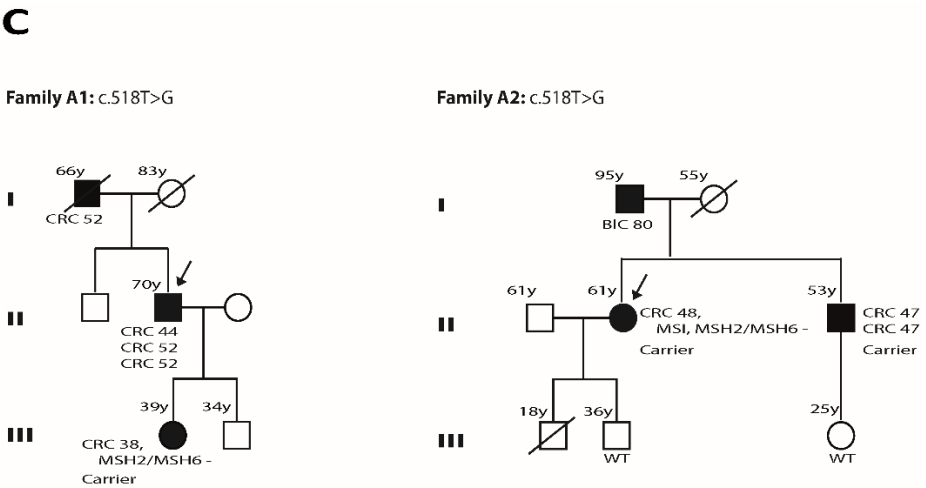


Figure S4. Family pedigrees from carriers of *MSH2* class 3 variants (A), class 4 variants (B) and further identified c.518T>G variant (C). **(Cont.)**

Table S1. Clinicopathological features of the included patients. A. Individual information. **B.** Clinical features according to the initial classification.

A.

Patient ID	Gender	Clinical criteria	Clinico-pathological characteristics*							
			Cancer 1 (age at diagnosis)	Cancer 2 (age at diagnosis)	Cancer 3 (age at diagnosis)	Cancer 4 (age at diagnosis)	Cancer 5 (age at diagnosis)	Cancer 6 (age at diagnosis)	IHC	MSI status
242	F	AC	CRC (28)	CRC (33)	**EC*** (50)	BIC (54)			MSH2/MSH6 loss	MSI
233	M	BC	**CRC*** (40)	PrC (51)					MSH2/MSH6 loss	NP
249	F	AC	CRC (43)	**CRC*** (44)	SC (51)				MSH2/MSH6 loss	MSI
253	M	BC	**CRC*** (31)						MSH2/MSH6 loss	NC
250	F	AC	**EC*** (43)	SA (36)	SA (48)	SA (50)	SA (54)		MSH2 loss/MSH6 NV	MSI
230	M	AC	**BIC*** (41)	SC (40)	SC (46)				MSH2 loss/MSH6 NP	NP
252	F	BC	**OC*** (42)						MSH2 loss/MSH6 NP	MSI
236	F	AC	**OC*** (43)	CRC (44)					MSH2 loss/MSH6 NV	MSI
231	F	BC	SA (50)	BC (49)	**EC*** (51)	SA (56)			MSH2 loss/MSH6 NP	NP
245	M	AC	CRC (39)	**CRC*** (45)					MSH2/MSH6 loss	MSI
260	M	BC	**CRC*** (42)	CRC (42)					MSH2/MSH6 loss	NP
262	M	BC	**CRC*** (34)	SA (44)	CRC (50)	CRC (51)			MSH2 loss/MSH6 NP	MSI
255	F	BC	**CRC*** (37)						MSH2/MSH6 loss	NP
257	M	BC	**CRC*** (41)						MSH2 loss/MSH6 NV	MSI
229	M	BC	SC (49)	SA (52)	**CRC*** (50)	UC (50)	BIC (50)	SC (52)	MSH2 loss/MSH6 NP	NP
256	M	AC	**CRC*** (59)	CRC (59)					MSH2/MSH6 loss	NP
263	F	BC	EC (56)	**CRC*** (64)					MSH2/MSH6 loss	NC
238	F	AC	CRC (33)	**UC*** (38)	UC* (38)	BIC (39)			MSH2/MSH6 loss	MSI
261	F	AC	EC (50)	**CRC*** (54)					MSH2 loss/MSH6 NP	NP
247	F	AC	CRC (51)	**EC*** (52)	RPC (60)	PC(60)			MSH2/MSH6 loss	MSI
237	M	AC	CRC (32)	CRC (34)	**CRC*** (42)	L (47)	CNSC (57)		MSH2/MSH6 loss	MSI
259	F	AC	CRC (37)	**CRC*** (48)	CRC(48)	SA (?)	SC (54)		MSH2/MSH6 loss	NP
241	F	AC	**CRC*** (27)						MSH2 loss/MSH6 NV	MSI
254	F	BC	**CRC*** (36)						MSH2/MSH6 loss	MSI
243	F	AC	OC (33)	**CRC*** (35)					MSH2/MSH6 loss	MSI
246	M	AC	**CRC*** (43)	CRC (43)	CRC (43)				MSH2/MSH6 loss	MSI
251	F	AC	**CRC*** (28)						MSH2/MSH6 loss	NP
234	M	BC	**CRC*** (45)						MSH2/MSH6 loss	NP
228	F	AC	**CRC*** (55)						MSH2/MSH6 loss	MSI
258	F	BC	**OC* (42)**						MSH2/MSH6 loss	NP

A. (cont.)

Patient ID	Gender	Clinical criteria	Clinico-pathological characteristics*							
			Cancer 1 (age at diagnosis)	Cancer 2 (age at diagnosis)	Cancer 3 (age at diagnosis)	Cancer 4 (age at diagnosis)	Cancer 5 (age at diagnosis)	Cancer 6 (age at diagnosis)	IHC	MSI status
248	F	BC	**CRC*** (56)	BC (60)	EC (61)				MSH2/MSH6 loss	MSI
239	M	BC	**CRC*** (36)						MSH2/MSH6 loss	MSI
232	M	BC	**CRC*** (56)						MSH2/MSH6 loss	NP
235	M	AC	**CRC*** (21)	PrC (50)					MSH2/MSH6 loss	MSI
240	M	BC	CRC (30)	**CRC*** (53)	CRC (53)				MSH2/MSH6 loss	MSI
122	F	BC	**CRC*** (41)						MSH2/MSH6 loss	NP
117	F	BC	**CRC*** (44)						MSH2/MSH6 loss	NP
264	F	BC	**CRC*** (29)	**CRC*** (51)					MSH2 loss/MSH6 NP & MSH2/MSH6 loss (respectively)	NP
120	M	AC	**CRC*** (54)	CRC (54)	CRC(54)				MSH2/MSH6 loss	NP
118	F	BC	CRC (31)	**CRC*** (35)	**CRC*** (52)	CRC (58)	SC (37)		MSH2 loss/MSH6 NV & MSH2/MSH6 loss (respectively)	MSI
121	F	PC	**CRC*** (77)						MSH2/MSH6 loss	NP
119	F	BC	**EC*** (45)						MSH2/MSH6 loss	MSI
102	F	BC	**CRC*** (55)						MSH2/MSH6 loss	NP
109	M	BC	**CRC*** (27)						MSH2/MSH6 loss	MSI
101	F	BC	**CRC*** (57)						MSH2/MSH6 loss	NP
103	F	PC	**CRC*** (73)						MSH2 loss/MSH6 NP	NP
104	F	BC	**CRC*** (51)						MSH2/MSH6 loss	NP
105	F	BC	**CRC*** (49)						MSH2/MSH6 loss	NP
123	M	BC	**CRC*** (59)						MSH2/MSH6 loss	NP
107	F	BC	**CRC*** (39)						MSH2 loss/MSH6 NV	MSI
108	F	BC	**CRC*** (32)	CRC (48)					MSH2 loss/MSH6 NV	MSI
110	M	BC	**CRC*** (43)						MSH2/MSH6 loss	NC
111	F	BC	CRC (51)	CRC (51)	**EC*** (56)				MSH2/MSH6 loss	NP
112	M	BC	**CRC*** (49)						MSH2/MSH6 loss	NC
113	F	BC	**CRC*** (49)	BC (55)					MSH2 loss/MSH6 NV	MSI
114	M	BC	**CRC*** (58)	CRC (58)					MSH2/MSH6 loss	MSI
115	F	AC	BC (62)	BC (69)	**EC*** (77)				MSH2/MSH6 loss	NP
116	F	BC	**CRC*** (48)						MSH2/MSH6 loss	MSI

Abbreviations: F, female; M, male; AC, Amsterdam criteria; BC, Bethesda criteria; PC, pathological criteria; CRC, colorectal cancer; EC, endometrial cancer; OC, ovarian cancer; SA, sebaceous adenoma; BC, breast cancer; SC, skin cancer; BlC, bladder cancer; PrC, prostate cancer; PC, pancreas cancer; CNSC, central nervous system cancer; L, lymphoma; UC, Ureteral Cancer; NP, not performed; MSI, microsatellite instability; NC, non-conclusive. Bold letter and (*) indicate tumors in which MSI/IHC was studied.

Table S1. Clinicopathological features of the included patients. A. Individual information. **B.** Clinical features according to the initial classification.

B.

Features	Total	LS	LLS		
	n (%)	n (%)	All n (%)	No variant identified n (%)	VUS (class 3) carrier n (%)
Total number of cases	58 (100)	35 (60.3)	23 (39.7)	18 (31.0)	5 (8.6)
Sex					
Female	36 (62.0)	19 (54.3)	17 (73.9)	13 (72.2)	4 (80)
Male	22 (38.0)	16 (45.7)	6 (26.1)	5 (27.8)	1 (20)
Mean age at diagnosisº* (range)	44 (21-77)^	45.8 (21-59)^	49.2 (31-77)^	51.7 (32-77)^	42.5 (31-54)^
Clinical criteria					
Amsterdam	20 (34.5)	18 (51.4)	2 (8.7)	1 (5.6)	1 (20)
Bethesda	36 (62.0)	17 (48.6)	19 (82.6)	15 (83.3)	4 (80)
Anatomo-pathological	2 (3.5)	0 (0)	2 (8.7)	2 (11.1)	0 (0)
Patients with multiple primary tumors*	26 (44.1)	20 (57.1)	6 (25)	3 (15.8)	3 (60)
MSH2-deficient analyzed tumors					
Colorectal cancer	47 (78.3)	25 (71.4)	22 (88.0)	15 (83.3)	7 (100)
Endometrial cancer	7 (11.7)	4 (11.4)	3 (12.0)	3 (16.7)	0 (0)
Ovarian cancer	4 (6.6)	4 (11.4)	0 (0)	0 (0)	0 (0)
Ureter cancer	1 (1.7)	1 (2.9)	0 (0)	0 (0)	0 (0)
Other	1 (1.7)	1 (2.9)	0 (0)	0 (0)	0 (0)

(º) First tumor diagnosis; (^) age range; (*) LS associated tumors

Table S2. Primers used in this study.

Gene	Analysis	Primer name	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon length (bp)	Number of CpGs interrogated
MSH2	MS-MCA	MS-MCA MSH2_PCR	TTTTTTTAATTAGGAGGTGAGGAG	CACCCCCTAAATCTTAAACACCT	221	24
		MS-MCA MSH2_Heminested	TTTTTTAGGGTATGTYGGAGAAG	CACCCCCTAAATCTTAAACACCT	125	13
	Sanger sequencing *MSH2* promoter (gDNA)	MSH2Pr-2_PCR&SEQ	GCCAAGAAGAGTCTGGGACA	ACGCGCATCCTTAGTAGAGC	404	
		MSH2Pr-2_SEQ	TTCAAGTTTCCTTCTGATG	GCCTTCCTCCTCCTCCCAG		
		MSH2Pr-1_PCR&SEQ	TCAAGCCTTGCAGCTGAGTA	CCATGTCGAAACCTCCTCAC	315	
		MSH2Ex1_PCR&SEQ	TCGCGCATTTTCTTCAACCA	GTCCCTCCCCAGCACG	285	
	c.211G>C (RNA splicing and stability)	gDNA_E1up_E1dw	TCGCGCATTTTCTTCAACCA	GTCCCTCCCCAGCACG	285	
		cDNA_MSH2_E1up/E4dw	TCGCGCATTTTCTTCAACCA	TCAAAGAGGAGGAATTCTGATCACAGA	712	
		cDNA_MSH2_E1up c.83/E4dw	AGAAGCCGACCACCACAGT	TCAAAGAGGAGGAATTCTGATCACAGA	590	
		ASE_c.211up	AAGTACATGGGGCCGGCAC			
	c.518T>G (RNA splicing and stability)	gDNA_E3up_E3dw	AATTTTAAAGTATGTTCAAGAG	CCTAGGCCTGGAATCTCCTC	379	
		cDNA_MSH2_E1up c.83/E4dw	TCGCGCATTTTCTTCAACCA	TCAAAGAGGAGGAATTCTGATCACAGA	590	
		ASE_c.518up	GTGGATTCCATACAGAGGAAAC			
	c.989T>C (RNA splicing and stability)	gDNA_E6up_E6dw	CGGATTAAGAGGTTGAAAGTTGGTC	CCCACGATTACACACAATATGAACA	590	
		cDNA_MSH2_E5up/E8dw	TCCAACTTTGGACAGTTTGAACT	TTCCTGAAACTTGGAGAAGTCA	507	
		ASE_c.989up	CTCAGTCTCTGGCTGCCTTGC			
	c.1276G>A (RNA splicing)	cDNA_MSH2_E6up/E8dw	TCAGTCTCTGGCTGCCTTG	TTCCTGAAACTTGGAGAAGTCA	388	
	c.2069A>G (RNA splicing and stability)	gDNA_E13up_E13dw	CGCGATTAATCATCAGTGT	CACAGGACAGAGACATACATT	357	
		cDNA_MSH2_E12up/E14dw	GCTATGTAGAACCAATGCAGACAC	CTCTTCAGTGGTGAGTGCTGT	705	
		ASE_c.2069dw		ATGGCACAAAACACCCAATT		
	duplication E11 (RNA splicing)	cDNA_MSH2_E6up/E13dw	TCAGTCTCTGGCTGCCTTG	AGCCCCTACTCGGGCTAAG	1174	
	duplication E11-16 (RNA splicing)	cDNA_MSH2_E15up/E12dw	CAGCAGCAAAGAAGTGCTATC	AGTGTCTGCATTGGTTCTACATAG	348	
MSH6	MS-MCA	MS-MCA MSH6_PCR	GGTAGGGYGGGTTTTTTAT	AAACTCCTAAAAACACCYCAT	238	29
		MS-MCA MSH6_Heminested	GGTAGGGYGGGTTTTTTAT	ACCCCAATAACCAATCAACA	154	18
	Sanger sequencing *MSH6* promoter (gDNA)	MSH6Pr-2_PCR&SEQ	GATTACAGGCGTGAGCCACT	CCTCTCTGGAGCGGAAGC	511	
		MSH6Pr-1_PCR&SEQ	CTCTAACGGCAGGAGGTCAC	CAGTGGCCAATCAACAGG	416	
		MSH6Pr-0.5_PCR&SEQ	GAAGGGGAGCTCAGCAGTTC	CTGTACAGGGTGCTCTGTCG	345	
EPCAM	Sanger sequencing *EPCAM* 3'UTR (gDNA)	EPCAM_3'UTR_PCR&SEQ	CCTGTTTCAGATAAAGGAGATGG	TTGAAATGTCAAAGTTAAGAAATTCAG	481	

Table S3. Genes and exons covered by NGS subexome panel

Gene	Transcript	Exons	Promoter
APC	NM_000038	All	Yes
BUB3	NM_004725	All	Yes
MUTYH	NM_001128425	All	Yes
STK11	NM_000455	All	Yes
POLE	NM_006231	All	Yes
POLD1	NM_002691	All	Yes
BMPR1A	NM_004329	All	Yes
SMAD4	NM_005359	All	Yes
PTEN	NM_000314	All	Yes
ENG	NM_000118	All	Yes
FAN1	NM_014967	All	Yes
TP53	NM_000546	All	Yes
CDH1	NM_004360	All	Yes
CHEK2	NM_001005735	All	Yes
BUB1B	NM_001211	All	Yes
BUB1	NM_004336	All	Yes
EXO1	NM_130398	All	Yes
AXIN2	NM_004655	All	Yes
EPCAM	NM_002354	All	Yes
MLH1	NM_000249	All	Yes
MLH3	NM_001040108	All	Yes
MSH2	NM_000251	All	Yes
MSH3	NM_002439	All	Yes
MSH6	NM_000179	All	Yes
PMS1	NM_000534	All	Yes
PMS2	NM_000535	All	Yes
AKT1	NM_005163	3	No
BRAF	NM_004333	11 and 15	No
CTNNB1	NM_001904	3	No
EGFR	NM_005228	3, 7, 15 and 18 to 21	No
FBXW7	NM_033632	8 to 12	No
GNAS	NM_000516	6 and 8	No
KRAS	NM_004985	2 to 4	No
MAP2K1 (MEK1)	NM_002755	2	No
MET	NM_000245	2, 5, 14, 16 to 19, and 21	No
NRAS	NM_002524	2, 3, 4 and 5	No
PIK3CA	NM_006218	2, 3, 8, 10, 14 and 21	No
SRC	NM_005417	14	No
SETD2	NM_014159	3	No
SETD1B	NM_015048	1	No
SETDB2	NM_031915	13	No

- Targeted regions of exons include -/+10 flanking bases.
- Promoter regions comprise 650bp upstream the TSS.

Table S4. Analysis of loss of heterozygosity at informative intragenic SNPs and nearby microsatellites

Patient ID - Tumor tested^	Intragenic MSH2 SNPs genotyped by our customized Next Generation Sequencing panel									Microsatellites (ratio tumor/blood)			Detection of LOH in *MSH2* locus
	Promoter		Intron 1	Intron 6	Intron 7	Intron 9	Intron 12		Intron 13	D2S2328	D2S288	D2S378	
	c.-433T>G	c.-118T>C	c.211+9	c.1077-80G>A	c.1277-118G>A	c.1511-9T	c.2006-265A>G	c.2006-6T	c.2210+175G>A	chr2:46,545,279-46,545,377	chr2:57,303,761-57,303,887	chr2:40,678,813-40,678,954	
	chr2:47629898	chr2:47630213	chr2:47630550	chr2:47656801	chr2:47672569	chr2:47693788	chr2:47703241	chr2:47703500	chr2:47703885				
108				0.47	0.51		0.54		0.47	_	_	_	Possible
108_C1				0.12 (20/167x)	NA		NA		0.07 (2/27x)	0,89	0,62	NI-MSI	
108_C2				0.18 (23/130x)	NA		NA		0.06 (4/72x)	0,77	0,83	0,99	Possible
114	0.504		0.486							_	_	_	Not apparent
114_C1	0.609		NA							0,93	NI - homo	NA	
121		0.47				0.448		0.466		_	_	_	Not apparent
121_C1		0.48				0.485		0.484		0,6	NI - homo	NI- MSI	
111					0.46		0.53		0.46	_	_	_	Not apparent
111_C3					0.53		NA		0.49	0,82	NI - homo	NI - homo	

Brown-coloured, SNPs and microsatellites showing allelic imbalance in tumor. Number of reads of imbalanced SNPs found in tumor samples is indicated between parenthesis (variant/total).

Abbreviations: LOH, loss of heterozygosity; NA, Nonassesable; NI, non informative; homo, homozygous; MSI, Microsatellite instability

Table S5. *In silico* predictions and result of the splicing analysis of class 3 and 4 *MSH2* variants identified in this study.
A. Result of the *in silico* predictions at the RNA level. **B.** Result of the *in silico* predictions at the protein level for non-truncating variants.

A

VUS	Exon	SS	Splice Site Prediction									Enhancer site prediction			cDNA splicing analysis	cDNA stability analysis (+/- puromicin)
			NNSplice		Spliceport		NetGene2		Softberry		Interpretation	Rescue ESE	ESE finder	Interpretation		
			wildtype	variant	wildtype	variant	wildtype	variant	wildtype	variant						
c.211G>C	E1	A	___	___	___	___	___	___	___	___	**Inconclusive**	No change	1 site destroyed / 3 created	**Inconclusive**	**r.195_211del; p.Tyr66Serfs*10**	NP
		D	0.95	0.59	NR	NR	0.00	0.00	11.56	NR						
c.518T>G	E3	A	0.98	0.98	NR	NR	0.00	0.00	4.5	4.5	**No effect**	1 created	1 created	**Aberrant ESE**	r.518T>G; p.Leu173Arg	Non allelic imbalance (1.1±0.2 / 1.2±0.2)
		D	1	1	NR	NR	0.00	0.00	14.64	14.64						
c.989T>C	E6	A	0.98	0.98	1.76	1.93	0.00	0.00	9.3	9.3	**No effect**	No change	1 destroyed	**Inconclusive**	r.989T>C; p.Leu330Pro	Non allelic imbalance (0.8±0.02 / 0.8±0.1)
		D	0.98	0.98	NR	NR	0.00	0.00	11.14	11.14						
c.1276G>A	E7	A	0.91	0.91	1.22	1.22	0.36	0.36	NR	NR	**Inconclusive**	No change	No change	**No change**	**r.1230_1277del; p.Ile411_Gly426del**	NP
		D	0.91	NR	NR	NR	0.00	NR	11.98	NR						
c.1511-1G>A	I9	A	NR	NR	1.54	0.67	0.34	NR	NR	NR	**Aberrant SS**	1 created	1 created	**Aberrant ESE**	NP	NP
		A2	__	__	__	__	__	0.44	__	__						
		D	0.6	0.6	NR	NR	NR	NR	NR	NR						
c.2069A>G	E13	A	0.95	0.95	1.2	1.44	0.77	0.77	7.88	7.88	**No effect**	No change	2 created	**Inconclusive**	r.2069A>G; p.Gln690Arg	Non allelic imbalance (1.1±0.2 / 1.1±0.1)
		D	1.00	1.00	1.64	1.64	0.00	0.00	15.06	15.06						
c.2074G>C	E13	A	0.95	0.95	1.20	1.32	0.77	0.77	7.88	7.88	**No effect**	No change	No change	**No change**	NP	NP
		D	1	1	1.64	1.64	0.53	0.53	15.06	15.06						
c.[2635-3C>T; 2635-5T>C]	I15	A	NR	NR	0.62	0.00	NR	NR	6.62	4.38	**Aberrant SS**	1 destroyed / 1 created	No change	**Inconclusive**	NP	NP
		D	NR	NR	NR	NR	NR	NR	NR	NR						
dup exon 11	E11		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	**r.1662_1759dup; p.Gly587Alafs*3**	NP
dup exons 11-16	E11-16		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	**r.1662_*23dup; p.?**	NP

B

VUS	Predicted impact on protein function							Structure prediction					
	Functional domain	PolyPhen-2 (score)	SIFT (score)	Condel (score)	MAPP_MMR (score)	PROVEAN (score)	Interpretation	PoPMuSic	CUPSAT	ERIS	I-MUTANT 3.0	FoldX 4	Interpretation
c.518T>G p.Leu173Arg	Connector domain	Probably Damaging (0.986)	Damaging (0)	Deleterious (0.68)	Deleterious (24.64)	Deleterious (-4,87)	**Impaired**	D (1.79)	D (-0.38)	D (3.86)	D (-2.34)	D (1.88)	Destabilizing
c.989T>C p.Leu330Pro	Lever domain	Probably Damaging (1.000)	Damaging (0)	Deleterious (0.69)	Deleterious (25.65)	Deleterious (-6,76)	**Impaired**	D (3.41)	D (-6.26)	D (>10)	D (-1.86)	D (7.93)	Destabilizing
c.1276G>A p.Ile411_Gly426del	Lever domain	NA	NA	NA	NA	Deleterious (-51.216)	**Impaired**	NA	NA	NA	NA	NA	NA
c.2069A>G p.Gln690Arg	ATPase domain	Probably Damaging (0,999)	Damaging (0)	Deleterious (0.75)	Deleterious (15.12)	Deleterious (-3,6)	**Impaired**	D (0.83)	D (-0.95)	D (>10)	D (-1.29)	D (4.61)	Destabilizing
c.2074G>C p.Gly692Arg	ATPase domain	Probably Damaging (1.000)	Damaging (0)	Deleterious (0.78)	Deleterious (36.520)	Deleterious (-7.734)	**Impaired**	D (2.28)	D (-6.98)	D (>10)	D (-1.44)	D (10.13)	Destabilizing

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; NP, not performed.

ARTÍCULO 6

Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals

Estela Dámaso, **Maribel González-Acosta**, Gardenia Vargas-Parra, Matilde Navarro, Judith Balmaña, Teresa Ramon y Cajal, Noemí Tuset, Fátima Marín, Anna Fernández, Carolina Gómez, Àngela Velasco, Ares Solanes, Sílvia Iglesias, Gisela Urgell, Consol López, Jesús del Valle, Olga Campos, Maria Santacana, Xavier Matias-Guiu, Conxi Lázaro, Laura Valle, Joan Brunet, Marta Pineda*, Gabriel Capellá*.
* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Manuscrito en preparación.

RESUMEN:

Este trabajo sostiene la hipótesis que la predisposición a desarrollar cáncer colorrectal (CCR) con deficiencia de reparación observada en los pacientes Lynch-like (SLL) podría ser causada por epimutaciones constitucionales no identificadas. Por tanto, el objetivo es dilucidar la causa subyacente a la deficiencia reparadora observada en individuos SLL mediante un análisis exhaustivo de los casos a nivel genético y epigenético.

En el estudio se incluyeron 115 pacientes que cumplieron con los criterios de SLL, 23 de los cuales habían sido previamente reportados en otro trabajo del grupo (Vargas-Parra et al 2017). El reanálisis mediante un panel personalizado de NGS de los casos con una fuerte historia familiar o personal de cáncer reveló la presencia de dos mutaciones truncantes en los genes MMR. En total se encontraron quince variantes de significado desconocido en los genes MMR, de las cuales 5 pudieron ser reclasificadas a patogénicas. También se encontraron 13 variantes presuntamente patogénicas por su estudio *in silico* en otros genes de predisposición a CCR. El análisis del metiloma identificó un nuevo caso de epimutación constitucional de *MLH1*. Sin embargo, no se identificaron regiones diferencialmente metiladas en los pacientes SLL al compararlos con individuos Lynch o controles sanos.

El estudio de subexoma, combinado con la evaluación de patogenicidad de variantes de significado desconocido, permitió la identificación de mutaciones deletéreas en los genes MMR, así como nuevos genes candidatos SLL. Las epimutaciones constitucionales fuera de los genes MMR no son responsables del fenotipo de deficiencia MMR observado en pacientes con SLL

Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals

Authors:

Estela Dámaso[1], Maribel González-Acosta[1,2], Gardenia Vargas-Parra[1,2], Matilde Navarro[1,2], Judith Balmaña[3], Teresa Ramon y Cajal[4], Noemí Tuset[5], Bryony A. Thompson[6], Fátima Marín[1,2], Anna Fernández[1], Carolina Gómez[1], Àngela Velasco[2,7], Ares Solanes[1], Sílvia Iglesias[1,2], Gisela Urgel[5], Consol López[4], Jesús del Valle[1,2], Olga Campos[1], Maria Santacana[8], Xavier Matias-Guiu[8,9], Conxi Lázaro[1,2], Laura Valle[1,2], Joan Brunet[1,2,7,10], Marta Pineda[1,2]*, Gabriel Capellá[1,2]*

Affiliations:

(1) Hereditary Cancer Program, Catalan Institute of Oncology, Insititut d'Investigació Biomèdica de Bellvitge (IDIBELL), ONCOBELL Program, L'Hospitalet de Llobregat, Barcelona, Spain.

(2) Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain

(3) High Risk and Cancer Prevention Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

(4) Medical Oncology Department, Hospital de Santa Creu i Sant Pau, Barcelona, Spain.

(5) Genetic Counseling Unit, Hospital Arnau de Vilanova, Lleida, Spain.

(6) Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia.

(7) Hereditary Cancer Program, Catalan Institute of Oncology, Institut d'Investigació Biomèdica de Girona (IDIBGI), Girona, Spain.

(8) Pathology Department, Hospital Arnau de Vilanova, Lleida, Spain.

(9) Pathology Department, Bellvitge University Hospital, Insititut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain.

(10) Department of Medical Sciences, School of Medicine, University of Girona, Girona, Spain

Short title:

Constitutional (epi)genetic characterization of Lynch-like syndrome

Disclosures

The authors declare no conflict of interest.

Catalan Institute of Oncology, IDIBELL
Av. Gran Via de l'Hospitalet, 199-203
08908 Hospitalet de Llobregat, Spain
Tel: (+34) 93 260 73193
E-mail: gcapella@iconcologia.net; mpineda@iconcologia.net

Correspondence:

Gabriel Capellá, M.D., Ph.D. and Marta Pineda, Ph.D.
Hereditary Cancer Program

Abstract

In ~50% of Lynch syndrome (LS)-suspected patients (also called Lynch-like syndrome, LLS), the causal mechanism for cancer predisposition remains unknown. Our aim was to elucidate the constitutional basis of mismatch repair (MMR) deficiency in LLS patients throughout a comprehensive genetic and epigenetic analysis.

One hundred and fifteen LLS patients harboring MMR deficient tumors and no pathogenic germline MMR gene mutations were included in this study. Mutational analysis of 26 colorectal cancer associated genes was performed by using a customized multigene panel and massively parallel sequencing. Pathogenicity of MMR variants was assessed by mRNA analysis and multifactorial likelihood calculations. Genome-wide methylome analysis was perfomed by using the Infinium HumanMethylation450K BeadChip.

The multigene panel analysis revealed the presence of two MMR gene truncating mutations not found in previous analysis. Of a total of 15 MMR variants of unknown significance identified, five -present in 6 unrelated individuals- were reclassified as pathogenic. In addition, 13 predicted deleterious variants in other CRC-predisposing genes (*MSH3, MUTYH, POLD1, APC, EPCAM, BUB1, FAN1, EXO1* or *PSM1*) were found in 12 probands. Methylome analysis detected one constitutional *MLH1* epimutation in an individual diagnosed with CRC at age 42, but no additional differentially methylated regions were identified in LLS compared to LS patients or healthy individuals.

In conclusion, the use of an ad-hoc designed gene panel combined with pathogenicity assessment of variants allows the identification of deleterious MMR mutations as well as new LLS candidate genes. Moreover, constitutional epimutations in non-LS-associated genes are not responsible for the MMR-deficient phenotype observed in LLS patients.

Grant support:
This work was funded by the Spanish Ministry of Economy and Competitiveness and cofunded by FEDER funds -a way to build Europe- (grants SAF2012-33636, SAF2015-68016-R and SAF2016-80888-R), CIBERONC, RTICC Network (RD12/0036/0031 and RD12/0036/0008), the Spanish Association Against Cancer (AECC) (080253), the Government of Catalonia (grant 2014SGR338 and 2017SGR1282), Fundación Mutua Madrileña (grant AP114252013). ED was supported by a grant from the Spanish Ministry of Economy and Competitiveness. The AECC fellowship to MG-A. AF was supported by a grant from the Catalonian Health Department. FM was supported by CIBERONC. The Mexican National Council for Science and Technology (CONACyT) fellowship to GV.

Acknowledgements
We are indebted to the patients and their families. We thank all members of the Hereditary Cancer Program at the Catalan Institute of Oncology. We thank Heleen van der Klift and Margaret Burton for their support with *MSH2* recurrent inversion analysis.

INTRODUCTION

Lynch syndrome (LS) is a hereditary cancer predisposition syndrome that increases the risk for colorectal and endometrial cancer as well other tumors (Lynch *et al*, 2015). It is mainly caused by pathogenic germline (epi)genetic alterations in mismatch repair (MMR) genes: *MLH1, MSH2, MSH6* and *PMS2* (Hitchins, 2015; Lynch *et al*, 2015). For tumor development, inactivation of the MMR wildtype allele is needed, leading to a MMR-deficient phenotype typically characterized by loss of expression of MMR proteins and microsatellite instability. In sporadic tumors, MLH1 loss of expression is mainly due to somatic *MLH1* methylation (Leung *et al*, 2007; Yamamoto & Imai, 2015; Young *et al*, 2005).

Nevertheless, even in the absence of somatic *MLH1* promoter methylation, no MMR germline mutations are identified as a causal mechanism in approximately 55% of patients showing MMR-deficiency in tumors, the so called Lynch-like syndrome (LLS) (Buchanan *et al*, 2014). LLS is considered a heterogeneous group showing intermediate risk of colorectal cancer (CRC) between LS and sporadic cancer (Rodríguez–Soler *et al*, 2013; Win *et al*, 2015). Thus the identification of causal mechanisms is crucial for guiding individualized surveillance strategies of LLS patients and their relatives.

Constitutional (germline) MMR cryptic mutations (usually associated to rearrangements or regulatory regions), somatic mosaicism and variants of unknown significance account for a proportion of LLS cases (Liu *et al*, 2016; Meyer *et al*, 2009; Morak *et al*, 2011; Mork *et al*, 2016; Rhees *et al*, 2014; Sourrouille *et al*, 2013; Vargas-Parra *et al*, 2017; Wagner *et al*, 2002). Furthermore, double somatic hits in MMR genes have been detected in a variable proportion (30–82%) of LLS (Geurts-Giele *et al*, 2014; Haraldsdottir *et al*, 2014; Jansen *et al*, 2016; Mensenkamp *et al*, 2014; Sourrouille *et al*, 2013; Vargas-Parra *et al*, 2017). However, even in the presence of double somatic MMR hits, an inherited predisposition to cancer -unrelated to MMR genes- cannot be totally excluded (Morak *et al*, 2017; Sourrouille *et al*, 2013). Biallelic *MUTYH* mutations, commonly associated with attenuated familial adenomatous polyposis, have been detected in 1 to 3% of LLS patients (Castillejo *et al*, 2014; Morak *et al*, 2014). Likewise, germline mutations in proofreading polymerases can lead to MMR-deficiency (Elsayed *et al*, 2014). Recently other genes are emerging as LLS candidate genes, such as *MCM9, FAN1*, *BUB1* and *SETD2* (de Voer *et al*, 2013; Goldberg *et al*, 2015; Seguí *et al*, 2015a; Vargas-Parra *et al*, 2017).

Constitutional epigenetic alterations in *MLH1* and *MSH2* genes are occasionally responsible for the MMR deficient phenotype in LS patients (Hitchins, 2015; Peltomäki, 2016). Similarly, constitutional epigenetic alterations have been rarely described in other cancer genes such as *BRCA1* and *RAD51C* in ovarian and breast cancer (Hansmann *et al*, 2012), *KILLIN* in Cowden syndrome (Bennett *et al*, 2010; Ngeow *et al*, 2011), *DAPK* in chronic lymphocytic leukemia (Raval *et al*, 2007) and *RB1* in retinoblastoma (Quiñonez-Silva *et al*, 2016). In contrast, the role of constitutional methylation in LLS has not been explored.

The aim of the current study is to elucidate the constitutional basis of MMR deficiency in a cohort of 115 LLS cases throughout a comprehensive genetic and epigenetic characterization. The obtained results contribute to the understanding of LLS by ruling out the presence of constitutional methylation

events as a common cause for LLS as well as highlighting the relevance of performing comprehensive genetic analyses in these patients.

METHODS

Patients

A total of 115 Caucasian Lynch-like syndrome patients harboring MMR deficient tumors MMR loss of expression and/or microsatellite instability (MSI) were included (Table S1). Twenty-three of them were reported in a previous publication (Vargas-Parra et al., 2017). The immunohistochemistry (IHC) pattern of MMR protein expression was as follows: 57 MLH1/PMS2 loss, 27 MSH2/MSH6 loss, 12 MSH6 loss, 5 PMS2 loss and 14 MMR conserved expression but MSI. In the 57 tumors showing loss of MLH1/PMS2 protein expression the presence of somatic *MLH1* promoter hypermethylation and/or *BRAF* V600E were excluded, except for 3 cases (7, 9 and 78) that had wildtype *BRAF* and non-informative tumor *MLH1* promoter methylation results.

Based on the IHC MMR expression pattern, the corresponding MMR genes were sequenced. Cases in whom no pathogenic variants in MMR genes had been identified were included in this study (Table S1). Of note, nine patients initially classified as LLS were excluded from this cohort due to the previous identification of germline biallelic *MUTYH* and *MSH2* pathogenic mutations (Castillejo *et al*, 2014; Seguí *et al*, 2015b; Vargas-Parra *et al*, 2017). Concerning clinical criteria fulfillment, 83 patients met Revised Bethesda guidelines (72,2%) and 11 the Amsterdam criteria (9,6%) for hereditary nonpolyposis CRC (Table S1). The remaining 21 (5.4%) were referred to the Genetic Counseling Unit because of histological features suggestive of MMR-deficiency and loss of MMR protein expression.

In addition to LLS patients, 61 LS cases harboring MMR genetic mutations, 12 constitutional *MLH1* epimutation carriers and 41 healthy controls were included as controls for genome-wide methylome analysis (Dámaso *et al*, 2018) (Table S2).

All patients were assessed at the Cancer Genetic Counseling Units of the Catalan Institute of Oncology, Santa Creu i Sant Pau, Arnau de Vilanova and Vall d'Hebron hospitals from 1998 to 2012. Informed consent was obtained from all individuals enrolled and internal Ethics Committees of participant hospitals approved this study.

Samples

Isolation of genomic DNA from blood of all included patients was performed using FlexiGene DNA kit (Qiagen, Hilden, Germany) or Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. FFPE blocks of normal colorectal mucosa and CRC tissue were obtained when available. For each FFPE specimen, 10-20 x 10-µm sections were cut from a single representative block per case, using macrodissection with a scalpel if needed to enrich for tumor cells. After deparaffinization using the Qiagen Deparaffinization Solution (Qiagen, Hilden, Germany), DNA

was isolated using the QIAmp DNA FFPE Tissue Kit (Qiagen) according to manufacturer's instructions. DNA quality was tested using NanoDrop ND 1000 Spectrophotometer (Thermo Fischer Scientific), electrophoresis in agarose gel and Qubit Fluorometer using dsDNA BR Assay (Invitrogen, Carlsbad, CA, USA).

Mismatch repair genes mutational analysis

Mutational analysis of coding regions of MMR genes

According to the IHC pattern in tumors, mutation analysis of candidate MMR genes (*MLH1* NM_000249.3, NG_007109.2; *MSH2*, NM_000251.2, NG_007110.1; *MSH6*, NM_000179.2, NG_007111.1; *PMS2* NM_000535.6, NG_008466.1) was initially performed on blood DNA by PCR amplification of exonic regions and exon–intron boundaries, followed by Sanger sequencing. Primers and conditions are available upon request. Genomic rearrangements in the MMR genes were analyzed by multiplex ligation dependent probe amplification (MLPA) using SALSA-*MLH1*/*MSH2* P003-B1, SALSA-*MLH1*/*MSH2* P248-B1, *MSH6* P072 and/or PMS2 P008-C1 kits (MRC-Holland), according to manufacturer's indications. Screening of gross rearrangements in MSH2-deficient cases was complemented by using the 2 available MLPA kits for *MSH2* gene analysis and by screening the recurrent *MSH2* inversion in exons 1-7 (Wagner *et al*, 2002). Annotation of variants was done following the Human Genome Variation Society recommendations.

Direct sequencing of MMR promoter regions and 3'UTR of the *EPCAM* gene

The regions encompassing 662 bases upstream of the transcriptional start site (TSS) of *MSH2*, 915bp of *MSH6* TSS, 1469bp of *MLH1* TSS and 429bp of the *EPCAM* 3'UTR were amplified by PCR using Megamix-Double (Microzone Ltd., UK) and sequenced using the BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems, CA, USA) (Table S3; conditions available upon request). Sequences were analyzed on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems).

Targeted next generation sequencing

Sixty-two LLS patients with strong individual and/or familial cancer history (Amsterdam or Bethesda 1, 2, 4 or 5 criteria) were analyzed using a NGS custom panel of 26 CRC associated genes, previously used for the characterization of MSH2/MSH6–deficient cases (Vargas-Parra *et al*, 2017). Agilent SureDesign web-based application (Agilent Technologies, USA) was used to design DNA capture probes of 509 target regions, including the coding exons plus 10 flanking bases of 26 genes associated to CRC, as well as their promoter regions (comprising 650 bases upstream their TSS), as previously reported (Vargas-Parra *et al*, 2017). Agilent SureCall application was used to trim, align and call variants. Variant filtering was performed based on Phred quality ≥30, alternative allele ratio ≥0.05, read depth ≥38x in PBL samples. Identified variants were then filtered against common single-nucleotide polymorphisms (MAF>1 according to ExAC and ESP databases) as well as class 1 and class 2 MMR variants according to InSight database. Predicted pathogenic germline rare variants were

further confirmed by Sanger sequencing using independent DNA samples. Primers and conditions are detailed in Table S3.

Pathogenicity assessment of genetic variants

Variant frequency and cosegregation analysis. Global population frequency of the identified variants was retrieved from the Exome Aggregation Consortium (ExAC; http://exac.broadinstitute.org/) and NHLBI Exome Sequencing Project (ESP; http://evs.gs.washington.edu/EVS) databases. Identified variants were also screened in DNA samples from family relatives by Sanger sequencing when available.

In silico prediction of the functional impact. Alamut Visual v2.9.0 software (Interactive Biosoftware, Rouen, France) was used for i*n silico* predictions. The potential effects of variants on splicing were evaluated by using SSF, MaxEnt, NN SPLICE and Gene Splicer. At the protein level the impact of variants was analyzed using the *in silico* algorithms PolyPhen-2, SIFT, Align GVGD and Mutation taster. Also, PROVEAN was used for in-frame indel variants. PROMO 3.0 software (Farré *et al*, 2003; Messeguer *et al*, 2002) was used to predict any changes in transcription factor binding between wildtype alleles and promoter variants. Only human transcription factors were considered and 5% was selected as maximum matrix dissimilarity rate.

Multifactorial likelihood analysis. For MMR variants, posterior probability of pathogenicity was calculated by multifactorial likelihood analysis as previously described (Thompson, 2014) based on estimated prior probabilities of pathogenicity and likelihood ratios (LR) for segregation and tumor characteristics. Variants were classified according to the 5 class IARC scheme (Plon *et al*, 2011) based on the calculated posterior probability.

mRNA splicing analysis and allele specific expression analysis. Available lymphocytes from variant carriers were cultured with and without puromycin after one week of culture with PB-MAX medium. Total RNA was extracted using Trizol Reagent. One microgram of RNA was retrotranscribed using iScript cDNA synthesis kit (Bio-Rad, USA). cDNA amplification of exon containing the variants and at least two exons up and downstream the main one was performed using specific primers provided in Table S3. Sequencing was performed using the BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems). Mutation Surveyor (SoftGenetics) was used for sequence visualization.

For allelic expression analyses, regions containing heterozygous variants were selected. The relative levels of both alleles were determined in genomic DNA and cDNA by single-nucleotide primer extension (SNuPE) as previously described (Pineda *et al*, 2012) (primers provided in Table S3). Allele-specific expression (ASE) was calculated by dividing the ratio of variant/wildtype allele in cDNA by the ratio of variant/wildtype allele in gDNA. Experiments were performed in quadruplicate. ASE values of 1.0 indicate equal levels of expression from both alleles. ASE values lower than 1.0 indicate reduced expression from one allele.

Tumor analysis

Whole exome sequencing of FFPE DNA extracted from the tumor of patient 53 -carrier of a germline variant in the exonuclease domain of POLE- and of his blood, was carried out in a Hi-Seq2000 (Illumina) with a coverage >100x, after library preparation using the Agilent Sure Select Human All Exon v5 kit. Sequence alignment was carried out with BWA (Li & Durbin, 2009) and variant calling with MuTect (Cibulskis *et al*, 2013). Variants identified in the patient's blood DNA were eliminated for the analysis of somatic mutations in the tumor. Variants present in at least 10% of the reads were considered for subsequent analyses. The contribution of the COSMIC mutational signatures (Alexandrov *et al*, 2018; Alexandrov & Stratton, 2014; Nik-Zainal *et al*, 2016) to the tumor was calculated with *deconstructSigs* (Rosenthal 2016).

MSH3 expression and elevated microsatellite instability at selected tetranucleotide repeats (EMAST) were evaluated in the normal and tumor samples from case 74, harboring two *MSH3* variants. Immunohistochemistry of MSH3 protein was performed using anti-MSH3 antibody at dilution 1:150 (Novus Biologicals, USA). The reaction was visualized with the EnVision™ FLEX Detection Kit (Agilent Technologies-DAKO, Santa Clara, United States) following standard protocols. For EMAST analysis, six previously reported tetranucleotide repeat markers were analyzed (Adam *et al*, 2016; Arai *et al*, 2013; Burger *et al*, 2006; Carethers *et al*, 2015; Stoehr *et al*, 2012). Primers and conditions are listed in Table S3. The amplification products were run on an ABI Prism 3130 DNA sequencer and analyzed using GeneMapper v4.0 (Applied Biosystems). EMAST was considered when two or more of the analyzed markers displayed instability.

Genome-wide methylation profiling

Blood DNA samples from LLS patients and controls, as well as available FFPE colorectal normal/tumor DNA, were included in the genome wide methylation profiling analysis using Infinium Human Methylation 450K Beadchip (Table S2), also including the LLS cases previously reported (Vargas-Parra *et al*, 2017). Array data processing and data analysis were performed as previously described (Dámaso *et al*, 2018). Blood DNA with an A260/A280 ratio between 1.7-2.0 were considered suitable for hybridization. DNAs from FFPE samples were analyzed by qPCR using Infinium FFPE QC (Illumina) in order to determine their suitability for FFPE restoration. All samples showing ΔCt values lower than 5 were restored using the Infinium HD FFPE Restore kit (Illumina), following the manufacturer's instructions. A total of 1000 ng blood DNA and 500 ng FFPE DNA were bisulfite converted using the EZ DNA Methylation™ Kit (Zymo Research), according to the manufacturer's instructions. To determine the efficiency of the bisulfite conversion, a predetermined genomic region was evaluated by Sanger sequencing in one methylated and one unmethylated control of each bisulfite conversion batchs. Genome wide methylation profiling was performed using the Infinium HumanMethylation 450K Beadchip (Illumina), which interrogates the methylation status of 485.764 CpG sites across the genome. For internal quality control, in vitro methylated and unmethylated DNAs were included in each batch. After hybridization, sample scanning was performed using the HiScan platform (Illumina), which has a laser scanner with two colours (532nm/660nm). The relative intensity of each dye was

analyzed using the GenomeStudio software (Methylation Module). For each analyzed CpG site, a β-value was obtained depending on the florescence intensity. B measures took values between 0 (unmethylated) and 1 (fully methylated). The analysis of batch effects was performed using RnBeads software (Max-Planck-Institut Informatik). Group comparisons and statistical analysis -based on differentially methylated CpG sites, CpG islands, promoters, genes and tiling- were performed using RnBeads software (Max-Planck-Institut Informatik). CpG methylation was visualized using the Integrative Genome Viewer (Broad Institute). GRCh37/hg19 was used as the reference genome (date of release: February 2009). Only positions that reached an FDR p-value<0.05 when comparisons are done between groups > 10 samples were considered.

RESULTS

Reassessment of germline genetic variants in the MMR genes

The presence of missed MMR genetic alterations was reassessed in blood samples from 42 LLS patients with strong individual and/or familial cancer history by means of a NGS custom panel of CRC-associated genes, previously used in the analysis of 23 MSH2-deficient LLS cases from the same series (Vargas-Parra *et al*, 2017) (Table 1). By using this approach two *bona fide* previously not identified germline pathogenic MMR variants were found in cases fulfilling Amsterdam criteria. Case 33 was a male who suffered from two CRC at age of 40 and 46. Immunohistochemical (IHC) staining displayed loss of MLH1 protein expression in his first tumor, being non-informative the second one. Previous Single Strand Conformation Polymorphism (SSCP) analysis was negative whereas NGS analysis identified a pathogenic *MLH1* mutation, c.676C>T (p.Arg226*) (Figure 1A). Case 92 was a woman who developed endometrial cancer at age 48. Her tumor displayed MSI with conserved MMR protein expression. No mutation was identified by Sanger sequencing in either *MLH1* or *MSH2*. The panel allowed the identification of a truncating mutation in *MSH6,* c.2219T>A (p.Leu740*) (Figure 1B). In addition to the 9 MMR variants of unknown significance identified in 10 LSS individuals in previous analyses, 4 additional variants (*MSH6* c.2092C>G, *MSH6* c.3150_3161dup, *PMS2* c.1320A>G and MSH2 c.2802G>A) were detected in 4 additional cases. (Table 1).

This re-analysis was complemented with the sequencing of the promoter regions of the four MMR genes, which identified an *MLH1* promoter variant (c.-574T>C, rs558088820, MAF <0.0001) in case 13 (Table 1). This variant was predicted to interfere with YY1 transcription factor binding, which directs histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress its activity (Gordon *et al*, 2005).

Regarding rearrangements, the presence of the germline recurrent inversion of exons 1–7 in MSH2-deficient cases (Mork *et al*, 2016; Rhees *et al*, 2014; Wagner *et al*, 2002) was evaluated with negative results (Table 1). In contrast, MLPA reanalysis using the P248 kit (MRC-Holland) revealed the presence of an *MSH2* exon 8 duplication in case 57.

Table 1: Results obtained in the characterization of LLS patients (including cases analyzed by NGS, VUS MMR carriers and epimutations). (*) Not previously reported MMR variant classified according to Insight variant classification rules.

Case ID	Results from previous MMR mutational analysis by Sanger sequencing / SSCP	Results from the analysis of CRC-associated genes obtained in this study		Pathogenicity assessment of VUS		Final case classification
	VUS (Insight classification)	Variants in LS-associated genes (Insight classification)	Predicted pathogenic variants in other predisposing genes (ClinVar classification)	VUS assessment	Final classification	
A. Results obtained in the analysis of 42 samples by using a NGS subexome panel of CRC-associated genes. Only the MMR variants and the predicted pathogenic variants in other genes were shown (see Table S7).						
5	-	*MSH6* c.2092C>G, p.Gln698Glu (Class 3)	-			LLS (MMR VUS carrier)
6	-	-	-			LLS
7	-	***MLH1* epimutation**	-	Confirmation by MS-MLPA (48%)		**LS (*MLH1* epimutation carrier)**
8	-	-	-			LLS
10	-	-	*EPCAM* c.811G>T, p.Val271Phe (not reported)			LLS (VUS carrier)
13	-	*MLH1* c.-574T>C, p.? (Class 3*)	-			LLS (MMR VUS carrier)
28	-	-	-			LLS
29	-	-	*POLD1* c.2275G>A, p.Val759Ile (Class 1,2,3)			LLS (VUS carrier)
30	-	-	*APC* c.7936C>G, p.Gln2646Glu (Class 3)			LLS (VUS carrier)
33	-	***MLH1* c.676C>T, p.Arg226* (Class 5)**	-			**LS**
39	*MSH2* c.1787A>G; p.Asn596Ser (Class 3)	*MSH2* c.1787A>G; p.Asn596Ser (Class 3)	*FAN1* c.149T>G, p.Met50Arg (not reported)			LLS (VUS carrier)
42	-	-	-			LLS
44	-	-	-			LLS (VUS carrier)
45	-	-	-			LLS
48	-	-	-			LLS
53	-	-	-			LLS
55	-	-	*PMS1* c.497A>C, p.Lys166Thr (not reported)			LLS (VUS carrier)
56	-	-	-			LLS
57	-	*MSH2* E8 duplication (Class 3*)	-	Aberrant splicing	***MSH2* E8 duplication (Class 5)**	**LS**
58	*MSH2* c.2045C>G; p.Thr682Ser (Class 3*)	*MSH2* c.2045C>G; p.Thr682Ser (Class 3*)	*EXO1* c.2212-1G>A (Class 3)			LLS (MMR VUS carrier)
59	-	-	*APC* c.1966C>G, p.Leu656Val (Class 3)			LLS (VUS carrier)
61	-	-	-			LLS
62	-	-	*MSH3* c.2732T>G, p.Leu911Trp (not reported)			LLS (VUS carrier)
63	*MSH2* c.2702A>T; p.Glu901Val (Class 3*)	*MSH2* c.2702A>T; p.Glu901Val (Class 3*)	-			LLS (MMR VUS carrier)
64	-	-	-			LLS
65	-	-	*MUTYH* c.1437_1439delGGA, p.Glu480del (Class 5)			LLS (monoallelic MUTYH carrier)
66	-	-	-			LLS
74	-	-	*MSH3* c.685T>C, p.Tyr229His (not reported); *MSH3* c.2732T>G, p.Leu911Trp (not reported)	MSH3 conserved expression / EMAST/ in cis	*MSH3* c.685T>C, p.Tyr229His (VUS); MSH3 c.2732T>G, p.Leu911Trp (VUS)	LLS (VUS carrier)
76	-	-	-			LLS
78	-	-	-			LLS
79	-	-	-			LLS
81	-	-	*BUB1* c.2473C>T, p.Pro825Ser (not reported)			LLS (VUS carrier)
82	-	*MSH6* c.3150_3161dup , p.(Val1051_Ile1054dup) (Class 3*)	-	Multifactorial (>0.99)/normal splicing	***MSH6* c.3150_3161dup , p.(Val1051_Ile1054dup) (Class 5)**	**LS**
85	-	*PMS2* c.1320A>G, p.Pro440= (Class 3*)	*MSH3* c.3072G>C, p.Gln1024His (not reported)			LLS (MMR VUS carrier)
87	-	-	-			LLS
92	-	***MSH6* c.2219T>A, p.Leu740* (Class 5)**	-			**LS**
93	-	-	-			LLS
94	-	-	-			LLS
95	-	-	-			LLS
96	-	-	*APC* c.7514G>A, p.Arg2505Gln (Class 1,2)			LLS (VUS carrier)
97	-	-	-			LLS
98	-	*MSH2* c.2802G>A, p.Thr934Thr (Class 3)	-			LLS (MMR VUS carrier)
B. Results obtained in 7 additional cases harboring MMR variants identified by previous Sanger sequencing						
35	*MLH1* c.25C>T, p.Arg9Trp, (Class 3) *APC* c.1958+3A>G (Class 5) (Borrás et al 2012)	-	-			**FAP (MMR VUS carrier)**
67	*MSH6* c.1153_1155delAGG p.Arg385del (Class 3 *)	-		Multifactorial (>0.99)/normal splicing	***MSH6* c.1153_1155delAGG p.Arg385del (Class 5*)**	**LS**
70	*MSH6* c.1618_1620delCTT; p.Leu540del (Class 3 *)	-	-	Multifactorial (>0.99)/aberrant splicing	***MSH6* c.1618_1620delCTT; p.Leu540del (Class 5*)**	**LS**
72	*MSH6* c.1450G>A; p.Glu487Lys (Class 3 *)	-	-			LLS (MMR VUS carrier)
73	*MSH6* c.3296T>A; p.Ile1099Asn (Class 3 *)	-	-			LLS (MMR VUS carrier)
75	*MSH6* c.1618_1620delCTT; p.Leu540del (Class 3 *)	-	-	Multifactorial (>0.99)/aberrant splicing	***MSH6* c.1618_1620delCTT; p.Leu540del (Class 5*)**	**LS**
77	*MSH6* c.3226C>T, p.Arg1076Cys (Class 3)	-	-	Insight variant classification revision	***MSH6* c.3226C>T, p.Arg1076Cys (Class 4, Insight March 2018)**	**LS**

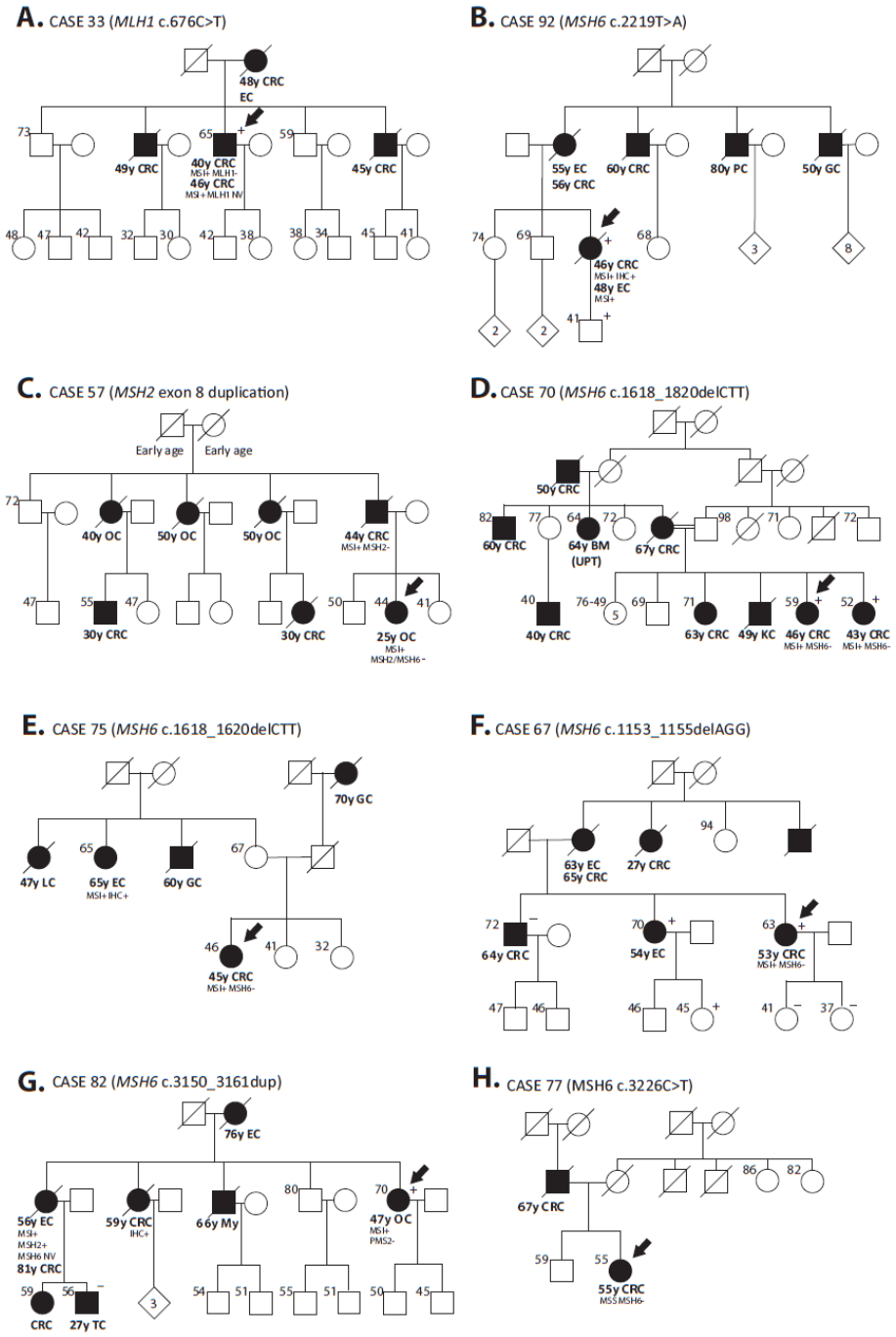


Figure 1: Pedigrees from patients reclassified as Lynch syndrome during the current study. Abbreviations: CRC=colorectal cancer, EC=endometrial cancer, PC=prostate cancer, GC=gastric cancer, OC=ovarian cancer, BM(UTP)=brain metastasis from unknown primary tumor, KC=kidney cancer, TC=testis cancer, My=myeloma, MSI+=microsatellite instability, NV=No valuable, +=mutation carrier, -=non carrier.

Pathogenicity assessment of MMR variants

In all, 15 MMR VUS were identified in 16 probands (Table 2A): 7 in *MSH6,* 5 in *MSH2*, 2 in *MLH1* and 1 in *PMS2*. mRNA splicing evaluation and stability analyses were possible for the *MSH6* variants c.1153_1155del (p.Arg385del), c.1618_1620del (p.Leu540del) and c.3150_3161dup (p.Val1051_Ile1054dup). An aberrant transcript was identified in the c.1618_1620del carriers (cases 70 and 75) corresponding to a partial out-of-frame deletion of exon 4 (r.1607_3172del, p.Ser536_Asp1058delinsAsn), that coexisted with the full-length transcript (r.1618_1620del, p.Leu540del) (Figure S1). This is in agreement with a partial allelic imbalance detected at the c.1618 position (Table 2A). The remaining 2 variants analyzed had no apparent effect on mRNA splicing and stability (Table 2A). Clinico-pathological data from the same families were used in multifactorial likelihood analyses. Since *MSH6* c.1153_1155del and c.3150_3161dup variants had been identified in families from other centers (AF1-3; Figure S2), they were also included in the multifactorial calculations. For the three *MSH6* variants, posterior probability of pathogenicity resulted >0.999, classifying them as pathogenic (Table 2B). In addition, *MSH6* c.3226C>T (p.Arg1076Cys) variant, initially classified as VUS, was reclassified as probably pathogenic (class 4) because of its co-occurrence *in trans* with *MSH6* pathogenic mutations in patients with constitutional MMR deficiency and loss of MSH6 expression in normal cells (Gardès *et al*, 2012; Jasperson *et al*, 2011; Okkels *et al*, 2006; Plaschke *et al*, 2006; Rahner *et al*, 2008)

No effect on splicing and transcript stability was detected in lymphocytes from the carrier of *MSH2* c.1787A>G (p.Asn596Ser) variant, as previously reported (Betz *et al*, 2009) (Table 2A). In case 57, splicing analysis confirmed the presence of an aberrant transcript containing the exon 8 duplication (r.1277_1387dup), predicted to generate a frameshift protein (p.Val463Glufs*11), allowing to classify the variant as pathogenic (Figure 1C and S3).

The functional impact of *MLH1* promoter c.-574T>C variant on *MLH1* transcription could not be assessed due to the absence of coding heterozygous *MLH1* variants, and it was classified as VUS. This *MLH1* variant together with the other 9 identified in MMR genes remained as VUS due to insufficient evidence, although *in silico* predictions suggested neutrality for 4 of them (*MSH2* c.1787A>G, c.2045G>C and c.2802G>A and *PMS2* c.1320A>G) (Table 2 and S4).

Table 2. Results of the pathogenicity assessment of MMR variants of unknown significance (VUS)

Table 2A. Results of in silico predictions and pathogenicity assessment of MMR VUS (see Table S7 and S8)

Case ID	MMR gene	*MMR* variant	Predicted protein	Protein functional domain	Insight Classification (2015)	ClinVar Classification	Frequency in controls (ExAC/ESP)	rs ID	In silico predictions		RNA analyses		Multifactorial calculations	Final classification
									Splicing	Protein function	cDNA splicing analysis	cDNA stability analysis (+/- puromicin)		
13	*MLH1*	c.-574T>C	p.?	_	Class 3	Not reported	0,000084/NR	rs558088820	NA	NA	NP	NP	NP	Class 3
35		c.25C>T	p.(Arg9Trp)	MutSα Interaction	Class 3	VUS (2) / +++	NR/NR	rs587779000	No changes	Damaging	r.25C>T^; p.Arg9Trp	NP	NP	Class 3
39	*MSH2*	c.1787A>G	p.(Asn596Ser)	Lever domain & MSH3/MSH6 interaction & EXO1 stabilisation and interaction	Class 3	VUS (3) vs Bening/Likely bening (3) / +++	NR/0.0002	rs41295288	No changes	Benign	r.1787A>G; p.Asn596Ser	Non allelic imbalance (Sanger seq)	NP	Class 3
57		exon 8 duplication	p.?	Helix-turn-helix & MSH3/MSH6 interaction	Not reported	Not reported	_	_	NA	NA	r.1277_1387dup; p.Val463Glufs*11	NP	NP	**Class 5**
58		c.2045C>G	p.(Thr682Ser)	ATPase domain & MutLα Interaction	Not reported	Not reported	NR/NR	_	No changes	Benign	NA	NA	NP	Class 3
63		c.2702A>T	p.(Glu901Val)	Helix-turn-helix & MSH3/MSH6 interaction	Not reported	Not reported	NR/NR	_	No changes	Damaging	NA	NA	NP	Class 3
98		c.2802G>A	p.(Thr934=)	Helix-turn-helix & MSH3/MSH6 interaction	Class 3	VUS (2) vs Bening/Likely Bening (5) / +++	0.000/0.0001	rs150259097	No changes	NA	NP	NP	NP	Class 3
5	*MSH6*	c.2092C>G	p.(Gln698Glu)	Connector domain	Class 3	VUS (5) / +++	NR/NR	rs63750832	Unconclusive (3/5)	Benign	r.2092C>G¨; p.Gln698Glu	NP	NP	Class 3
67		c.1153_1155delAGG	p.(Arg385del)	DNA binding & MSH2 interaction	Class 3	VUS (2) / +++	NR/NR	rs267608043	No changes	Damaging	r.1153_1155delAGG (NP); p.Arg385del	Non allelic imbalance (NP / 1.02±0.09)	>0,99	**Class 5**
72		c.1450G>A	p.(Glu484Lys)	Connector domain & MSH2 interaction	Not reported	VUS (1) / +	NR/NR	_	No changes	Damaging	NP	NP	NP	Class 3
70 & 75		c.1618_1620delCTT	p.(Leu540del)	Connector domain & MSH2 interaction	Not reported	VUS (2) vs Pathogenic (1) / +	NR/NR	_	No changes	Damaging	r.[1618_1620del;1607_3172del]; p.[Leu540del;Ser536_Asp1058delinsAsn]	Destabilization (0.69±0.03 / 0.65±0.06)	>0,99	**Class 5**
82		c.3150_3161dup	p.(Val1051_Ile1054dup)	Lever domain	Not reported	Not reported	NR/NR	_	No changes	Damaging	r.3150_3161dup; p.Val1051_Ile1054dup	Non allelic imbalance (1.04±0.14 / 1.16±0.26)	>0,99	**Class 5**
77		c.3226C>T	p.(Arg1076Cys)	Lever domain	Class 3	Pathogenic/Likely pathogenic (6) / +++	NR/NR	rs63750617	No changes	Damaging	r.3226C>T*; p.Arg1076Cys	NP	NP	**Class 4****
73		c.3296T>A	p.(Ile1099Asn)	Lever domain	Not reported	Not reported	NR/NR	_	No changes	Damaging	NP	NP	NP	Class 3
85	*PMS2*	c.1320A>G	p.(Pro440=)	_	Not reported	VUS (1) vs Bening/Likely bening (5) / +	NR/0.0001	rs138697590	No changes	NA	NP	NP	NP	Class 3

^Borràs et al., Hum Mut 2012; *Thompson et al., 2013; ¨Wang et al., 1999; **Insight classification, March 2018
NA: Not available; NP: Not performed

Table 2B. Results of multifactorial likelihood analyses of MMR VUS

MSH6 variant	Frequency in controls (ExAC / ESP)	Initial classification (March 2018)	Multifactorial likelihood analysis													Final classification
			Prior probability of pathogenicity	Prior used	Case ID	Ascertainment	Cancer (age)	MSI/ IHC status	MSI CRC LR	Tumor Characteristics LR	Bayes	Segregation LR	Odds for causality	Posterior Odds	Posterior probability of pathogenicity	
c.1153_1155delAGG; p.Arg385del	NR /NR	Class 3	0,133556728	0,5	67	clinic	CRC (53)	MSI-H & MSH6 loss	6,96	6,96	2,1493	15,21811865	105,918106	105,918106	0,990647047	**Class 5 Pathogenic**
					AF1	clinic	EC (59)	MSH6 loss			7,0805					
					AF1	clinic	CRC (59)	MSH6 loss								
c.1618_1620delCTT; p.[Leu540del;Ser536_Asp1058delinsAsn]	NR /NR	Not reported	0,95943958	0,9	70	clinic	CRC (46)	MSI-H & MSH6 loss		6,96	1,84534682	1,84534682	12,8436139	115,592525	0,991423121	**Class 5 Pathogenic**
					70	clinic	CRC (43)	MSI-H & MSH6 loss	6,96							
					75	clinic	CRC (45)	MSI-H & MSH6 loss								
c.3150_3161dup p.Val1051_Ile1054dup	NR /NR	Not reported	0,9608	0,9	82	clinic	OC (47)	MSI-H & PMS2 loss		649,461896	0,9887	28,74734233	18670,3035	168032,731	0,999994049	**Class 5 Pathogenic**
					AF2	clinic	CRC (61)	MSI-H & MSH6 loss	8,66		29,0759					
					AF3	clinic	CRC (47)	MSH6 loss								
					AF3	clinic	EC (56)	MSH6 loss								
					AF3		CRC (75)	MSI-H & MSH6 loss	8,66							
					AF3	clinic	CRC (68)	MSI-H	8,66							

Abbreviations: LR, likelihood ratio; NR, not reported; NE, not evaluable; CRC, colorectal cancer; EC, endometrial cancer; MSI-H, microsatellite instability high; MSS, microsatellite stable .

Identification of variants in other CRC-predisposing genes

The multigene panel analysis allowed the identification of rare germline variants in other CRC-predisposing genes in 32 LLS cases (32/42, 76.2%) (Table S5). Thirteen of them were variants predicted as pathogenic by *in silico* tools, identified in well-known CRC predisposing genes such as *APC* and *MUTYH*, as well as variants in newly emerging cancer predisposing genes such as *MSH3* and *FAN1* (Table 3 and S6). Among them, four variants were identified in the *MSH3* gene (Table 3), two of them coexisting in *cis* in the same patient (case 74; Figure S4). One of these two variants, c.2732T>G (p.Leu911Trp) affects a highly conserved residue along MutS proteins, and the other one, c.685T>C (p.Tyr229His), is located next to the DNA recognition domain of the protein and affects a highly conserved residue (Adzhubei *et al*, 2010; Obmolova *et al*, 2000). While immunohistochemical staining showed conserved MSH3 nuclear expression in normal and tumor tissue from case 74, tetranucleotide repeats analysis displayed instability in 2 out of 6 microsatellites, indicating EMAST (Figure S4).

The *FAN1* c.149T>G (p.Met50Arg) variant was found in heterozygosity in case 39, diagnosed with CRC at 49 years of age. This variant, localized at the ubiquitin-binding domain, was previously associated to pancreatic cancer predisposition (Smith *et al*, 2016). Functional assays demonstrated that c.149T>G variant affects *FAN1* nuclease activity, impeding the repair of chromosome abnormalities when forks stall after hydroxyurea and mitomycin treatment (Lachaud *et al*, 2016). Conversely, homozygous carries of this *FAN1* variant have been reported in the Genome Aggregation Database (GnomAD).

POLE c.898A>G (p.Ile300Val) variant, located in the region coding the exonuclease domain of the polymerase, was identified in patient 53, diagnosed with CRC at age 51 and two synchronous CRC at age 81. Tumor WES revealed a major contribution of COSMIC mutational signature 6 (56.2%), associated with MMR deficiency, and complete absence of the POLE-associated COSMIC mutational signature 10, or signature 14, identified in tumors with concurrent POLE mutation and MMR deficiency (Alexandrov *et al*, 2015) (Figure S5). The evidence gathered indicates lack of causal association of the *POLE* c.898A>G with the patient's CRC, and supports a benign nature of the variant, as suggested by the *in silico* tools.

EXO1 c.2212-1G>A was identified in case 58, diagnosed with CRC at age 58 and 61. The splice-site variant causes an in-frame deletion of 6 amino acids in the MSH2 interaction domain (Table 3). The absence of family history prevented cosegregation analysis.

No rare (population MAF<0.01) germline variants were identified in *BUB1B, CHEK2, PTEN, STK11* or *TP53* genes (Table S5).

Table 3. Variants identified in non LS-associated genes and predicted pathogenic by *in silico* predictions (see Table S7 and S9)

Case ID	Variant calling			rs ID	MAF	In silico predictions						ClinVar Classification
						Splicing	Protein function					
	Gene	cDNA change	Predicted protein change		ExAC/ESP		SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	
10	*EPCAM*	c.811G>T	p.(Val271Phe)		NR/NR	No changes	D (0)	D (1)	PrD (1.000)	PrD (0.989)	NP	Not reported
29	*POLD1*	c.2275G>A	p.(Val759Ile)	rs145473716	0.002/0.001	No changes	D (0)	D (1)	PrD (1.000)	PrD (0.988)	NP	VUS (1) vs Benign/Likely benign (6) /+
30	*APC*	c.7936C>G	p.(Gln2646Glu)		NR/NR	No changes	D (0.02)	D (1)	PsD (0.688)	B (0.182)	NP	VUS (1)/+
39	*FAN1*	c.149T>G	p.(Met50Arg)	rs148404807	0.002/0.002	No changes	T (0.08)	D (1)	PrD (0.991)	PsD (0.690)	NP	Not reported
55	*PMS1*	c.497A>C	p.(Lys166Thr)		NR/NR	No changes	D (0)	D (1)	PsD (0.757)	PsD (0.599)	NP	Not reported
58	*EXO1*	c.2212-1G>A	p.Val738_Lys743del	rs4150000	0.0019/0.0028	Loss of ASS	NA	NA	NA	NA	NA	Lhotaa et al., 2016: r.2212_2229del; p.Val738_Lys743del
59	*APC*	c.1966C>G	p.(Leu656Val)	rs577466163	NR/NR	Gain of DSS	D (0)	D (1)	PrD(0.999)	PrD (0.998)	NP	VUS (1)/+
62 and 74	*MSH3*	c.2732T>G	p.(Leu911Trp)	rs41545019	0.002/0.004	No changes	D (0)	D (0.999)	PrD (1.000)	PrD (0.978)	NP	Not reported
65	*MUTYH*	c.1437_1439delGGA	p.Glu480del	rs587778541	NR/0.000	No changes	NA	NA	NA	NA	D (-7.78)	Pathogenic (9)/**
74	*MSH3*	c.685T>C	p.(Tyr229His)		NR/NR	No changes	D (0.01)	D (0.999)	PrD (1.000)	PrD (0.973)	NP	Not reported
81	*BUB1*	c.2473C>T	p.(Pro825Ser)	rs748392521	NR/NR	No changes	D (0)	D(1)	PrD (1.000)	PrD (0.997)	NP	Not reported
85	*MSH3*	c.3072G>C	p.(Gln1024His)	rs147640909	0.000/0.000	Loss of DSS / Inconclusive at ASS	T (0.39)	P (0.996)	B (0.007)	B (0.013)	NP	Not reported
96	*APC*	c.7514G>A	p.(Arg2505Gln)	rs147549623	0.001/0.001	No changes	D (0.04)	D (1)	PrD (1.000)	PrD (0.961)	NP	Benign/Likely benign (8)/++

Constitutional epigenetic alterations in MMR genes

Methylome analysis was firstly used to evaluate the existence of constitutional epigenetic alterations in the MMR genes. Blood DNA from case 7 displayed *MLH1* promoter hypermethylation that was further validated in blood using MS-MLPA (mean methylation in the *MLH1* C/D regions 48%; data not shown) (Deng *et al*, 1999). The *MLH1* epimutation carrier developed a *BRAF* wildtype CRC at age 42 (Figure 2A). Blood methylation pattern matched in extension with the 1.6 Kb differentially methylated region (DMR) previously described in constitutional epimutation carriers (Dámaso *et al*, 2018) (Figure 2B). The constitutional epimutation was also detected in normal colorectal mucosa of the carrier (Figure 2C). No other cases with MMR promoter hypermethylation were found.

Global epigenetic characterization of Lynch-like cases

Constitutional genome-wide epigenetic characterization of LLS cases was carried out with the aim of assessing the contribution of constitutional epimutations in other non-LS genes to LLS. No differentially methylated (DM) CpG islands were evidenced when LLS blood samples were compared to LS or healthy individuals (Table S7A). As expected the *EPM2AIP1-MLH1* CpG island was the only DM region identified in blood when the LLS group was compared to *MLH1* constitutional epimutations (Table S7A). The subsequent analysis of individual CpG sites identified a number of DM sites in the genome (Table S7B). Among them, only a single CpG located within *KHDC1* gene showed methylation differences higher than 20% in *MLH1*-deficient LLS cases in comparison to constitutional *MLH1* epimutations. However, this CpG site, located in a boundary between a non-methylated and a fully methylated region, evidenced high dispersion within groups (Figure S6). No constitutional epigenetic aberrations were evidenced in the LLS group when methylome data was reanalyzed after excluding LS variant carriers and carriers of predicted pathogenic variants in CCR predisposing genes.

Next, we investigated the presence of tissue-specific epigenetic alterations in normal colorectal mucosa. Similar to the results obtained in blood samples, no DM CpG islands or CpG sites were identified in LLS when compared to LS or healthy control samples (Table S8). No further differences were observed when analyzing the colorectal tumors from LLS and LS patients (Table S9). Methylome analysis of DM CpG islands in paired normal-tumor colonic samples from LLS individuals resulted in the identification of a high number of DM CpG islands (n=4380), most of them (n=3076) also identified as DM in normal-tumor samples from LS individuals (Figure 3), pointing to similar tumor methylation patterns in both groups. As expected (Pfeifer, 2018), strong hypermethylation of CpG islands and moderate hypomethylation of CpG sites within body genes was observed in tumors from both groups.

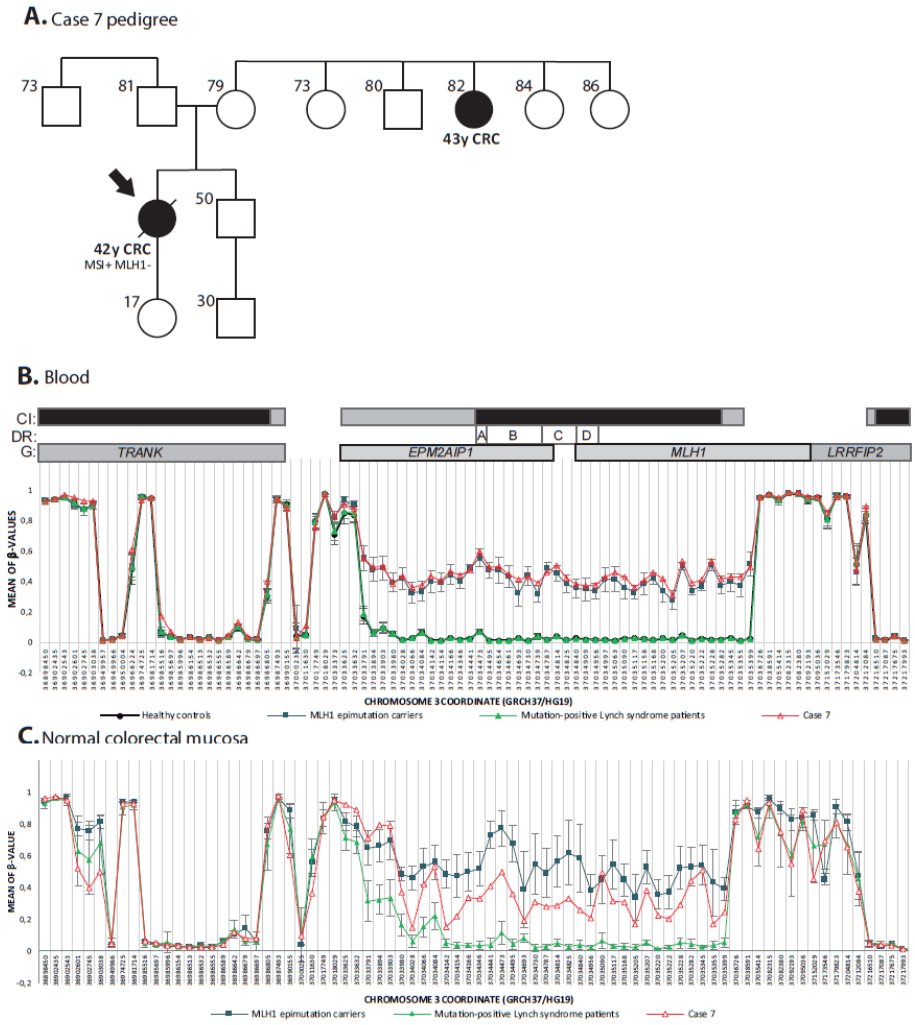


Figure 2: Identification of a new case of constitutional *MLH1* epimutation. **(A)** Pedigree of case 7. Representation of mean β-values in blood DNA **(B)** and FFPE normal colorectal mucosa **(C)** from case 7 against *MLH1* epimutation carriers, mutation-positive Lynch syndrome patients and healthy controls at differentially methylated region described for constitutional *MLH1* epimutation carriers. Chromosome coordinates of CpG sites are graphed at axis of abscissa. The locations of the CpG sites are not drawn to scale. CpG islands (CI) are represented as black rectangles and their shores are represented in grey. Location of Deng's promoter regions (DR) are indicated as white rectangles. Genes (G) including displayed CpG sites are represented as grey rectangles. Cytoband divisions (CB) are displayed as grey rectangles. Ensembl GRCh37 was taken as reference for gene coordinates.

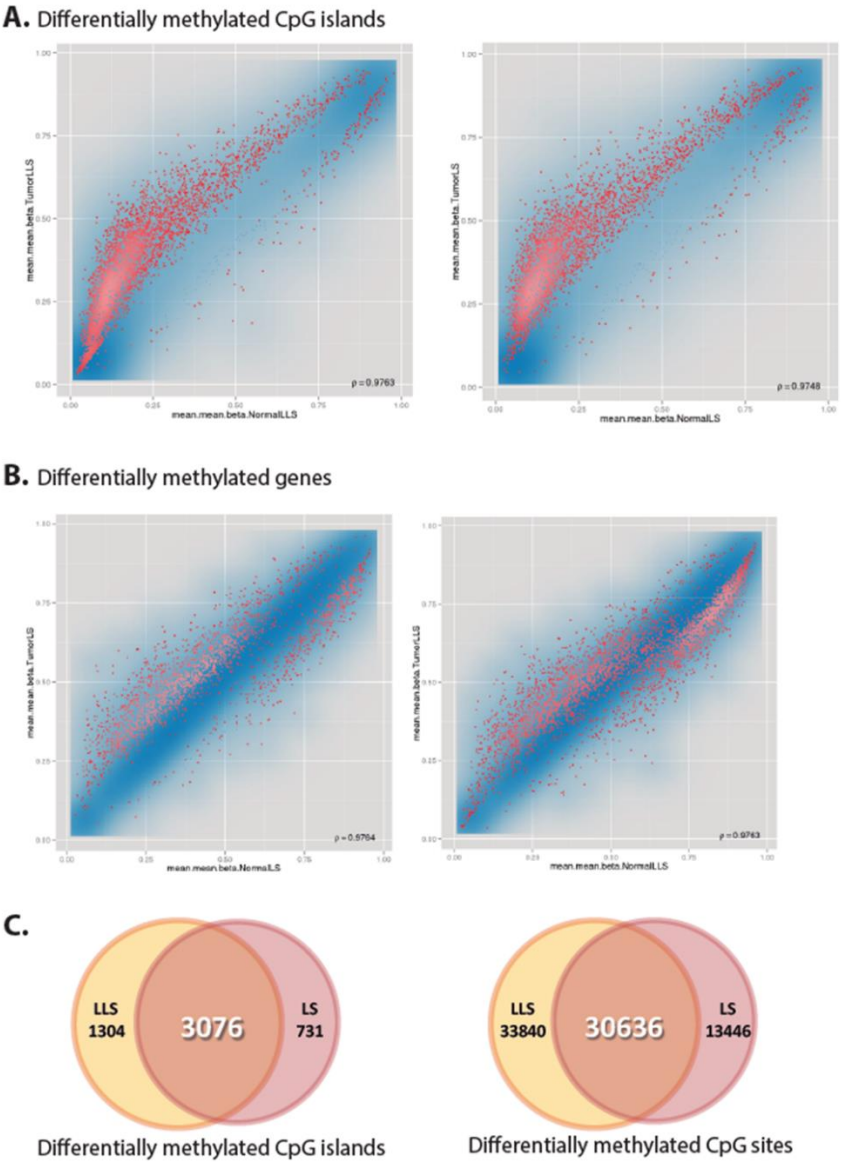


Figure 3: Scatterplot of the normalized mean B-values obtained using the Infinium 450k Human Methylation array to identify differentially methylated CpG islands **(A)** and genes **(B)** in tumors from LLS cases (left) and LS controls (right). The transparency corresponds to point density. One % of the points in the sparsest populated plot regions are drawn explicitly. The colored points represent differentially methylated CpG islands and genes with an FDR adjusted p-values lower than 0,05. **(C)** Venn diagrams of the differentially methylated CpG islands (left) and CpG sites (right), which shown the overlapping of epigenetic changes during tumorogenesis in LLS cases (yellow) and LS controls (red).

DISCUSSION

Individuals with MMR deficient tumors and no identified germline MMR mutations, account for more than a half of the cases being attended at genetic counseling units because of LS suspicion. They encompass a heterogeneous group of patients that may benefit from further stratification after comprehensive (epi)genetic characterization. By combining the use of variant pathogenicity assessment with ad-hoc designed panel and a global epigenetic characterization we have reclassified 9 of 115 cases as LS, one secondary to a constitutional epimutation. These results, together with the 5 cases from the same series reclassified in a previous work (Vargas-Parra *et al*, 2017) yielded a 12% (14/120) reclassification rate. Also, predicted deleterious variants in other CRC predisposing genes were found, which might explain an additional 11% of LLS cases. Except for the *MLH1* constitutional epimutation, no other clinically relevant differentially methylated regions were identified in LLS after a genome-wide methylome analysis.

In the present work, a customized NGS panel for the analysis of 26CRC associated genes allowed us to identify 2 previously missed *bona fide* MMR pathogenic variants in two families fulfilling the Amsterdam criteria. Fifteen additional MMR variants (9 identified by previous Sanger sequencing and 6 in the current MMR gene re-analysis) were also found in 16 individuals. RNA analyses in combination with multifactorial likelihood calculations resulted in the classification of 5 of them as pathogenic mutations. These results highlight the benefit of applying quantitative and qualitative analyses for variant interpretation and classification. Of note, four out of the 17 MMR variants identified (including pathogenic mutations and VUS) were not found in the candidate MMR gene according to the IHC pattern (cases 5, 82, 92 and 98), two of them finally classified as disease causing in the family (cases 82 and 92). These observations highlight the benefit of multiplex MMR genes panel testing in the presence of discordant IHC results.

Copy number variant (CNV) reanalysis using an updated MLPA test identified an *MSH2* exon 8 duplication in an additional case fulfilling Amsterdam criteria. These results further reinforce the notion that reanalysis of MMR genes using updated testing strategies should be considered in former LLS cases with strong individual and/or familial cancer history. While our NGS panel was not designed for CNV identification, recent advances in bioinformatic analysis have allowed the robust identification of rearrangements in other cancer gene panels, making it closer the routine use of NGS for CNV identification (Schmidt *et al*, 2017).

The use of subexome gene panels allowed the identification of additional candidate genes for LLS (de Voer *et al*, 2013; Goldberg *et al*, 2015; Vargas-Parra *et al*, 2017). In our cohort, variants were found in well-known CRC predisposition genes such as *APC* and *MUTYH,* as well as in newly emerging cancer predisposition genes, such as *MSH3, EXO1* and *FAN1*. Since patients with biallelic mutations in *MUTYH* were previously discarded in our LLS series (Castillejo *et al*, 2014; Seguí *et al*, 2015b), only 3 heterozygous *MUTYH* carriers were found (current study and Vargas-Parra *et al*, 2017). The estimated risk for monoallelic *MUTYH* mutation carriers does not support an earlier initiation of colonoscopy

screening, in line with current National Comprehensive Cancer Network recommendations (Katona *et al*, 2018).

There are a few reports of germline variants in *EXO1* and *MSH3* in LS suspected families, although the clinical significance of these variants was not determined (Jagmohan-Changur *et al*, 2003; Peltomäki, 2003). Moreover, *MSH3* variants have been found in combination with variants in LS-associated genes (Duraturo *et al*, 2011; Morak *et al*, 2017). Recently, biallelic *MSH3* mutations have been described to drive to adenomatous polyposis and CRC (Adam *et al*, 2016). In our cohort, 4 patients were carriers of monoallelic predicted pathogenic variants in *EXO1* or *MSH3* genes, and one *MSH3* carrier case harbored a tumor showing EMAST. These findings suggested the possibility of an oligogenic effect of *MSH3* and *EXO1* variants. Further studies are needed in order to elucidate the role of *MSH3* and *EXO1* in LLS.

Recent reports implicate *FAN1* as a colorectal cancer (CRC) and high-risk pancreatic cancer (PC) susceptibility gene (Seguí *et al*, 2015a; Smith *et al*, 2016). We found a patient carrying the *FAN*1 c.149T>G (p.Met50Arg) variant which was previously associated to functional defects and pancreatic cancer predisposition (Lachaud *et al*, 2016; Smith *et al*, 2016). However, the role of *FAN1* in cancer predisposition is currently a matter of controversy since no significant increase in the burden of *FAN1* mutations are detected in CRC cases versus controls (Broderick *et al*, 2017).

At the epigenetic level genome-wide methylation profiling was performed in DNA from blood and available colorectal tissue of all probands of our series. Individual methylation analysis of MMR genes allowed the identification of a new case of constitutional *MLH1* epimutation (Barrington *et al*, 2018; Dámaso *et al*, 2018; Hitchins, 2015; Morak *et al*, 2018; Pinto *et al*, 2018). This finding reinforces the need to rule out suggestive *MLH1* epimutation cases by analyzing DNA blood methylation in all early-onset cancer patients, irrespective of family history, where somatic methylation has not been assessed.

Genome-wide methylome analysis has ruled out other common constitutional epigenetic alterations associated with LLS individuals. This analysis also discarded the presence of colorectal tissue specific epimutations, as described for *MSH2* epimutations (Ligtenberg *et al*, 2009). However, we cannot completely rule out the existence of methylation aberrations in specific groups taking into account the diversity of IHC MMR patterns. Moreover, methylome analysis was not able to discriminate between tumors from LLS and LS individuals in line with the strong homogeneity of the epigenetic and genetic profile of MSI tumors previously reported (Hinoue *et al*, 2012; The Cancer Genome Atlas, 2012).

In all, germline reassessment of LS suspected cases is useful for the elucidation of the molecular basis of a relevant proportion of LLS cases. Subexome panels of cancer predisposing genes in combination with pathogenicity assessment of variants offered a good yield in reclassification, unmasking the limitations of IHC testing and the difficulty of detecting cryptic MMR mutations. The availability of advanced sequencing technologies will shed light on the molecular classification of LLS at the germline

level. When combined with somatic testing these technologies will likely fulfill their anticipated potential.

REFERENCES

Adam R, Spier I, Zhao B, Kloth M, Marquez J, Hinrichsen I, Kirfel J, Tafazzoli A, Horpaopan S, Uhlhaas S, Stienen D, Friedrichs N, Altmüller J, Laner A, Holzapfel S, Peters S, Kayser K, Thiele H, Holinski-Feder E, Marra G, Kristiansen G, Nöthen Markus M, Büttner R, Möslein G, Betz Regina C, Brieger A, Lifton Richard P, Aretz S (2016) Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *The American Journal of Human Genetics* **99**(2)**:** 337-351

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

Alexandrov L, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Boot A, Covington KR, Gordenin DA, Bergstrom E, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA, Mustonen V, Getz G, Rozen SG, Stratton MR (2018) The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv*

Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, Stratton MR (2015) Clock-like mutational processes in human somatic cells. *Nature Genetics* **47:** 1402

Alexandrov LB, Stratton MR (2014) Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current Opinion in Genetics & Development* **24:** 52-60

Arai H, Okudela K, Oshiro H, Komitsu N, Mitsui H, Nishii T, Tsuboi M, Nozawa A, Noishiki Y, Ohashi K, Inui K, Masuda M (2013) Elevated microsatellite alterations at selected tetra-nucleotide (EMAST) in non-small cell lung cancers--a potential determinant of susceptibility to multiple malignancies. *International Journal of Clinical and Experimental Pathology* **6**(3)**:** 395-410

Barrington DA, Champion ML, Boitano TKL, Walters-Haygood CL, Farmer MB, Alvarez RD, Estes JM, Leath CA, III (2018) Characteristics of African American women at high-risk for ovarian cancer in the southeast: Results from a Gynecologic Cancer Risk Assessment Clinic. *Gynecologic Oncology* **149**(2)**:** 337-340

Bennett KL, Mester J, Eng C (2010) Germline epigenetic regulation of killin in cowden and cowden-like syndrome. *JAMA* **304**(24)**:** 2724-2731

Betz B, Theiss S, Aktas M, Konermann C, Goecke TO, Möslein G, Schaal H, Royer-Pokora B (2009) 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**(1)**:** 123

Broderick P, Dobbins SE, Chubb D, Kinnersley B, Dunlop MG, Tomlinson I, Houlston RS (2017) Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients: a Systematic Review. *Gastroenterology* **152**(1)**:** 75-77.e4

Buchanan DD, Rosty C, Clendenning M, Spurdle AB, Win AK (2014) Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). *The Application of Clinical Genetics* **7:** 183-193

Burger M, Burger SJ, Denzinger S, Wild PJ, Wieland WF, Blaszyk H, Obermann EC, Stoehr R, Hartmann A (2006) Elevated Microsatellite Instability at Selected Tetranucleotide Repeats does not Correlate with Clinicopathologic Features of Bladder Cancer. *European Urology* **50**(4)**:** 770-776

Carethers J, Koi M, Tseng-Rogenski S (2015) EMAST is a Form of Microsatellite Instability That is Initiated by Inflammation and Modulates Colorectal Cancer Progression. *Genes* **6**(2)**:** 185

Castillejo A, Vargas G, Castillejo MI, Navarro M, Barberá VM, González S, Hernández-Illán E, Brunet J, Ramón y Cajal T, Balmaña J, Oltra S, Iglesias S, Velasco À, Solanes A, Campos O, Sánchez Heras AB, Gallego J, Carrasco E, González Juan D, Segura Á, Chirivella I, Juan MJ, Tena I, Lázaro C, Blanco I, Pineda M, Capellá G, Soto JL (2014) Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. *European Journal of Cancer* **50**(13)**:** 2241-2250

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* **31:** 213

Dámaso E, Castillejo A, Arias MdM, Canet-Hermida J, Navarro M, Del Valle J, Campos O, Fernández A, Marín F, Turchetti D, García-Díaz JdD, Lázaro C, Genuardi M, Rueda D, Alonso Á, Soto JL, Hitchins M, Pineda M, Capellá G (2018) Primary constitutional *MLH1* epimutations: a focal epigenetic event. *British Journal of Cancer* **In press**

de Voer RM, Geurts van Kessel A, Weren RDA, Ligtenberg MJL, Smeets D, Fu L, Vreede L, Kamping EJ, Verwiel ETP, Hahn MM, Ariaans M, Spruijt L, van Essen T, Houge G, Schackert HK, Sheng JQ, Venselaar H, van Ravenswaaij–Arts CMA, van Krieken JHJM, Hoogerbrugge N, Kuiper RP (2013) Germline Mutations in the Spindle Assembly Checkpoint Genes BUB1 and BUB3 Are Risk Factors for Colorectal Cancer. *Gastroenterology* **145**(3)**:** 544-547

Deng G, Chen A, Hong J, Chae HS, Kim YS (1999) Methylation of CpG in a Small Region of the hMLH1 Promoter Invariably Correlates with the Absence of Gene Expression. *Cancer Research* **59**(9)**:** 2029-2023

Duraturo F, Liccardo R, Cavallo A, Rosa MD, Grosso M, Izzo P (2011) Association of low-risk MSH3 and MSH2 variant alleles with Lynch syndrome: Probability of synergistic effects. *International Journal of Cancer* **129**(7)**:** 1643-1650

Elsayed FA, Kets CM, Ruano D, van den Akker B, Mensenkamp AR, Schrumpf M, Nielsen M, Wijnen JT, Tops CM, Ligtenberg MJ, Vasen HFA, Hes FJ, Morreau H, van Wezel T (2014) Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. *European Journal Of Human Genetics* **23:** 1080

Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM, Messeguer X (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Research* **31**(13)**:** 3651-3653

Gardès P, Forveille M, Alyanakian M-A, Aucouturier P, Ilencikova D, Leroux D, Rahner N, Mazerolles F, Fischer A, Kracker S, Durandy A (2012) Human MSH6 Deficiency Is Associated with Impaired Antibody Maturation. *The Journal of Immunology* **188**(4)**:** 2023-2029

Geurts-Giele WR, Leenen CH, Dubbink HJ, Meijssen IC, Post E, Sleddens HF, Kuipers EJ, Goverde A, van den Ouweland AM, van Lier MG, Steyerberg EW, van Leerdam ME, Wagner A, Dinjens WN (2014) Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. *The Journal of Pathology* **234**(4)**:** 548-559

Goldberg Y, Halpern N, Hubert A, Adler SN, Cohen S, Plesser-Duvdevani M, Pappo O, Shaag A, Meiner V (2015) Mutated MCM9 is associated with predisposition to hereditary mixed polyposis and colorectal cancer in addition to primary ovarian failure. *Cancer Genetics* **208**(12)**:** 621-624

Gordon S, Akopyan G, Garban H, Bonavida B (2005) Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25:** 1125

Hansmann T, Pliushch G, Leubner M, Kroll P, Endt D, Gehrig A, Preisler-Adams S, Wieacker P, Haaf T (2012) Constitutive promoter methylation of BRCA1 and RAD51C in patients with familial ovarian cancer and early-onset sporadic breast cancer. *Human Molecular Genetics* **21**(21)**:** 4669-4679

Haraldsdottir S, Hampel H, Tomsic J, Frankel WL, Pearlman R, de la Chapelle A, Pritchard CC (2014) Colon and Endometrial Cancers With Mismatch Repair Deficiency Can Arise From Somatic, Rather Than Germline, Mutations. *Gastroenterology* **147**(6)**:** 1308-1316.e1

Hinoue T, Weisenberger D, Lange C, Shen H, Byun H-M, Van Den Berg D, Malik S, Pan F, Noushmehr H, van Dijk C, Tollenaar R, Laird P (2012) Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Research* **22**(2)**:** 271-282

Hitchins MP (2015) Constitutional epimutation as a mechanism for cancer causality and heritability? *Nat Rev Cancer* **15**(10)**:** 625-634

Jagmohan-Changur S, Poikonen T, Vilkki S, Launonen V, Wikman F, Orntoft TF, Møller P, Vasen H, Tops C, Kolodner RD, Mecklin J-P, Järvinen H, Bevan S, Houlston RS, Aaltonen LA, Fodde R, Wijnen J, Karhu A (2003) EXO1 Variants Occur Commonly in Normal Population. *Evidence against a Role in Hereditary Nonpolyposis Colorectal Cancer* **63**(1)**:** 154-158

Jansen AML, van Wezel T, van den Akker BEWM, Ventayol Garcia M, Ruano D, Tops CMJ, Wagner A, Letteboer TGW, Gómez-García EB, Devilee P, Wijnen JT, Hes FJ, Morreau H (2016) Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. *European Journal Of Human Genetics* **24:** 1089

Jasperson K, Samowitz W, Burt R (2011) Constitutional mismatch repair-deficiency syndrome presenting as colonic adenomatous polyposis: clues from the skin. *Clinical Genetics* **80**(4)**:** 394-397

Katona BW, Yurgelun MB, Garber JE, Offit K, Domchek SM, Robson ME, Stadler ZK (2018) A counseling framework for moderate-penetrance colorectal cancer susceptibility genes. *Genetics In Medicine*

Lachaud C, Moreno A, Marchesi F, Toth R, Blow JJ, Rouse J (2016) Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science* **351**(6275)**:** 846-849

Leung SY, Chan TL, Yuen ST (2007) Reply to "Heritable germline epimutation is not the same as transgenerational epigenetic inheritance". *Nature Genetics* **39:** 576

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**(14)**:** 1754-1760

Ligtenberg MJL, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TYH, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJB, Tsui WY, Kong CK, Brunner HG, van Kessel AG, Yuen ST, van Krieken JHJM, Leung SY, Hoogerbrugge N (2009) Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3[prime] exons of TACSTD1. *Nat Genet* **41**(1)**:** 112-117

Liu Q, Hesson LB, Nunez AC, Packham D, Williams R, Ward RL, Sloane MA (2016) A cryptic paracentric inversion of MSH2 exons 2–6 causes Lynch syndrome. *Carcinogenesis* **37**(1)**:** 10-17

Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP (2015) Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* **15**(3)**:** 181-194

Mensenkamp AR, Vogelaar IP, van Zelst–Stams WAG, Goossens M, Ouchene H, Hendriks–Cornelissen SJB, Kwint MP, Hoogerbrugge N, Nagtegaal ID, Ligtenberg MJL (2014) Somatic Mutations in MLH1 and MSH2 Are a Frequent Cause of Mismatch-Repair Deficiency in Lynch Syndrome-Like Tumors. *Gastroenterology* **146**(3)**:** 643-646.e8

Messeguer X, Escudero R, Farré D, Nuñez O, Martínez J, Albà MM (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**(2)**:** 333-334

Meyer C, Brieger A, Plotz G, Weber N, Passmann S, Dingermann T, Zeuzem S, Trojan J, Marschalek R (2009) An Interstitial Deletion at 3p21.3 Results in the Genetic Fusion of MLH1 and ITGA9 in a Lynch Syndrome Family. *Clinical Cancer Research* **15**(3)**:** 762-769

Morak M, Heidenreich B, Keller G, Hampel H, Laner A, de la Chapelle A, Holinski-Feder E (2014) Biallelic MUTYH mutations can mimic Lynch syndrome. *European Journal Of Human Genetics* **22:** 1334

Morak M, Ibisler A, Keller G, Jessen E, Laner A, Gonzales-Fassrainer D, Locher M, Massdorf T, Nissen AM, Benet-Pagès A, Holinski-Feder E (2018) Comprehensive analysis of the MLH1 promoter region in 480 patients with colorectal cancer and 1150 controls reveals new variants including one with a heritable constitutional MLH1 epimutation. *Journal of Medical Genetics* **55**(4)**:** 240-248

Morak M, Käsbauer S, Kerscher M, Laner A, Nissen AM, Benet-Pagès A, Schackert HK, Keller G, Massdorf T, Holinski-Feder E (2017) Loss of MSH2 and MSH6 due to heterozygous germline defects in MSH3 and MSH6. *Familial Cancer* **16**(4)**:** 491-500

Morak M, Koehler U, Schackert HK, Steinke V, Royer-Pokora B, Schulmann K, Kloor M, Höchter W, Weingart J, Keiling C, Massdorf T, Holinski-Feder E (2011) Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing Lynch syndrome. *Journal of Medical Genetics* **48**(8)**:** 513-519

Mork ME, Borras E, Taggart MW, Cuddy A, Bannon SA, You YN, Lynch PM, Ramirez PT, Rodriguez-Bigas MA, Vilar E (2016) Identification of a novel PMS2 alteration c.505C>G (R169G) in trans with a PMS2 pathogenic mutation in a patient with constitutional mismatch repair deficiency. *Familial Cancer* **15**(4)**:** 587-591

Ngeow J, Mester J, Rybicki L, Ni Y, Milas M, Eng C (2011) Incidence and Clinical Characteristics of Thyroid Cancer in Prospective Series of Individuals with Cowden and Cowden-Like Syndrome Characterized by Germline PTEN, SDH, or KLLN Alterations. *The Journal of Clinical Endocrinology & Metabolism* **96**(12)**:** E2063–E2071

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn S-M, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GKJ, Jang SJ, Jones DR, Kim H-Y, King TA, Krishnamurthy S, Lee HJ, Lee J-Y, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, Veer Lvt, Tutt A, Knappskog S, Tan BKT, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JWM,

Børresen-Dale A-L, Richardson AL, Kong G, Thomas G, Stratton MR (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534:** 47

Obmolova G, Ban C, Hsieh P, Yang W (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407:** 703

Okkels H, Sunde L, Lindorff-Larsen K, Thorlacius-Ussing O, Gandrup P, Lindebjerg J, StubbeTeglbjærg P, Oestergaard JR, Nielsen FC, Krarup HB (2006) Polyposis and early cancer in a patient with low penetrant mutations in MSH6 and APC: hereditary colorectal cancer as a polygenic trait. *International Journal of Colorectal Disease* **21**(8)**:** 847-850

Peltomäki P (2003) Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer. *Journal of Clinical Oncology* **21**(6)**:** 1174-1179

Peltomäki P (2016) Update on Lynch syndrome genomics. *Familial Cancer* **15**(3)**:** 385-393

Pfeifer G (2018) Defining Driver DNA Methylation Changes in Human Cancer. *International Journal of Molecular Sciences* **19**(4)**:** 1166

Pineda M, Mur P, Iniesta MD, Borras E, Campos O, Vargas G, Iglesias S, Fernandez A, Gruber SB, Lazaro C, Brunet J, Navarro M, Blanco I, Capella G (2012) MLH1 methylation screening is effective in identifying epimutation carriers. *Eur J Hum Genet* **20**(12)**:** 1256-1264

Pinto D, Pinto C, Guerra J, Pinheiro M, Santos R, Vedeld HM, Yohannes Z, Peixoto A, Santos C, Pinto P, Lopes P, Lothe R, Lind Guro E, Henrique R, Teixeira MR (2018) Contribution of MLH1 constitutional methylation for Lynch syndrome diagnosis in patients with tumor MLH1 downregulation. *Cancer Medicine* **7**(2)**:** 433-444

Plaschke J, Linnebacher M, Kloor M, Gebert J, Cremer FW, Tinschert S, Aust DE, von Knebel Doeberitz M, Schackert HK (2006) Compound heterozygosity for two MSH6 mutations in a patient with early onset of HNPCC-associated cancers, but without hematological malignancy and brain tumor. *European Journal Of Human Genetics* **14:** 561

Plon SE, Cooper HP, Parks B, Dhar SU, Kelly PA, Weinberg AD, Staggs S, Wang T, Hilsenbeck S (2011) Genetic testing and cancer risk management recommendations by physicians for at-risk relatives. *Genetics In Medicine* **13:** 148

Quiñonez-Silva G, Dávalos-Salas M, Recillas-Targa F, Ostrosky-Wegman P, Aranda DA, Benítez-Bribiesca L (2016) "Monoallelic germline methylation and sequence variant in the promoter of the RB1 gene: a possible constitutive epimutation in hereditary retinoblastoma". *Clinical Epigenetics* **8**(1)**:** 1

Rahner N, Höefler G, Högenauer C, Lackner C, Steinke V, Sengteller M, Friedl W, Aretz S, Propping P, Mangold E, Walldorf C (2008) Compound heterozygosity for two MSH6 mutations in a patient with early onset colorectal cancer, vitiligo and systemic lupus erythematosus. *American Journal of Medical Genetics Part A* **146A**(10)**:** 1314-1319

Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen S-S, Hackanson B, Grever MR, Lucas DM, Matkovic JJ, Lin TS, Kipps TJ, Murray F, Weisenburger D, Sanger W, Lynch J, Watson P, Jansen M, Yoshinaga Y, Rosenquist R, de Jong PJ, Coggill P, Beck S, Lynch H, de la Chapelle A, Plass C (2007) Downregulation of Death-Associated Protein Kinase 1 (DAPK1) in Chronic Lymphocytic Leukemia. *Cell* **129**(5)**:** 879-890

Rhees J, Arnold M, Boland CR (2014) Inversion of exons 1–7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. *Familial Cancer* **13**(2)**:** 219-225

Rodríguez–Soler M, Pérez–Carbonell L, Guarinos C, Zapater P, Castillejo A, Barberá VM, Juárez M, Bessa X, Xicola RM, Clofent J, Bujanda L, Balaguer F, Reñé JM, de–Castro L, Marín–Gabriel JC, Lanas A, Cubiella J, Nicolás–Pérez D, Brea–Fernández A, Castellví–Bel S, Alenda C, Ruiz–Ponte C, Carracedo A, Castells A, Andreu M, Llor X, Soto JL, Payá A, Jover R (2013) Risk of Cancer in Cases of Suspected Lynch Syndrome Without Germline Mutation. *Gastroenterology* **144**(5)**:** 926-932.e1

Schmidt AY, Hansen TvO, Ahlborn LB, Jønson L, Yde CW, Nielsen FC (2017) Next-Generation Sequencing-Based Detection ofmline Copy Number Variations in BRCA1/BRCA2. *The Journal of Molecular Diagnostics* **19**(6)**:** 809-816

Seguí N, Mina LB, Lázaro C, Sanz-Pamplona R, Pons T, Navarro M, Bellido F, López-Doriga A, Valdés-Mas R, Pineda M, Guinó E, Vidal A, Soto JL, Caldés T, Durán M, Urioste M, Rueda D, Brunet J, Balbín M, Blay P, Iglesias S, Garré P, Lastra E, Sánchez-Heras AB, Valencia A, Moreno V, Pujana MÁ, Villanueva A, Blanco I, Capellá G, Surrallés J, Puente XS, Valle L (2015a) Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology* **149**(3)**:** 563-566

Seguí N, Navarro M, Pineda M, Köger N, Bellido F, González S, Campos O, Iglesias S, Valdés-Mas R, López-Doriga A, Gut M, Blanco I, Lázaro C, Capellá G, Puente XS, Plotz G, Valle L (2015b) Exome sequencing identifies MUTYH mutations in a family with colorectal cancer and an atypical phenotype. *Gut* **64**(2)**:** 355-356

Smith AL, Alirezaie N, Connor A, Chan-Seng-Yue M, Grant R, Selander I, Bascuñana C, Borgida A, Hall A, Whelan T, Holter S, McPherson T, Cleary S, Petersen GM, Omeroglu A, Saloustros E, McPherson J, Stein LD, Foulkes WD, Majewski J, Gallinger S, Zogopoulos G (2016) Candidate DNA repair susceptibility genes identified by exome sequencing in high-risk pancreatic cancer. *Cancer Letters* **370**(2)**:** 302-312

Sourrouille I, Coulet F, Lefevre JH, Colas C, Eyries M, Svrcek M, Bardier-Dupas A, Parc Y, Soubrier F (2013) Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. *Familial Cancer* **12**(1)**:** 27-33

Stoehr C, Burger M, Stoehr R, Bertz S, Ruemmele P, Hofstaedter F, Denzinger S, Wieland WF, Hartmann A, Walter B (2012) Mismatch Repair Proteins hMLH1 and hMSH2 Are Differently Expressed in the Three Main Subtypes of Sporadic Renal Cell Carcinoma. *Pathobiology* **79**(3)**:** 162-168

The Cancer Genome Atlas N (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487:** 330

Thompson BA. Interpreting the clinical significance of mismatch repair gene sequence variants. PhD Thesis PhD Thesis, The University of Queensland, 2014

Vargas-Parra GM, González-Acosta M, Thompson BA, Gómez C, Fernández A, Dámaso E, Pons T, Morak M, del Valle J, Iglesias S, Velasco À, Solanes A, Sanjuan X, Padilla N, de la Cruz X, Valencia A, Holinski-Feder E, Brunet J, Feliubadaló L, Lázaro C, Navarro M, Pineda M, Capellá G (2017) Elucidating the molecular basis of MSH2-deficient tumors by combined germline and somatic analysis. *International Journal of Cancer* **141**(7)**:** 1365-1380

Wagner A, van der Klift H, Franken P, Wijnen J, Breukel C, Bezrookove V, Smits R, Kinarsky Y, Barrows A, Franklin B, Lynch J, Lynch H, Fodde R (2002) A 10-Mb paracentric inversion of chromosome arm 2p inactivates MSH2 and is responsible for hereditary nonpolyposis colorectal cancer in a North-American kindred. *Genes Chromosomes Cancer* **35**(1)**:** 49-57

Win AK, Buchanan DD, Rosty C, MacInnis RJ, Dowty JG, Dite GS, Giles GG, Southey MC, Young JP, Clendenning M, Walsh MD, Walters RJ, Boussioutas A, Smyrk TC, Thibodeau SN, Baron JA, Potter JD, Newcomb PA, Le Marchand L, Haile RW, Gallinger S, Lindor NM, Hopper JL, Ahnen DJ, Jenkins MA (2015) Role of tumour

molecular and pathology features to estimate colorectal cancer risk for first-degree relatives. *Gut* **64**(1)**:** 101-110

Yamamoto H, Imai K (2015) Microsatellite instability: an update. *Archives of Toxicology* **89**(6)**:** 899-921

Young J, Barker MA, Simms LA, Walsh MD, Biden KG, Buchanan D, Buttenshaw R, Whitehall VLJ, Arnold S, Jackson L, Kambara T, Spring KJ, Jenkins MA, Walker GJ, Hopper JL, Leggett BA, Jass JR (2005) Evidence for BRAF mutation and variable levels of microsatellite instability in a syndrome of familial colorectal cancer. *Clinical Gastroenterology and Hepatology* **3**(3)**:** 254-263

SUPPLEMENTARY MATERIAL

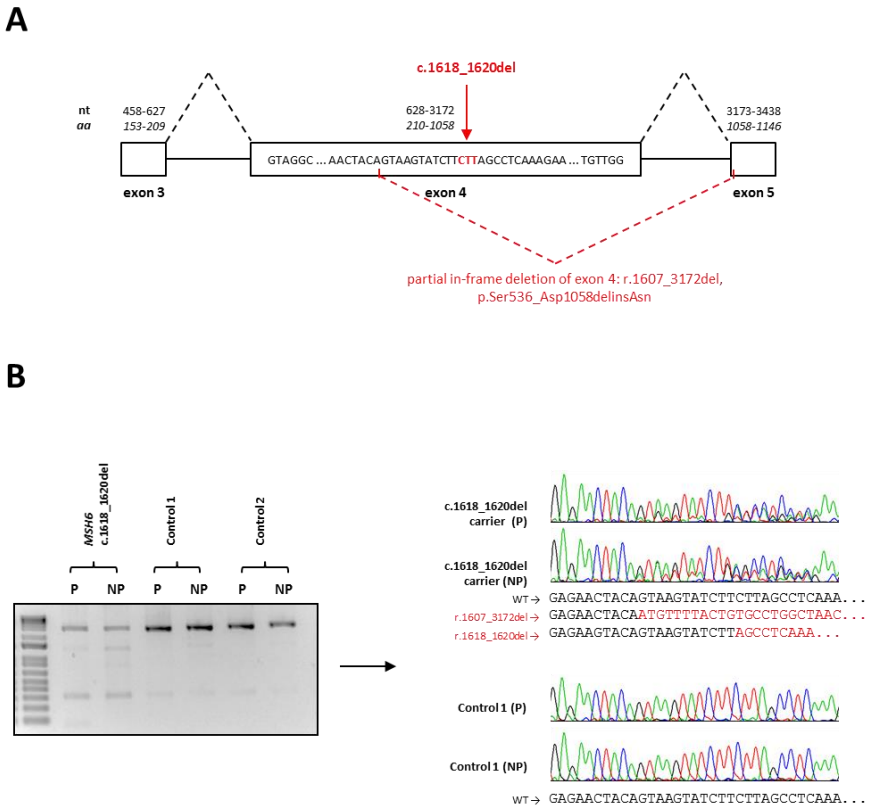
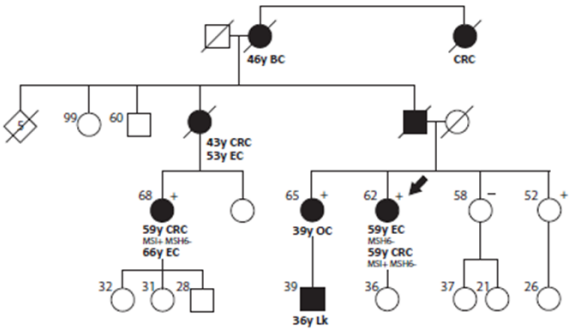
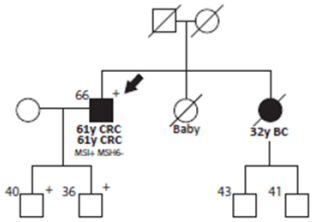


Figure S1. Splicing analysis of *MSH6* c.1618_1620del variant. A) Schematic overview of *MSH6* exons. **B)** Left: agarose gel showing RT-PCR products. Right: direct sequencing of RT-PCR products showing the wt transcript and the aberrant transcript corresponding to a partial in-frame deletion of exon 4 (r.1607_3172del; p.Ser536_Asp1058delinsAsn) coexisting with the full-length variant transcript (r.1618_1620del). Abbreviations: nt, nucleotide sequence; aa, amino acid sequence; P, puromycin; NP, no puromycin.

A. AF1 (*MSH6* c.1153_1155delAGG)



B. AF2 (*MSH6* c.3150_3161dup)



C. AF3 (*MSH6* c.3150_3161dup)

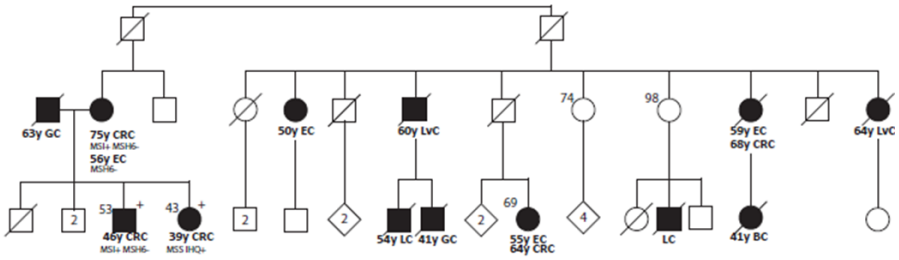


Figure S2: Pedigrees from additional families included for pathogenecity assessment of MMR VUS by multifactorial analysis. Abbreviations: CRC=colorectal cancer, BC=breast cancer, LvC=liver cancer, LC=lung cancer, EC=endometrial cancer, GC=gastric cancer.

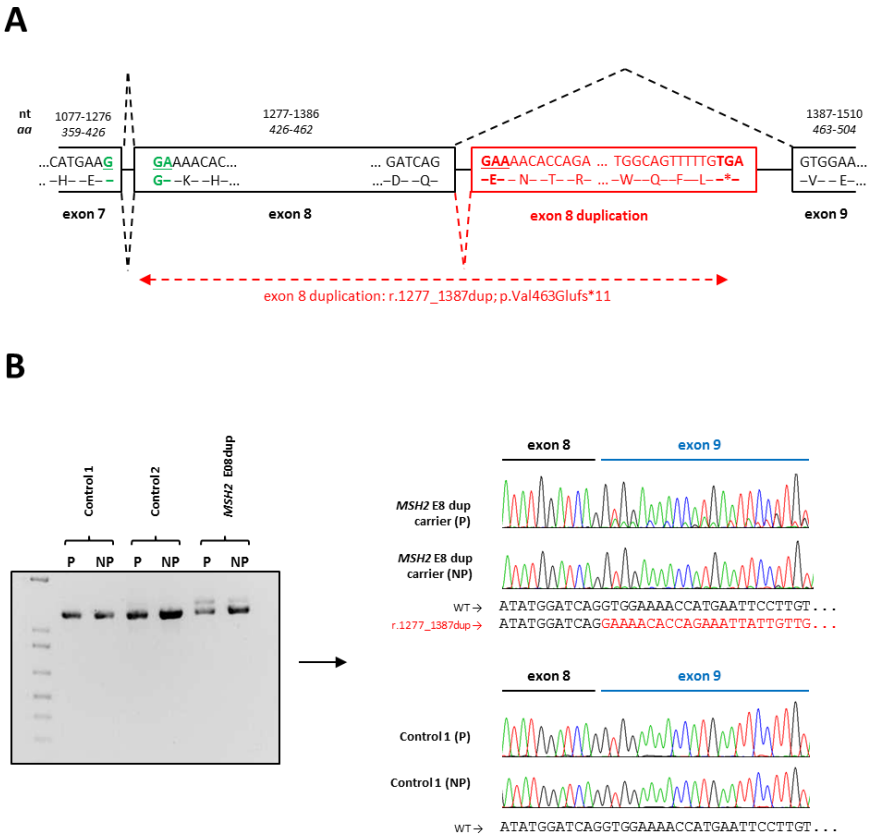


Figure S3: Splicing analysis of *MSH2* exon 8 duplication. A) Schematic overview of *MSH2* exons. **B)** Left: agarose gel showing RT-PCR products. Right: direct sequencing of RT-PCR products showing the wt transcript and the aberrant transcript corresponding to a duplication of exon 8. Abbreviations: nt, nucleotide sequence; aa, amino acid sequence; P, puromycin; NP, no puromycin.

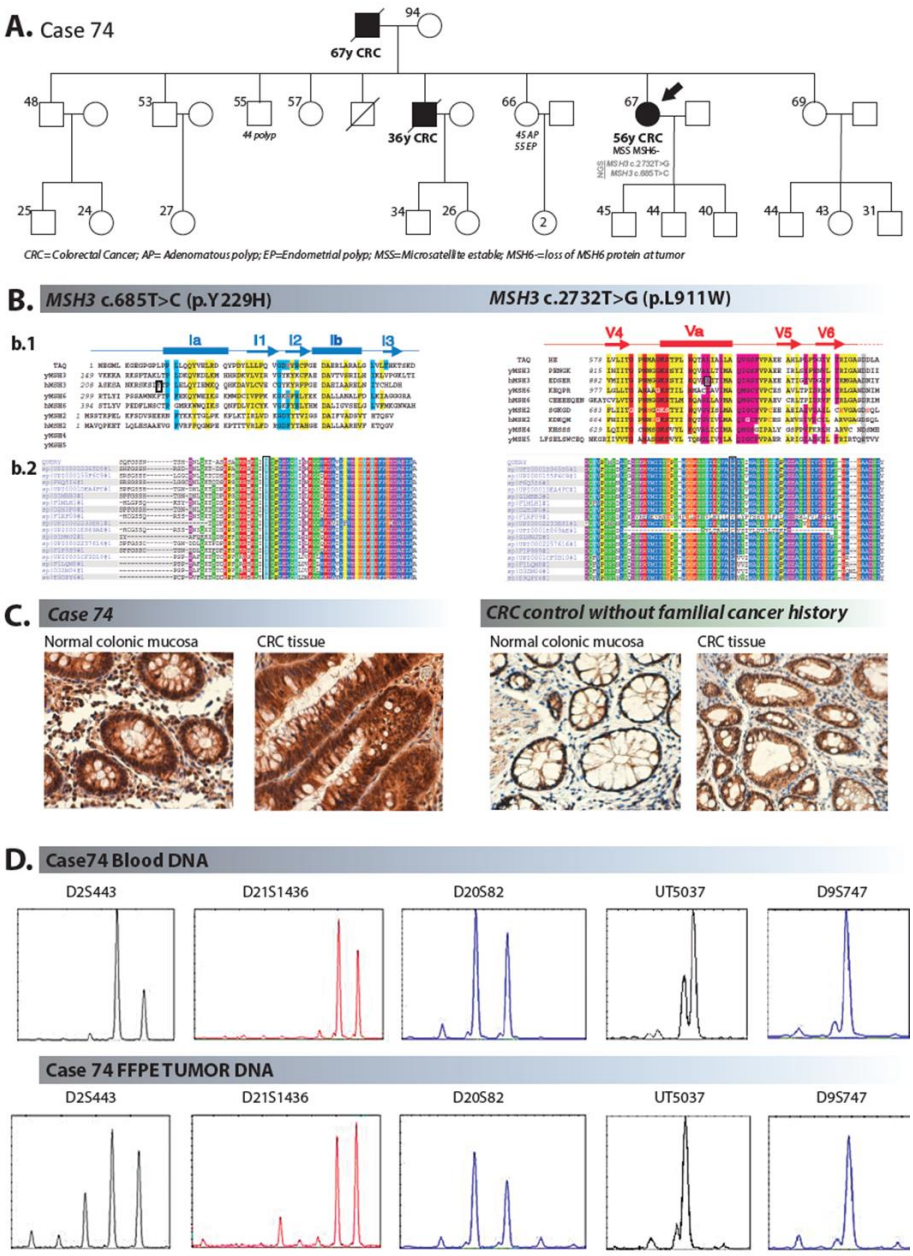
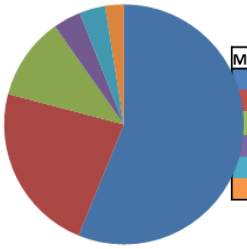


Figure S4: Pathogenicity assessment of *MSH3* variants found in case 74. **(A)** Pedigree of case 74. **(B)** Structure-based sequence alignment of Mut S homologues (upper panel) and multiple sequence alignment (lower panel). **(B.1)** MutS homologues from human (hMSH), S. cerevisiae (yMSH) and T. aquaticus are shown. The secondary structures observed in TAQ MutS are indicated above the aligned sequence. Conserved residues for structural integrity are highlighted in yellow, for DNA recognition in blue, for protein dimerization in green, for ATPase activity in red and for interdomain interactions in purple. Residues that, when mutated, cause defective mismatch repair in yeast, or HNPCC in humans are coloured red, or, if they are highlighted in purple and red,

white. The five nucleotide-binding motifs are indicated beneath the sequence alignment. Variants are marked with a black box on human *MSH3* protein. Image modified from Obmolova et al 2000. **(B.2)** Polyphen-2/UniProtKB/UniRef100 aligment against different animal species. 75 amino acids surrounding the variant position (marked with a black box) are shown. **(C)** MSH3 IHC staining of patient 74 and a CRC control without family history of cancer. **(D)** EMAST analysis in blood and FFPE tumor DNA from case 74.



Mutational signature	Contribution (%)	Proposed etiology
Signature 6	56,19	Defective DNA mismatch repair
Signature 1	22,89	Deamination of 5-methylcytosine (present in all cancer types)
Signature 16	11,16	-
Signature 26	3,73	Defective DNA mismatch repair
Signature 12	3,51	-
Signature 7	2,52	UV exposure

Figure S5: Contribution of COSMIC mutational signatures to the MMR-deficient tumor developed by patient 53, carrier of the POLE variant c.898A>G (p.Ile300Val). Signatures contributions were calculated with deconstructSigs (Rosenthal 2016) from whole-exome sequencing data. The proposed etiology for the different signatures was obtained from Alexandrov *et al*, 2018.

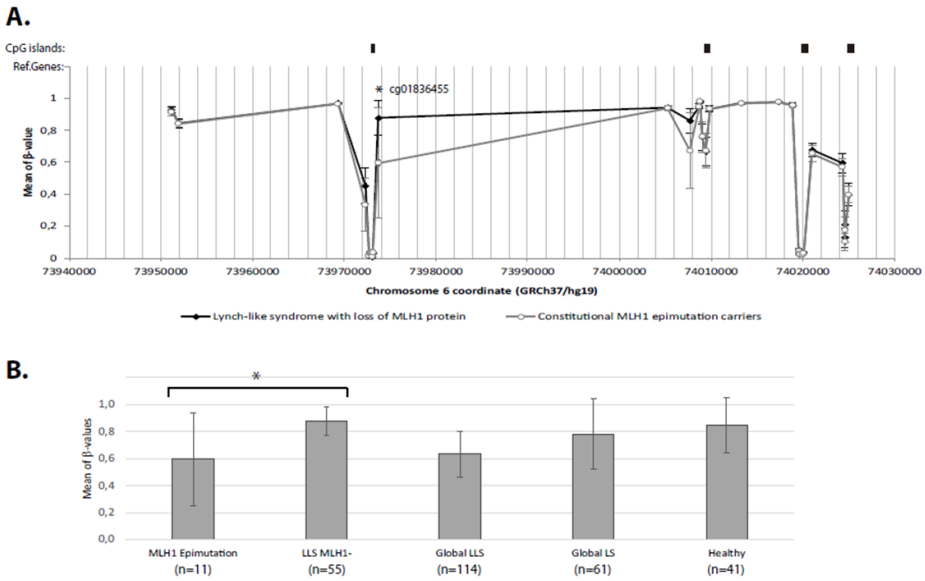


Figure S6: Differentially methylated CpG site was found inside *KHDC1* gene. **(A)** Methylation pattern of the locus in MLH1-deficient Lynch-like syndrome patients (black) and constitutional MLH1 epimutation carriers (grey). The location of reference genes and CpG islands is represented above according to UCSC genome browser. Significative differentially methylated CpG site is marked up with an asterisk and its probe ID is given. **(B)** Mean of Beta-values of each group of study for cg01836455 probe. Significant differences are marked up with an asterisk.

Table S1- Clinico-pathological data of the LLS patients included in the study. (#) Previously included in Vargas-Parra et al 2016. *) Not previously reported MMR variants were classified according to the Insight rules. Abbreviations: M=Male; F=Female; A=Amsterdam criteria; B+number=Revised Bethesda criteria; CRC=Colorectal Cancer; EC=Endometrial Cancer; OC=Ovarian Cancer; SBC=Small Bowell Cancer; MSI+=Microsatellite Instabillity; NP=Not performed; WT=Wildtype; UM=Unmethylated; P=Positive; N=Negative; NV=Not Valuable; PR=Promoter Region; CR=Coding Region; P+CR=Promoter and Coding Region; B=Blood; N=Normal colonic mucosa; T=Colorectal Tumor

Case ID	GENDER	CLINICAL CRITERIA	TUMOR DATA									MMR Mutational Analysis by Sanger Sequencing / SSCP					Available samples
			TUMOR TYPE	AGE AT DIAGNOSIS	MSI	BRAF	MLH1 Methylation	MLH1 IHC	MSH2 IHC	MSH6 IHC	PMS2 IHC	MLH1	MSH2	MSH6	PMS2	Identified VUS	
1	F	B1-3	CRC	44	MSI+	WT	UM	N	P	P	NP	P+CR	CR	NP	NP	-	B, T
2	F	B1	CRC	31	MSI+	WT	UM	N	P	P	NP	P+CR	CR	NP	NP	-	B, T
3	M	B1	CRC	49	MSI+	WT	UM	N	P	P	NP	P+CR	CR	NP	NP	-	B
4	F	B1	CRC	48	MSI+	WT	UM	N	P	P	NP	P+CR	CR	NP	NP	-	B
5	M	B1	CRC	40	MSI+	WT	UM	N	P	P	NP	P+CR	NP	NP	NP	-	B, N, T
6	F	B1	CRC	30	MSI+	WT	UM	N	P	NV	N	P+CR	NP	NP	NP	-	B, T
7	F	B1	CRC	42	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B, N, T
8	F	B1	CRC	35	MSI+	WT	UM	N	P	P	N	P+CR	NP	NP	NP	-	B, T
9	F	B1	CRC	25	NP	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B, N, T
10	M	B1	CRC	48	NP	WT	UM	N	NP	NP	NP	P+CR	NP	NP	NP	-	B, T
11	M	B1	SBC	30	NP	WT	UM	N	NP	NP	NP	P+CR	NP	NP	NP	-	B
12	M	AP	CRC	66	MSI+	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
13	M	B4	CRC	70	MSI+	WT	UM	N	P	P	NP	P+CR	NP	NP	NP	-	B
14	M	B3	CRC	59	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
15	M	B3	CRC	39	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
16	M	AP	CRC	63	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
17	F	B4	CRC	71	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
18	F	AP	CRC	61	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
19	F	B1	CRC	52	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
20	F	B1	CRC	32	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
21	F	AP	CRC	70	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
22	M	AP	CRC	73	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
23	M	B3	CRC	54	MSI+	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
24	M	B3	CRC	55	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
25	M	AP	CRC	82	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
26	M	AP	CRC	66	NP	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
27	F	B1	CRC	39	MSI+	NV	NP	N	P	P	NP	CR	NP	NP	NP	-	B, T
28	F	B1	CRC	41	NV	WT	UM	N	P	P	N	P+CR	NP	NP	NP	-	B, N, T
29	M	B1	CRC	41	MSI+	NP	UM	N	P	P	N	P+CR	NP	NP	NP	-	B
30	M	B2	CRC CRC	61 68	NP MSI+	WT	NP UM	NP N	NP P	NP P	NP N	P+CR	NP	NP	NP	-	B
31	F	B3	CRC	55	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B, N
32	F	B5	CRC	78	MSI+	WT	NP	N	P	P	P	P+CR	NP	NP	NP	-	B
33	M	A	CRC	40	MSI+	WT	UM	N	NP	NP	NP	P+CR	NP	NP	NP	-	B
34	F	AP	Unkown	Unkown	MSI+	NP	NP	N	NP	NP	NP	CR	NP	NP	NP	-	B
35	M	A	CRC	49	MSI+	NP	NP	NV	P	NP	NP	CR	CR	NP	NP	*MLH1* c.25C>T, p.Arg9Trp, (Class 3) *APC* c.1958+3A>G (class 5 (Borrás et al 2012)	B
36	M	B1	CRC	50	MSI+	WT	NP	N	P	P	NP	P+CR	CR	NP	NP	-	B
37	F	AP	CRC	52	MSI+	NP	NP	N	P	NP	NP	P+CR	NP	NP	NP	-	B
38	F	B5	CRC	75	MSI+	NP	NP	N	P	P	NP	P+CR	CR	NP	NP	-	B
39	M	B1	CRC	49	MSI+	NV	UM	N	P	P	N	P+CR	CR	NP	NP	*MSH2* c.1787A>G; p.Asn596Ser (Class 3)	B
40	M	AP	CRC	61	NP	NP	NP	N	P	P	N	P+CR	NP	NP	NP	-	B
41	F	B5	CRC	70	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B
42	F	B1	EC	50	NP	NP	UM	N	NP	NP	N	P+CR	NP	NP	CR	-	B
43	F	B1	CRC	42	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	CR	-	B
44	F	B1	CRC	49	MSI+	WT	UM	N	P	N	N	P+CR	NP	NP	NP	-	B
45	F	A	EC	60	MSI+	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B, T
46	F	AP	CRC	66	MSI+	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
47	F	B5	CRC	51	MSI+	WT	NP	N	NP	NP	N	P+CR	NP	NP	NP	-	B, N
48	M	B1	CRC	16	MSI+	WT	NP	N	P	P	P	P+CR	NP	NP	NP	-	B
49	F	B2	CRC EC	69 69	MSI+ NP	WT NP	NP NP	N NP	P NP	P NP	N NP	P+CR	NP	NP	NP	-	B
50	M	B1	CRC	41	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B, N, T

Table S1 (Cont.)

Case ID	GENDER	CLINICAL CRITERIA	TUMOR DATA									MMR Mutational Analysis by Sanger Sequencing / SSCP					Available samples
			TUMOR TYPE	AGE AT DIAGNOSIS	MSI	BRAF	MLH1 Methylation	MLH1 IHC	MSH2 IHC	MSH6 IHC	PMS2 IHC	MLH1	MSH2	MSH6	PMS2	Identified VUS	
51	F	AP	EC	56	MSI+	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
52	F	B2	EC	50	MSI+	WT	NP	N	P	P	NP	P+CR	NP	NP	NP	-	B
53	M	B2	CRC	51	MSI+	WT	UM	N	P	P	N						B
			CRC	81	NP	NP	NP	NP	NP	NP	NP						
			CRC	81	NP	NP	NP	NP	NP	NP	NP	NP	NM	NP	NP	-	
54	M	AP	CRC	65	MSI+	WT	UM	N	P	P	N	P+CR	NP	NP	CR	-	B
55	M	B1,3	CRC	47	MSI+	WT	UM	N	P	P	N	P+CR	NP	NP	NP	-	B
56	M	B1	CRC	49	MSI+	WT	UM	N	P	P	N	P+CR	NP	NP	NP	-	B
57	F	A	OC	25	MSI+	NP	NP	P	N	NP	NP	CR	P+CR	P+CR	NP	-	B
58	M	B2	CRC	58	NP	NP	NP	NP	NP	NP	NP						B
			CRC	61	NP	NP	NP	P	N	N	NP	CR	P+CR	P+CR	NP	*MSH2* c.2045C>G; p.Thr682Ser (Class 3*)	
59	M	B1,4	CRC	21	MSI+	NP	NP	P	N	N	P	NP	P+CR	P+CR	NP	-	B
60	M	B2	CRC	58	NP	NP	NP	P	N	N	P	NP	P+CR	PR	NP	-	B
61	M	A	CRC	44	MSI+	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B, N
62	M	B1,B4	CRC	45	MSI+	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
63	F	B1	CRC	31	MSI+	NP	NP	P	N	N	P	NP	P+CR	PR	NP	*MSH2* c.2702A>T; p.Glu901Val (Class 3*)	B, N, T
64	F	B1	CRC	50	MSI+	NP	NP	P	N	N	P						B, N, T,
			CRC	63	NS	NP	NP	NP	NP	NP	NP						
			BC	78	NS	NP	NP	NP	NP	NP	NP	NP	P+CR	P+CR	NP	-	
65	M	B5	CRC	74	MSI+	NP	NP	P	N	N	P	NP	P+CR	P+CR	NP	-	B, N
66	M	A	CRC	58	NP	NP	UM	P	P	N	NP	NP	P+CR	P+CR	NP	-	B, T
67	F	B2,5	CRC	53	MSI+	NP	UM	P	P	N	P	NP	PR	P+CR	NP	*MSH6* c.1153_1155delAGG p.Arg385del (Class 3 *)	B, T
68	M	AP	CRC	85	MSI+	NP	NP	P	P	N	NP	NP	PR	P+CR	NP	-	B
69	M	AP	CRC	56	NP	NP	NP	P	P	N	NP	NP	PR	P+CR	NP	-	B
70	F	B1,4	CRC	46	MSI+	NP	NP	P	P	N	NP	NP	PR	P+CR	NP	MSH6 c.1618_1620delCTT; p.Leu540del (Class 3 *)	B, T
71	F	B1	CRC	36	MSI+	NP	NP	P	P	N	P	CR	P+CR	P+CR	NP	-	B
72	M	AP	CRC	65	MSI+	NP	NP	P	P	N	NP	NP	PR	P+CR	NP	MSH6 c.1450G>A; p.Glu487Lys (Class 3 *)	B
73	F	B2	EC	52	NP	NP	NP	NP	NP	N	NP						B
			CRC	53	NP	NP	NP	NP	NP	NP	NP	NP	PR	P+CR	NP	MSH6 c.3296T>A; p.Ile1099Asn (Class 3 *)	
74	F	A	CRC	56	MSI+	NP	NP	P	P	N	NP	NP	PR	P+CR	NP	-	B
75	F	B1	CRC	45	MSI+	NP	NP	P	P	N	P	NP	P+CR	P+CR	NP	MSH6 c.1618_1620delCTT; p.Leu540del (Class 3 *)	B, N, T
76	F	B1,4	EC	48	MSI+	NP	NP	P	P	N	NP						B
			BC	69	NP	NP	NP	NP	NP	NP	NP	NP	PR	P+CR	NP	-	
77	F	B3	CRC	55	MSS	NP	NP	NP	NP	N	NP	NP	NP	CR	NP	MSH6 c.3226C>T, p.Arg1076Cys (Class 3)	B
78	M	B1	CRC	47	MSI+	WT	NP	p	P	P	N	P+CR	CR	NP	CR	-	B, N
79	M	B5	CRC	59	MSI+	NP	UM	P	P	P	N	PR	NP	NP	CR	-	B
80	F	AP	CRC	57	MSI+	WT	UM	NV	P	P	N	P+CR	NP	NP	CR	-	B, T
81	F	B1	CRC	45	MSI+	WT	NP	NV	P	NV	N	P+CR	CR	CR	CR	-	B
82	F	A	OC	47	MSI+	NP	NP	P	P	P	N	P+CR	CR	NP	CR	-	B
83	M	B3	CRC	58	MSI+	NP	NP	P	P	P	P	CR	CR	NP	NP	-	B, N, T
84	M	AP	CRC	72	MSI+	NP	NP	NP	NP	NP	NP	CR	CR	NP	NP	-	B
85	M	B4	GC	29	MSI+	NP	NP	NP	NP	NP	NP	CR	CR	NP	NP	-	B
86	F	B1,4	CRC	36	MSI+	WT	NP	NV	P	NV	NP	CR	CR	CR	NP	-	B
87	F	B4	CRC	68	MSI+	WT	NP	NV	P	NV	N	NP	NP	NP	CR	-	B
88	M	B5	CRC	65	MSI+	WT	NP	P	P	P	P	CR	CR	NP	NP	-	B, T
89	M	B3	CRC	58	MSI+	WT	NP	P	P	P	P	CR	CR	NP	NP	-	B
90	M	B5	CRC	56	MSI+	NV	NP	P	P	P	P	CR	CR	NP	NP	-	B, T
91	M	AP	CRC	73	MSI+	NP	NP	P	NV	NV	NP	NP	CR	NP	NP	-	B
92	F	A	CRC	46	MSI+	NP	NP	P	P	P	NP	CR	CR	CR	NP	-	B
93	F	A	CRC	65	MSI+	NP	NP	P	P	P	NP	CR	CR	NP	CR	-	B
94	M	B5	CRC	55	MSI+	NP	NP	P	P	P	NP	CR	CR	NP	NP	-	B
95	M	B5	CRC	54	MSI+	NP	NP	P	P	P	NP	CR	CR	NP	NP	-	B
96	F	B1	CRC	48	MSI+	WT	NP	P	P	P	P	CR	CR	NP	NP	-	B
97	F	B5	CRC	65	MSI+	NP	NP	P	P	P	P	NP	NP	CR	NP	-	B
98	F	B5	OC	55	MSI+	NP	NP	P	P	P	P	NP	NP	CR	NP	-	B
#101	F	B3	CRC*	57	NP	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#102	F	B3	CRC*	55	NP	NP	NP	P	N	N	NP	NP	P+CR	PR	NP	-	B

Table S1 (Cont.)

Case ID	GENDER	CLINICAL CRITERIA	TUMOR DATA									MMR Mutational Analysis by Sanger Sequencing / SSCP					Available samples
			TUMOR TYPE	AGE AT DIAGNOSIS	MSI	BRAF	MLH1 Methylation	MLH1 IHC	MSH2 IHC	MSH6 IHC	PMS2 IHC	MLH1	MSH2	MSH6	PMS2	Identified VUS	
#103	F	AP	CRC*	73	NP	NP	NP	P	N	NP	NP	NP	P+CR	P+CR	NP	-	B
#104	F	B3	CRC*	51	NP	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#105	F	B1, 3	CRC*	49	NP	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#107	F	B1	CRC*	39	MSI+	NP	NP	P	N	NV	NP	NP	P+CR	P+CR	NP	-	B, N
#108	F	B1, 2	CRC*	32	MSI+	NP	NP	P	N	NV	NP	NP	P+CR	PR	NP	-	B, T
			CRC	48	MSI-	NP	NP	NV	P	P	NP						
#110	M	B1	CRC*	43	NV	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#111	F	B2	CRC	51	NP	NP	NP	NP	NP	NP	NP	NP	P+CR	P+CR	NP	-	B
			CRC	51	MSI-	NP	NP	NP	NP	NP	NP						
			EC	56	NP	NP	NP	P	N	N	P						
#112	M	B1	CRC*	49	NV	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#113	F	B1	CRC*	49	MSI+	NP	NP	P	N	NV	NP	NP	P+CR	P+CR	NP	-	B, N
#114	M	B2, 3, 5	CRC	58	MSI+	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B, N, T
			CRC	58	NP	NP	NP	NP	NP	NP	NP						
#115	F	A	BC	62	NP	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#116	F	B1	CRC*	48	MSI+	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B, N
#119	F	B4	EC*	45	MSI+	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#121	F	AP	CRC*	77	NP	NP	NP	P	N	N	NP	NP	P+CR	P+CR	NP	-	B, N, T
#123	M	B3	CRC*	59	NP	NP	NP	P	N	N	NP	NP	P+CR	PR	NP	-	B

Table S2-Summary of the samples included in the genome-wide methylation analysis.

Group	Included Males	Included Females	Blood samples	Mean age at blood extraction (years)	FFPE Normal Colon mucosa	Mean age Normal Colon (years)	FFPE CRC	Mean age CRC (years)	Paired Normal/ Tumor
Lynch-like syndrome patients (n=115)	**53**	**62**	**114**	**57±14**	**15**	**52±11**	**25**	**48±12**	**8**
tumors with loss of expression of MLH1/PMS2 proteins	28	29	56	56±16	5	46±7	11	42±11	3
tumors with loss of expression of MSH2/MSH6 proteins	10	17	27	56±15	7	57±13	6	50±17	3
tumors with loss of expression of MSH6 protein	4	8	12	60±11	1	45±0	4	51±6	1
tumors with loss of expression of PMS2 protein	2	3	5	57±6	1	47±0	1	57±0	0
tumors with MSI (conserved expression of MMR proteins)	9	5	14	63±13	1	58±0	3	60±5	1
Lynch syndrome patients (carriers of genetic mutations) (n=61)	**30**	**31**	**61**	**53±8**	**17**	**50±10**	**21**	**50±9**	**14**
MLH1 mutation carriers	9	12	21	49±8	6	47±8	8	46±7	6
MSH2 mutation carriers	10	18	28	47±9	5	40±9	7	43±10	4
MSH6 mutation carriers	8	1	9	55±14	3	50±12	3	48±6	1
PMS2 mutation carriers	3	0	3	64±10	3	63±13	3	63±13	3
Constitutional *MLH1* Epimutation carriers (n=12)	**6**	**6**	**11**	**59±12**	**4**	**35±7**	**6**	**42±8**	**3**
Healthy controls (n=41)	**12**	**29**	**41**	**51±11**	**0**	**NA**	**0**	**NA**	**0**

Table S3- Primers and conditions used at current study

Validation of *bona fide* pathogenic MMR variants found by Haloplex custom panel					
Gene	**Change**	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Amplicon size**	**Annealing temperature**
MLH1	NM_000249.3:c.676C>T	GTTTCAGTCTCAGCCATGAG	ACACATGATTCACGCCACAG	376	55
MSH6	NM_000179.2:c.2219T>A	GAGGCACGATGTAGAAAGATGGCA	TCCTGGTGTCAACCCAATGGAA	582	60
MSH6	NM_000179.2:c.3150_3161dup	TGGGATACAGCCTTTGACCATGA	CTTAAATTGCTGTGGGCAGCCT	633	60
Validation of *in-silico* predicted damaging variants found by Haloplex custom panel					
Gene	**Change**	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Amplicon size**	**Annealing temperature**
MSH2	NM_000251.2:c.2702A>T	ATGTGTGATATGTTTAGATGGAA	GCACTGACAGTTAACACTATGGA	285	50
MSH6	NM_000179.2:c.2092C>G	CCATTGGGTTGACACCAGGAGA	TTGAATCCTTCCAGAGCAGAAAGA	595	60
MSH3	NM_002439.4:c.685T>C	GTTAGCTTTTTGCCAGATTTGC	TAAAATAGTGCCTGAAAAGAC	333	56
MSH3	NM_002439.4:c.1862T>C	GAAGGAGGAGTTTCCTTTGT	AGAACACTGTCAGCTTTAATAG	295	56
MSH3	NM_002439.4:c.2732T>C	TCACACAGTTCAGGTTGAAG	CTCCTAAATGTTGAGTGCTTT	499	58
MSH3	NM_002439.4:c.3072G>C	GTTGTACTTTTCTTGTGACT	CCAACAAACTTTGAGTTATCA	317	55
APC	NM_000038.5:c.1959G>A	GTTACTGCATACACATTGTGAC	ACTTCTATCTTTTTCAGAACGAG	649	57
APC	NM_000038.4:c.1966C>G	GTTACTGCATACACATTGTGAC	ACTTCTATCTTTTTCAGAACGAG	649	57
APC	NM_000038.4:c.3173A>G	AGTCTTAAATATTCAGATGAGCAG	AATTCCATGATTAGAACCCAC	282	60
APC	NM_000038.4:c.7514G>A	TCCACACATTCGTCTGTTCA	CTCACCCAAACATCCTCTGTT	500	55
APC	NM_000038.4:c.7936C>G	TCCACACATTCGTCTGTTCA	CTCACCCAAACATCCTCTGTT	500	55
POLD1	NM_002691.3:c.2275G>A	GCTTCACTCCGCATGATTCT	ATGAGGGCACCTAAATGCAG	527	55
BUB1	NM_004336.4:c.2473C>T	CCACATTGCAGCAACAGTTC	TGTAGAATTCCCAGGGGTTG	167	56
PMS1	NM_000534.4:c.497A>C	GACGTTCCTTCCAAATCTAAATG	GTCATAGCCCATATCTAACTGATTT	300	61
EPCAM	NM_002354.2:c.811G>T	TTCCTTTTCTCCTTTTCAATACA	CGCCCAGCCACTATTACTTT	295	55
RNA analysis					
Gene	**Change**	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Amplicon size**	**Annealing temperature**
MSH2	c.1787A>G_*splicing*	TCAGTCTCTGGCTGCCTTG	AGCCCCTACTCGGGCTAAG	1140	64
MSH2	exon 8 duplication_*splicing*	TCAGTCTCTGGCTGCCTTG	AGCCCCTACTCGGGCTAAG	1140	64
MSH6	c.1153_1155delAGG_*splicing*	ACTGAGAGCAATGCAACGTG	CAGGAAAATGGCAAAGCCTA	2868	64
MSH6	c.1618_1620delCTT_*splicing*	ACTGAGAGCAATGCAACGTG	CAGGAAAATGGCAAAGCCTA	2868	64
MSH6	c.3150_3161dup_*splicing*	ACTGAGAGCAATGCAACGTG	CAGGAAAATGGCAAAGCCTA	2868	64
MSH6	c.1153_1155delAGG_*ASE*	GAAGAGATGAGCACAGGAGG	_	_	_
MSH6	c.1618_1620delCTT_*ASE*	CCTCTTTTTCTTTGAGGCTAAGA	_	_	_
MSH6	c.3150_3161dup_*ASE at c.2633T>C*	AAAATTATAGGGATCATGGAAGAAG	_	_	_
EMAST analysis					
Gene	**Location**	**Forward primer (5'-3')**	**Reverse primer (5'-3')**	**Amplicon size**	**Annealing temperature**
D20S82	20p12.3	FAM-GCCTTGATCACACCACTACA	GTGGTCACTAAAGTTTCTGCT	246–270bp	61
D21S1436	21q21.1	PET-AGGAAAGAGAAAGAAAGGAAGG	TATATGATGAAAGTATATTGGGGG	appr.178bp	58
UT5037	Chr.8	NED-TTCCTGTGAACCATTAGGTCA	GGGAGACAGAGCAAGACTC	appr.145bp	60
D2S443	2p13.2–2p13.1	NED-GAGAGGGCAAGACTTGGAAG	ATGGAAGAGCGTTCTAAAACA	appr.251bp	58
D9S747	9q32	FAM-GCCATTATTGACTCTGGAAAAGAC	CAGGCTCTCAAAATATGAACAAAAT	182-202bp	56
MYCL1	1p34.1	FAM-TGGCGAGACTCCATCAAAG	CTTTTTAAGCTGCAACAATTT	appr.150bp	53

Table S4. Detailed *in silico* predictions of MMR class 3 variants identified in this study. Abbreviations: NP: not performed; NA: not available; DSS=Consensus Donor Splice Site; DSS*=Alternative Donor Splice Site; ASS=Consensus Aceptor Splice Site; ASS*=Alternative Aceptor Splice Site; D=Damaging; PrD=Probably Damaging; PsD=Possibly Damaging; T=Tolerated. Gain or Loss of Splice sites are considered when 4 of the 5 predictors are in agreement of their calculation. Inconclusive interpretation is given when 3 of the 5 predictors predicted changes. Less than 3 similar predictions are considered as no changes.

Case ID	Variant calling				Splice Site Prediction												Protein function Prediction					
	Gene	cDNA change	Predicted protein change	rs ID/ExAC/ESP	Splice Site	SSF [0-100]		MaxEnt [0-16]		NNSPLICE [0-1]		GeneSplicer [0-21]		HSF [0-100]		Interpretation	SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	Interpretation
						wildtype	variant	wildtype	variant	wildtype	variant	wildtype	variant	wildtype	variant							
35	*MLH1*	c.25C>T	p.Arg9Trp	rs587779000/NR/NR	DSS	75,68	=	8.60	=	0.93	=	5.52	=	84,51	=	No changes	D (0)	D (1)	PrD (1.000)	PrD (0.973)	D (-4.59)	**Damaging**
					ASS																	
					ASS*	81,95	=	5,85	=	0,51	0.54 (+5.8%)	2,2	1.77 (-19.4%)	86,97	=							
39	*MSH2*	c.1787A>G	p.Asn596Ser	rs41295288/NR/0.0002	DSS											No changes	T (0.63)	D (1)	B (0.012)	B (0.059)	N (-2.16)	Benign
					ASS	98,45	=	9,05	=	0,97	=	10,47	10.50 (+0.2%)	90,8	=							
58	*MSH2*	c.2045C>G	p.Thr682Ser	NR/NR/NR	DSS											No changes	T (0.06)	D (1)	B (0.240)	B (0.216)	N (-1.59)	Benign
					ASS	87,38	=	8,23	=	0,95	=	3,34	3.21 (-3.8%)	86,07	=							
63	*MSH2*	c.2702A>T	p.Glu901Val	NR/NR/NR	DSS											No changes	D (0.04)	D (1)	PsD (0.911)	B (0.408)	D (-4.11)	**Damaging**
					DSS*	71,87	77.20 (+7.4%)	—	1,93					80,82	83.15 (+2.9%)							
					ASS	75,61	=	6,11	=					82,26	=							
98	*MSH2*	c.2802G>A	P.Thr934=	rs150259097/0.000/0.001	DSS	_		_		_		_		_		No changes	NA	NA	NA	NA	NA	NA
					ASS	_		_		_		_		_								
67	*MSH6*	c.1153_1155delAGG	p.Arg385del	rs267608043/NR/NR	DSS	_		_		_		_		_		No changes	NP	NP	NP	NP	D (-9.02)	Benign
					ASS	_		_		_		_		_								
72	*MSH6*	c.1450G>A	p.Glu487Lys	NR/NR/NR	DSS	_		_		_		_		_		No changes	D (0)	D (1)	PrD (1.000)	PrD (1.000)	D (-3.93)	Damaging
					ASS	_		_		_		_		_								
70	*MSH6*	c.1618_1620delCTT	p.Leu540del	NR/NR/NR	DSS	_		_		_		_		_		No changes	NP	NP	NP	NP	D (-10.18)	Benign
75	*MSH6*	c.1618_1620delCTT			ASS	_		_		_		_		_								
5	*MSH6*	c.2092C>G	p.Gln698Glu	rs63750832/NR/NR	DSS											Inconclusive	T (1)	D (1)	B (0.205)	B (0.098)	N (-1.34)	Benign
					ASS																	
					ASS*	80,66	—	4	—					86,02	75.73 (-12.0%)						
82	*MSH6*	c.3150_3161dup	p.(Val1051_Ile1054dup)	NR/NR/NR	DSS	81,58	=	8,91	=	0,86	=			85,86	90.83 (+5.8%)	No changes	NP	NP	NP	NP	D (-11.89)	**Damaging**
					ASS																	
77	*MSH6*	c.3226C>T	p.Arg1076Cys	rs63750617/NR/NR	DSS											No changes	D (0)	D (1)	PrD (0.991)	PsD (0.679)	D (-3.95)	**Damaging**
					ASS	87,78	=	12,58	=	0,98	=	10,42	9.81 (-5.9%)	89,11	=							
					ASS*	75,52	77.03 (+2.0%)	2,21	2.80 (+26.6%)					83,19	82.77 (-0.5%)							
73	*MSH6*	c.3296T>A	p.Ile1099Asn	NR/NR/NR	DSS											No changes	D (0)	D (1)	PsD (0.860)	PsD (0.797)	D (-4.54)	**Damaging**
					ASS																	
					ASS*			1,66	0.56 (-66.3%)					73,09	71.09 (-2.7%)							
85	*PMS2*	c.1320A>G	p.Pro440Pro	rs138697590/NR/0.000	DSS	_		_		_		_				No changes	NA	NA	NA	NA	NA	NA
					ASS	_		_		_		_										

Table S5. Variants identified in the mutational analysis of CRC-predisposing genes of 42 LLS cases. Pathogenic mutations and predicted pathogenic mutations are highlighted in bold. NP: not performed; NA: not available.

Case ID	Variant calling				Coverage		rs ID	MAF	In silico predictions						ClinVar Classification
									Splicing	Protein function					
	Gene	Transcript	cDNA change	Predicted protein change	Allele Frequency	Read Depth		ExAC/ESP		SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	
5	*MSH6*	NM_000179.2	c.2092C>G	p.Gln698Glu	0.427	5024	rs63750832	NR/NR	Inconclusive	Tolerated (1)	Disease causing (1)	Benign (0.205)	Benign (0.098)	NP	VUS (5) /+++
	ENG	NM_000118.3	c.1539G>A	p.Lys513=	0.474	3696	NA	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
8	*EPCAM*	NM_002354.2	c.345G>A	p.Met115Ile	0.511	1307	rs115212523	0.001/NR	No changes	Deleterious (0.003)	Polymorphism (1)	Benign (0.013)	Benign (0.016)	NP	Not reported
	MSH3	NM_002439.4	c.1160T>A	p.Phe387Tyr	0.523	1265	rs140543135	0.000/0.001	No changes	Tolerated (0.48)	Polymorphism (1)	Benign (0.408)	Benign (0.259)	NP	Not reported
	POLD1	NM_002691.3	c.971-4G>A	p.?	0.523	7307	rs200144991	0.000/NR	No changes	NA	NA	NA	NA	NA	Benign/Likely benign (5) / ++
	POLD1	NM_002691.3	c.2953+8G>T	p.?	0.474	4182	rs777560428	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
10	***EPCAM***	**NM_002354.2**	**c.811G>T**	**p.Val271Phe**	0.5	2358	NA	NR/NR	No changes	**Deleterious (0)**	**Disease causing (1)**	**Probably damaging (1.000)**	**Probably damaging (0.989)**	NP	Not reported
	POLE	NM_006231.3	c.4259C>T	p.Ala1420Val	0.573	1054	rs41561818	0.003/0.002	No changes	Tolerated (0.32)	Disease causing (1)	Benign (0.006)	Benign (0.003)	NP	Benign/Likely benign (4) / ++
	APC	NM_000038.5	c.2640C>T	p.Ile880=	0.439	2680	rs200184105	0.000/NR	No changes	NA	NA	NA	NA	NA	Likely benign (6) / ++
13	*EXO1*	NM_130398.3	c.1828A>G	p.Ser610Gly	0.462	7454	rs12122770	0.002/0.003	No changes	Tolerated (0.27)	Disease causing (0.739)	Benign (0.004)	Benign (0.004)	NP	Not reported
	SETD2	NM_014159.6	c.2251C>A	p.Pro751Thr	0.485	3223	rs115788094	0.000/0.001	No changes	Tolerated (0.32)	Polymorphism (0.867)	Benign (0,220)	Benign (0.024)	NP	Likely benign(1) vs VUS (1) / +
28	*MLH3*	NM_001040108.1	c.3315C>A	p.Asp1105Glu	0.474	5490	rs28757008	0.003/0.003	No changes	Deleterious (0)	Polymorphism (0.839)	Benign (0.060)	Benign (0.031)	NP	Benign (1) / +
	FAN1	NM_014967.4	c.783G>A	p.Ala261=	0.415	3101	rs142437586	0.026/0.000	No changes	NA	NA	NA	NA	NA	Not reported
	MLH3	NM_001040108.1	c.2940C>T	p.Thr980=	0.446	3285	rs750994816	NR/NR	No changes	NA	NA	NA	NA	NA	Likely benign (1) / +
29	***POLD1***	**NM_002691.3**	**c.2275G>A**	**p.Val759Ile**	0.494	4773	rs145473716	0.002/0.001	No changes	**Deleterious (0)**	**Disease causing (1)**	**Probably damaging (1.000)**	**Probably damaging (0.988)**	NP	Benign/Likely benign(6) vs VUS (1) / +
30	***APC***	**NM_000038.4**	**c.7936C>G**	**p.Gln2646Glu**	0.506	2384	NA	NR/NR	Inconclusive	**Deleterious (0.02)**	**Disease causing (1)**	**Possibly damaging (0.688)**	Benign (0.182)	NP	VUS (1) / +
33	***MLH1***	**NM_000249.3**	**c.676C>T**	**p.Arg226Ter**	0.273	43	rs63751615	0.000004/NR	No changes	NA	NA	NA	NA	NA	Pathogenic (9) / +++
	EPCAM	NM_002354.2	c.345G>A	p.Met115Ile	0.0629	349	rs115212523	0.001/NR	No changes	Deleterious (0.003)	Polymorphism (1)	Benign (0.013)	Benign (0.016)	NP	Not reported
	POLD1	NM_002691.3	c.971-4G>A	p.?	0.0656	2057	rs200144991	0.000/NR	No changes	NA	NA	NA	NA	NA	Not reported
	POLD1	NM_002691.3	c.2953+8G>T	p.?	0.0556	1223	rs777560428	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
39	***FAN1***	**NM_014967.4**	**c.149T>G**	**p.Met50Arg**	0.465	2417	rs148404807	0.002/0.0037	No changes	Tolerated (0.08)	**Disease causing (1)**	**Probably damaging (0.991)**	**Possibly damaging (0.690)**	NP	Not reported
	MSH2	NM_000251.2	c.1787A>G	p.Asn596Ser	0.437	3103	rs41295288	NR/NR	No changes	Tolerated (0.63)	Disease causing (1)	Benign (0.012)	Benign (0.059)	NP	VUS (3) vs Likely benign (3) /+++
	APC	NM_000038.5	c.1959G>A	p.Arg653=	0.471	1580	rs72541809	0.004/0.003	No changes	NA	NA	NA	NA	NA	Benign/Likely benign (11) / ++
44	*MSH3*	NM_002439.4	c.1862T>C	p.Ile621Thr	0.433	990	rs750694528	NR/NR	Inconclusive	Deleterious (0.03)	Polymorphism (0.821)	Benign (0.053)	Benign (0.082)	NP	Not reported
45	*AXIN2*	NM_004655.3	c.1985T>C	p.Leu662Pro	0.49	6827	rs142476324	0.001/0.001	No changes	Tolerated (0.1)	Disease causing (1)	Benign (0.009)	Benign (0.004)	NP	Likely benign(3) vs VUS (1) / ++
48	*SMAD4*	NM_005359.5	c.582A>G	p.Thr194=	0.439	5771	rs145805120	0.000/0.000	No changes	NA	NA	NA	NA	NA	Benign/Likely benign (6) / ++
53	*POLE*	NM_006231.3	c.898A>G	p.Ile300Val	0.498	7250	NA	NR/NR	No changes	Tolerated (0.26)	Disease causing (1)	Benign (0.003)	Benign (0.016)	NP	Not reported
	BUB3	NM_004725.3	c.*371A>G	p.?	0.366	1735	rs527569482	NR/NR	NA	NA	NA	NA	NA	NA	Not reported
55	***PMS1***	**NM_000534.4**	**c.497A>C**	**p.Lys166Thr**	0.472	323	NA	NR/NR	No changes	**Deleterious (0)**	**Disease causing (1)**	**Possibly damaging (0.757)**	**Possibly damaging (0.599)**	NP	Not reported
	POLE	NM_006231.3	c.3925G>A	p.Asp1309Asn	0.475	4354	NA	NR/NR	No changes	Tolerated (0.24)	Disease causing (0.996)	Benign (0.000)	Benign (0.001)	NP	VUS (1) / +
57	*FAN1*	NM_014967.4	c.1520G>A	p.Arg507His	0.408	1979	rs150393409	0.006/0.005	No changes	Tolerated (0.06)	Disease causing (0.996)	Possibly damaging (0.765)	Benign (0,052)	NP	Not reported
58	***EXO1***	**NM_130398.3**	**c.2212-1G>A**	**p.?**	0.447	690	rs4150000	0.0019/0.0028	**Loss of an ASS**	NA	NA	NA	NA	NA	Not reported
	MSH2	NM_000251.2	c.2045C>G	p.Thr682Ser	0.447	2651	rs587779130	NR/NR	No changes	Tolerated (0.06)	Disease causing (1)	Benign (0.240)	Benign (0.216)	NP	VUS (1) / +++
	FAN1	NM_014967.4	c.603C>T	p.Asp201=	0.489	3301	rs142084532	0.001/0.001	No changes	NA	NA	NA	NA	NA	Not reported
	SETD2	NM_014159.6	c.1404A>G	p.Thr468=	0.444	2718	rs368082274	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
59	***APC***	**NM_000038.4**	**c.1966C>G**	**p.Leu656Val**	0.471	1617	rs577466163	NR/NR	**Gain of a donor splicing site**	**Deleterious (0)**	**Disease causing (1)**	**Probably damaging (0.999)**	**Probably damaging (0.998)**	NP	VUS (2) / ++
60	*AXIN2*	NM_004655.3	c.1985T>C	p.Leu662Pro	0.495	10201	rs142476324	0.001/0.001	No changes	Tolerated (0.1)	Disease causing (1)	Benign (0.009)	Benign (0.004)	NP	Likely benign (3) vs VUS (1) / ++
	AXIN2	NM_004655.3	c.1083G>A	p.Glu361=	0.495	14191	NA	0.000/NR	No changes	NA	NA	NA	NA	NA	Likely (1) / +
62	*CDH1*	NM_004360.4	c.88C>A	p.Pro30Thr	0.496	2288	rs139866691	0.001/0.001	No changes	Tolerated (0.08)	Polymorphism (1)	Probably damaging (0.986)	Possibly damaging (0.886)	NP	Benign/Likely benign(10) vs VUS (2) / +
	MSH3	**NM_002439.4**	**c.1862T>C**	**p.Leu911Trp**	0.464	4684	rs41545019	0.002/0.004	No changes	**Deleterious (0)**	**Disease causing (0.999)**	**Probably damaging (1.000)**	**Probably damaging (0.978)**	NP	Not reported
63	*APC*	NM_000038.4	c.3460_3462delGAA	p.Glu1157del	0.543	1290	rs567584401		No changes	NA	NA	NA	NA	Neutral	Benign/Likely benign(11) vs VUS (1) / +
	MSH2	**NM_000251.2**	**c.2702A>T**	**p.Glu901Val**	0.51	1937	NA	NR/NR	No changes	**Deleterious (0.04)**	**Disease causing (1)**	**Possibly damaging (0.911)**	Benign (0.408)	NP	Not reported
	SETD1B	NM_015048.1	c.23A>C	p.His8Pro	0.125	2435	rs773794483	NR/NR	No changes	Tolerated (0.28)	Polymorphism (0.505)	Benign (0.000)	Benign (0.000)	NP	Not reported
65	*MUTYH*	NM_001128425.1	c.1437_1439delGGA	p.Glu480del	0.43	1365	rs587778541	NR/0,000	No changes	NA	NA	NA	NA	**Deleterious**	Pathogenic (10) / ++
	BMPR1A	NM_004329.2	c.1560G>A	p.Thr520=	0.473	2268	rs142775086	0.000/0.002	No changes	NA	NA	NA	NA	NA	Benign/Likely benign (7) / ++
	ENG	NM_000118.3	c.585G>A	p.Glu195=	0.456	2883	NA	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
74	***MSH3***	**NM_002439.4**	**c.685T>C**	**p.Tyr229His**	0.476	4670	NA	NR/NR	No changes	**Deleterious (0.01)**	**Disease causing (0.999)**	**Probably damaging (1.000)**	**Probably damaging (0.973)**	NP	Not reported
	MSH3	**NM_002439.4**	**c.2732T>G**	**p.Leu911Trp**	0.447	5498	rs41545019	0.002/0.004	No changes	**Deleterious (0)**	**Disease causing (0.999)**	**Probably damaging (1.000)**	**Probably damaging (0.978)**	NP	Not reported

Table S6. Detailed in silico predictions of rare germline variants in other CRC-predisposing genes identified in this study. Abbreviations: NP: not performed; NA: not available; DSS=Consensus Donor Splice Site; DSS*=Alternative Donor Splice Site; ASS=Consensus Aceptor Splice Site; ASS*=Alternative Aceptor Splice Site; D=Damaging; PrD=Probably Damaging; PsD=Possibly Damaging; T=Tolerated. Gain or Loss of Splice sites are considered when 4 of the 5 predictors are in agreement of their calculation. Inconclusive interpretation is given when 3 of the 5 predictors predicted changes. Less than 3 similar predictions are considered as no changes.

Case ID	Variant calling				Splice Site Prediction													Protein function					
	Gene	cDNA change	Predicted protein change	rs ID/ExAC/ESP	Splice Site	SSF [0-100] wildtype	SSF [0-100] variant	MaxEnt [0-16] wildtype	MaxEnt [0-16] variant	NNSPLICE [0-1] wildtype	NNSPLICE [0-1] variant	GeneSplicer [0-21] wildtype	GeneSplicer [0-21] variant	HSF [0-100] wildtype	HSF [0-100] variant	Interpretation	SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	Interpretation
10	*EPCAM*	c.811G>T	p.Val271Phe	NR/NR/NR	DSS	87,13	=	10,1	=	0,99	=	1,26	0,75 (-40,5%)	94,09		No changes	D (0)	D (1)	PrD (1.000)	PrD (0.989)	NP	Damaging
					ASS																	
29	*POLD1*	c.2275G>A	p.Val759Ile	rs145473716/0.002/0.001	DSS											No changes	D (0)	D (1)	PrD (1.000)	PrD (0.988)	NP	Damaging
					ASS	76,26	=	9,25	=	0,95	=	14,26	14,00 (-1,8%)	88,88								
30	*APC*	c.7936C>G	p.Gln2646Glu	NR/NR/NR	DSS	NADA										No changes	D (0.02)	D (1)	PsD (0.688)	B (0.182)	NP	Damaging
					ASS	NADA	8828	NA	0.001/NR	No changes		Tolerated (0.06)		Polymorphism (1)		Benign (0.133)						
39	*FAN1*	c.149T>G	p.Met50Arg	rs148404807/0.002/0.0037	DSS	508	1865	rs377041426	0.016/0.000	No changes		NA		NA		NA	T (0.08)	D (1)	PrD (0.991)	PsD (0.690)	NP	Damaging
					ASS											No changes						
55	*PMS1*	c.497A>C	p.Lys166Thr	NR/NR/NR	DSS	474	3567	rs748392521	NR/NR	No changes	0.98	=	Deleterious (0)		Disease causing (1)		Probably damaging (1.000)					Damaging
					ASS	81,76	=	9,24	=	0,91	=	5,81	6,19 (+6,4%)	89,48		No changes	D (0)	D (1)	PsD (0.757)	PsD (0.599)	NP	
					ASS*	428	1607	NA	—	0,09	NA	NA	NA	66,74	76,13 (+14,1%)							
58	*EXO1*	c.2212-1G>C	p.Val738_Lys743del	rs415000/0.0019/0.0028	DSS											Loss of an ASS	D (0.019)	NA	NP	NA	NP	NA
					ASS	91,77	—	8,44	—	0,62	—			93,10	—							
59	*APC*	c.1966C>G	p.Leu656Val	rs577466163/NR/NR	DSS	196	1407	rs147640999	0.000/0.000	Loss of a DSS / Inconclusive at ASS		Tolerated (0.39)		Polymorphism (0.996)		Benign (0.007)						
					DSS*	—	72,54	—	6,29	—	0,71			—	80,95	Gain of a DSS	D (0)	D (1)	PrD(0.999)	PrD (0.998)	NP	Damaging
					ASS	93,56	3059	rs138697590	NR/0.000	0,81	0,87 (+7,0%)	5,87	6,02 (+2,5%)	91,31								
65	*MUTYH*	c.1437_1439delGGA	p.Glu480del	rs587778541/NR/0.000	DSS	42	4486	rs556549304	NR/NR	No changes		NA		NA		No changes	NA	NA	NA	NA	D (-7.78)	Damaging
					DSS*	373	3385	rs745483465	NR/NR	No changes	0,56	NA	3,06 (-33,6%)	75,54								
					ASS*	51	3424	rs373486660	NR/NR	No changes		Tolerated (0.44)		Disease causing (0.912)		Benign (0.125)						
					ASS*	29	850	rs900010840	NR/NR	NA		NA		NA								
					ASS*	49	2284	rs[illegible]	[illegible]	No changes		NA		NA								
62	*MSH3*	c.2732T>G	p.Leu911Trp	rs415450019/0.002/0.004	DSS	99,69	=	11	=	1	=	4,61	=	99,04		No changes	D (0)	D (0.999)	PrD (1.000)	PrD (0.978)	NP	Damaging
74					ASS	95,37	=	8,44	=	0,94	=	6,12	6,33 (+3,4%)	89,93								
					ASS*	70,86	=	4,17	2,56 (-38,9%)					81,32								
74	*MSH3*	c.685T>C	p.Tyr229His	NR/NR/NR	DSS	NADA	2043	rs1801166	0.004/0.009	No changes		Tolerated (0.49)		Disease causing (0.995)		Benign (0.001)						
					ASS	NADA										No changes	D (0.01)	D (0.999)	PrD (1.000)	PrD (0.973)	NP	Damaging
81	*BUB1*	c.2473C>T	p.Pro825Ser	rs748392521/NR/NR	DSS	419	269	rs34506289	NR/0.010	No changes		NA		NA		NA						
					ASS	476	1064	rs150259097	0.000/0.001	No changes		NA		NA		No changes	D (0)	D (1)	PrD (1.000)	PrD (0.997)	NP	Damaging
85	*MSH3*	c.3072G>C	p.Gln1024His	rs147640909/0.000/0.000	DSS	86,44	=	8,20	=	0,77	0,80 (+4,0%)	5,58	5,54 (-0,8%)	89,5	=							
					DSS*	84,5	=	9,46	=	0,98	=	1,52	2,21 (+45,4%)	88,86	=	Loss of a DSS / Inconclusive loss of an ASS	T (0.39)	P (0.996)	B (0.007)	B (0.013)	NP	Benign
					ASS	80,54	—	6,92	—	0,75	—			88,57	77,55 (-12,4%)							
					ASS*	87,76	=	10,22	=	0,95	=	6,83	7,40 (+8,3%)	85,23	=							
96	*APC*	c.7514G>A	p.Arg2505Gln	rs147549623/0.001/0.001	DSS	79,71	—	3,82	—					89,25	—	No changes	D (0.04)	D (1)	PrD (1.000)	PrD (0.961)	NP	Damaging
					ASS																	

Coverage | MAF | rs ID | Allele Frequency | Read Depth | ExAC/ESP | Splicing | In silico predictions | Protein function

Table S7: Differentially methylated CpG islands **(A)** and CpG sites **(B)** found by methylome analysis of blood DNA from LLS cases against controls (FDR p-value<0.05).

A) Differentially methylated CpG islands

Chromosome	Coordinates	Comparison Group	Mean difference	p.value	FDR p.value	Total sites	UCSC
chr3	37034229-37035356	LLS-MLH1vsMC	-0.383121502	1.39E-73	3.55E-69	30	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LLS-NOLOSSvsMC	-0.383140078	3.98E-31	1.02E-26	30	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LLSvsMC	-0.383164698	1.02E-119	2.60E-115	30	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LS-MLH1vsMC	-0.383133504	2.51E-39	6.41E-35	30	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LSvsMC	-0.383109405	1.35E-79	3.45E-75	30	*EPM2AIP1-MLH1*
chr3	37034229-37035356	MCvsHealthy	0.383091852	9.59E-61	2.45E-56	30	*EPM2AIP1-MLH1*

B) Differentially methylated CpG sites found in LLS cases (upper rows) and LS controls (lower rows). Significative CpG sites located at the differentially methylated region previously described for MLH1 epimutation carriers have been removed to facilitate its interpretation.

CG ID	Chromosome	Start	Comparison Group	Mean difference	p.value	FDR p.value	Gene
cg06806862	chr3	184934396	LLS-MLH1vsMC	0.1021287	5.58E-06	0.04913797	*EHHADH*
cg06806862	chr3	184934396	LLSvsMC	0.09807676	5.26E-07	0.00516911	*EHHADH*
cg04565255	chr13	77565759	LLSvsMC	0.09494056	1.33E-06	0.01249859	*CLN5*
cg04565255	chr13	77565759	LLS-MLH1vsMC	0.09914193	4.64E-06	0.04270578	*CLN5*
cg03901784	chr13	103451536	LLSvsMC	-0.0147134	1.21E-08	0.00012169	*BIVM*
cg03901784	chr13	103451536	LLS-MLH1vsMC	-0.01502137	3.24E-07	0.00318417	*BIVM*
cg23961842	chr6	85483800	LLSvsMC	-0.01816033	2.59E-06	0.02378466	*TBX18*
cg23961842	chr6	85483800	LLS-MLH1vsMC	-0.02055483	1.22E-08	0.00012301	*TBX18*
cg08948338	chr6	32936102	LLSvsMC	-0.0368752	3.63E-06	0.03197485	*BRD2*
cg10436026	chr13	37453429	LLS-MSH2vsLS-MSH2	0.07689572	1.91E-07	0.04265141	*SMAD9*
cg24760577	chr19	45201793	LLSvsMC	0.05556465	3.03E-06	0.02723409	*171A8.1*
cg04521543	chr2	669505	LLS-MSH2vsLS-MSH2	0.04809198	2.02E-07	0.04265141	*TMEM18*
cg01836455	chr6	73973719	LLS-MLH1vsMC	**0.28227418**	5.13E-06	0.04614998	*KHDC1*
cg14755019	chr4	41749810	LS-MLH1vsMC	0.07434478	1.90E-06	0.01916646	*PHOX2B*
cg23944298	chr3	127248864	LS-MSH2vsSANO	-0.07036139	9.76E-09	0.00412677	*BX537548*
cg23944298	chr3	127248864	LS-MSH6vsSANO	-0.10034633	9.88E-08	0.04180574	*BX537548*

Table S8: Differentially methylated CpG islands **(A)** and CpG sites **(B)** found by methylome analysis of normal colon mucosa DNA from LLS cases against controls (FDR p-value<0.05).

A) Differentially methylated CpG islands

Chromosome	Coordinates	Comparison Group	Mean difference	p.value	FDR p.value	Num sites	Gene
chr3	37034229-37035356	LLS-MLH1vsMC	-0.413333422	2.81E-10	7.12E-06	22	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LLSvsMC	-0.413333422	2.81E-10	7.12E-06	22	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LS-MLH1vsMC	-0.423811959	5.75E-12	1.46E-07	22	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LSvsMC	-0.423811959	5.75E-12	1.46E-07	22	*EPM2AIP1-MLH1*

B) Differentially methylated CpG sites

CG ID	Chromosome	Start	Comparison Group	Mean difference	p.value	FDR p.value	Gene
cg02279071	*chr3*	*37034154*	*LLS-MLH1vsMC*	*-0,360782456*	*1,0476E-07*	*0,00217751*	*EPM2AIP1-MLH1*
cg14751544	*chr3*	*37034166*	*LLS-MLH1vsMC*	*-0,400209867*	*1,343E-07*	*0,00264453*	*EPM2AIP1-MLH1*
cg16764580	*chr3*	*37034346*	*LLS-MLH1vsMC*	*-0,425835333*	*2,5295E-09*	*0,0001352*	*EPM2AIP1-MLH1*
cg01302270	*chr3*	*37034441*	*LLS-MLH1vsMC*	*-0,597660269*	*1,9095E-10*	*2,3814E-05*	*EPM2AIP1-MLH1*
cg17641046	*chr3*	*37034473*	*LLS-MLH1vsMC*	*-0,593696318*	*1,4669E-09*	*0,00010409*	*EPM2AIP1-MLH1*
cg07101782	*chr3*	*37034495*	*LLS-MLH1vsMC*	*-0,557054297*	*4,3468E-11*	*8,1315E-06*	*EPM2AIP1-MLH1*
cg10769891	*chr3*	*37034730*	*LLS-MLH1vsMC*	*-0,458920768*	*5,3222E-09*	*0,00023309*	*EPM2AIP1-MLH1*
cg23658326	*chr3*	*37034787*	*LLS-MLH1vsMC*	*-0,403278702*	*1,6692E-09*	*0,00010409*	*EPM2AIP1-MLH1*
cg11600697	*chr3*	*37034814*	*LLS-MLH1vsMC*	*-0,428970019*	*2,9372E-08*	*0,00091577*	*EPM2AIP1-MLH1*
cg21490561	*chr3*	*37034825*	*LLS-MLH1vsMC*	*-0,494376267*	*2,2595E-07*	*0,0042268*	*EPM2AIP1-MLH1*
cg00893636	*chr3*	*37034840*	*LLS-MLH1vsMC*	*-0,446345427*	*5,3118E-07*	*0,00864066*	*EPM2AIP1-MLH1*
cg06791151	*chr3*	*37034956*	*LLS-MLH1vsMC*	*-0,303127322*	*5,1985E-08*	*0,00149613*	*EPM2AIP1-MLH1*
cg24985459	*chr3*	*37035090*	*LLS-MLH1vsMC*	*-0,397349923*	*5,6069E-09*	*0,00023309*	*EPM2AIP1-MLH1*
cg12790037	*chr3*	*37035117*	*LLS-MLH1vsMC*	*-0,425614612*	*3,7725E-07*	*0,00641571*	*EPM2AIP1-MLH1*
cg17621259	*chr3*	*37035168*	*LLS-MLH1vsMC*	*-0,366040011*	*8,862E-08*	*0,00207227*	*EPM2AIP1-MLH1*
cg05906740	*chr3*	*37035205*	*LLS-MLH1vsMC*	*-0,266285456*	*1,6468E-08*	*0,00056013*	*EPM2AIP1-MLH1*
cg27331401	*chr3*	*37035207*	*LLS-MLH1vsMC*	*-0,411753569*	*2,4069E-07*	*0,00428822*	*EPM2AIP1-MLH1*
cg25837710	*chr3*	*37035220*	*LLS-MLH1vsMC*	*-0,295424543*	*9,534E-09*	*0,00035671*	*EPM2AIP1-MLH1*
cg12851504	*chr3*	*37035222*	*LLS-MLH1vsMC*	*-0,283628981*	*9,5791E-08*	*0,00210819*	*EPM2AIP1-MLH1*
cg06590608	*chr3*	*37035228*	*LLS-MLH1vsMC*	*-0,396755002*	*8,5926E-07*	*0,01339522*	*EPM2AIP1-MLH1*
cg11224603	*chr3*	*37035282*	*LLS-MLH1vsMC*	*-0,459588369*	*1,5039E-09*	*0,00010409*	*EPM2AIP1-MLH1*
cg19208331	*chr3*	*37035345*	*LLS-MLH1vsMC*	*-0,496556114*	*1,3073E-11*	*4,891E-06*	*EPM2AIP1-MLH1*
cg14598950	*chr3*	*37035355*	*LLS-MLH1vsMC*	*-0,331707848*	*7,8005E-08*	*0,00194565*	*EPM2AIP1-MLH1*
cg13846866	*chr3*	*37035399*	*LLS-MLH1vsMC*	*-0,298094031*	*6,8328E-08*	*0,00182602*	*EPM2AIP1-MLH1*
cg02279071	*chr3*	*37034154*	*LLSvsMC*	*-0,360782456*	*1,0476E-07*	*0,00217751*	*EPM2AIP1-MLH1*
cg14751544	*chr3*	*37034166*	*LLSvsMC*	*-0,400209867*	*1,343E-07*	*0,00264453*	*EPM2AIP1-MLH1*
cg16764580	*chr3*	*37034346*	*LLSvsMC*	*-0,425835333*	*2,5295E-09*	*0,0001352*	*EPM2AIP1-MLH1*
cg01302270	*chr3*	*37034441*	*LLSvsMC*	*-0,597660269*	*1,9095E-10*	*2,3814E-05*	*EPM2AIP1-MLH1*
cg17641046	*chr3*	*37034473*	*LLSvsMC*	*-0,593696318*	*1,4669E-09*	*0,00010409*	*EPM2AIP1-MLH1*
cg07101782	*chr3*	*37034495*	*LLSvsMC*	*-0,557054297*	*4,3468E-11*	*8,1315E-06*	*EPM2AIP1-MLH1*
cg10769891	*chr3*	*37034730*	*LLSvsMC*	*-0,458920768*	*5,3222E-09*	*0,00023309*	*EPM2AIP1-MLH1*
cg23658326	*chr3*	*37034787*	*LLSvsMC*	*-0,403278702*	*1,6692E-09*	*0,00010409*	*EPM2AIP1-MLH1*
cg11600697	*chr3*	*37034814*	*LLSvsMC*	*-0,428970019*	*2,9372E-08*	*0,00091577*	*EPM2AIP1-MLH1*
cg21490561	*chr3*	*37034825*	*LLSvsMC*	*-0,494376267*	*2,2595E-07*	*0,0042268*	*EPM2AIP1-MLH1*
cg00893636	*chr3*	*37034840*	*LLSvsMC*	*-0,446345427*	*5,3118E-07*	*0,00864066*	*EPM2AIP1-MLH1*
cg06791151	*chr3*	*37034956*	*LLSvsMC*	*-0,303127322*	*5,1985E-08*	*0,00149613*	*EPM2AIP1-MLH1*
cg24985459	*chr3*	*37035090*	*LLSvsMC*	*-0,397349923*	*5,6069E-09*	*0,00023309*	*EPM2AIP1-MLH1*
cg12790037	*chr3*	*37035117*	*LLSvsMC*	*-0,425614612*	*3,7725E-07*	*0,00641571*	*EPM2AIP1-MLH1*
cg17621259	*chr3*	*37035168*	*LLSvsMC*	*-0,366040011*	*8,862E-08*	*0,00207227*	*EPM2AIP1-MLH1*
cg05906740	*chr3*	*37035205*	*LLSvsMC*	*-0,266285456*	*1,6468E-08*	*0,00056013*	*EPM2AIP1-MLH1*
cg27331401	*chr3*	*37035207*	*LLSvsMC*	*-0,411753569*	*2,4069E-07*	*0,00428822*	*EPM2AIP1-MLH1*
cg25837710	*chr3*	*37035220*	*LLSvsMC*	*-0,295424543*	*9,534E-09*	*0,00035671*	*EPM2AIP1-MLH1*
cg12851504	*chr3*	*37035222*	*LLSvsMC*	*-0,283628981*	*9,5791E-08*	*0,00210819*	*EPM2AIP1-MLH1*
cg06590608	*chr3*	*37035228*	*LLSvsMC*	*-0,396755002*	*8,5926E-07*	*0,01339522*	*EPM2AIP1-MLH1*
cg11224603	*chr3*	*37035282*	*LLSvsMC*	*-0,459588369*	*1,5039E-09*	*0,00010409*	*EPM2AIP1-MLH1*
cg19208331	*chr3*	*37035345*	*LLSvsMC*	*-0,496556114*	*1,3073E-11*	*4,891E-06*	*EPM2AIP1-MLH1*
cg14598950	*chr3*	*37035355*	*LLSvsMC*	*-0,331707848*	*7,8005E-08*	*0,00194565*	*EPM2AIP1-MLH1*
cg13846866	*chr3*	*37035399*	*LLSvsMC*	*-0,298094031*	*6,8328E-08*	*0,00182602*	*EPM2AIP1-MLH1*
cg11291081	*chr3*	*37033894*	*LS-MLH1vsMC*	*-0,371115245*	*2,492E-06*	*0,03453129*	*EPM2AIP1-MLH1*
cg05845319	*chr3*	*37034066*	*LS-MLH1vsMC*	*-0,351275962*	*1,037E-06*	*0,01492238*	*EPM2AIP1-MLH1*
cg03901257	*chr3*	*37034142*	*LS-MLH1vsMC*	*-0,353001987*	*1,1978E-07*	*0,0019484*	*EPM2AIP1-MLH1*
cg02279071	*chr3*	*37034154*	*LS-MLH1vsMC*	*-0,371376384*	*2,0226E-08*	*0,00037837*	*EPM2AIP1-MLH1*
cg14751544	*chr3*	*37034166*	*LS-MLH1vsMC*	*-0,420766362*	*1,6435E-11*	*2,0497E-06*	*EPM2AIP1-MLH1*
cg16764580	*chr3*	*37034346*	*LS-MLH1vsMC*	*-0,434410819*	*4,5234E-10*	*1,3018E-05*	*EPM2AIP1-MLH1*
cg01302270	*chr3*	*37034441*	*LS-MLH1vsMC*	*-0,610600717*	*1,2046E-10*	*5,9011E-06*	*EPM2AIP1-MLH1*
cg17641046	*chr3*	*37034473*	*LS-MLH1vsMC*	*-0,58834543*	*1,9475E-08*	*0,00037837*	*EPM2AIP1-MLH1*
cg07101782	*chr3*	*37034495*	*LS-MLH1vsMC*	*-0,552071465*	*3,7198E-10*	*1,2652E-05*	*EPM2AIP1-MLH1*
cg10769891	*chr3*	*37034730*	*LS-MLH1vsMC*	*-0,465400466*	*1,21E-09*	*3,0181E-05*	*EPM2AIP1-MLH1*
cg23658326	*chr3*	*37034787*	*LS-MLH1vsMC*	*-0,411265051*	*3,6414E-11*	*3,406E-06*	*EPM2AIP1-MLH1*
cg11600697	*chr3*	*37034814*	*LS-MLH1vsMC*	*-0,445471307*	*7,6436E-10*	*2,0427E-05*	*EPM2AIP1-MLH1*
cg21490561	*chr3*	*37034825*	*LS-MLH1vsMC*	*-0,516735633*	*2,052E-10*	*8,5305E-06*	*EPM2AIP1-MLH1*
cg00893636	*chr3*	*37034840*	*LS-MLH1vsMC*	*-0,46385969*	*2,3849E-08*	*0,00042489*	*EPM2AIP1-MLH1*

Table S8-B (Cont.)

CG ID	Chromosome	Start	Comparison Group	Mean difference	p.value	FDR p.value	Gene
cg06791151	*chr3*	*37034956*	*LS-MLH1vsMC*	*-0,317113868*	*9,7243E-11*	*5,9011E-06*	*EPM2AIP1-MLH1*
cg24985459	*chr3*	*37035090*	*LS-MLH1vsMC*	*-0,394538719*	*5,5892E-08*	*0,00095052*	*EPM2AIP1-MLH1*
cg12790037	*chr3*	*37035117*	*LS-MLH1vsMC*	*-0,456526498*	*9,5236E-09*	*0,0002096*	*EPM2AIP1-MLH1*
cg17621259	*chr3*	*37035168*	*LS-MLH1vsMC*	*-0,382172302*	*4,4043E-10*	*1,3018E-05*	*EPM2AIP1-MLH1*
cg05906740	*chr3*	*37035205*	*LS-MLH1vsMC*	*-0,271068372*	*1,7629E-09*	*4,1224E-05*	*EPM2AIP1-MLH1*
cg27331401	*chr3*	*37035207*	*LS-MLH1vsMC*	*-0,437689445*	*5,4097E-11*	*4,048E-06*	*EPM2AIP1-MLH1*
cg25837710	*chr3*	*37035220*	*LS-MLH1vsMC*	*-0,307499026*	*2,8123E-13*	*5,261E-08*	*EPM2AIP1-MLH1*
cg12851504	*chr3*	*37035222*	*LS-MLH1vsMC*	*-0,300720203*	*1,2618E-10*	*5,9011E-06*	*EPM2AIP1-MLH1*
cg06590608	*chr3*	*37035228*	*LS-MLH1vsMC*	*-0,414447475*	*1,8062E-08*	*0,00037543*	*EPM2AIP1-MLH1*
cg11224603	*chr3*	*37035282*	*LS-MLH1vsMC*	*-0,463126031*	*2,6519E-10*	*9,9217E-06*	*EPM2AIP1-MLH1*
cg19208331	*chr3*	*37035345*	*LS-MLH1vsMC*	*-0,502994144*	*4,4572E-14*	*1,6676E-08*	*EPM2AIP1-MLH1*
cg14598950	*chr3*	*37035355*	*LS-MLH1vsMC*	*-0,331178728*	*2,2839E-07*	*0,00341796*	*EPM2AIP1-MLH1*
cg13846866	*chr3*	*37035399*	*LS-MLH1vsMC*	*-0,302244693*	*1,6035E-07*	*0,00249977*	*EPM2AIP1-MLH1*
cg11291081	*chr3*	*37033894*	*LSvsMC*	*-0,371115245*	*2,492E-06*	*0,03453129*	*EPM2AIP1-MLH1*
cg05845319	*chr3*	*37034066*	*LSvsMC*	*-0,351275962*	*1,037E-06*	*0,01492238*	*EPM2AIP1-MLH1*
cg03901257	*chr3*	*37034142*	*LSvsMC*	*-0,353001987*	*1,1978E-07*	*0,0019484*	*EPM2AIP1-MLH1*
cg02279071	*chr3*	*37034154*	*LSvsMC*	*-0,371376384*	*2,0226E-08*	*0,00037837*	*EPM2AIP1-MLH1*
cg14751544	*chr3*	*37034166*	*LSvsMC*	*-0,420766362*	*1,6435E-11*	*2,0497E-06*	*EPM2AIP1-MLH1*
cg16764580	*chr3*	*37034346*	*LSvsMC*	*-0,434410819*	*4,5234E-10*	*1,3018E-05*	*EPM2AIP1-MLH1*
cg01302270	*chr3*	*37034441*	*LSvsMC*	*-0,610600717*	*1,2046E-10*	*5,9011E-06*	*EPM2AIP1-MLH1*
cg17641046	*chr3*	*37034473*	*LSvsMC*	*-0,58834543*	*1,9475E-08*	*0,00037837*	*EPM2AIP1-MLH1*
cg07101782	*chr3*	*37034495*	*LSvsMC*	*-0,552071465*	*3,7198E-10*	*1,2652E-05*	*EPM2AIP1-MLH1*
cg10769891	*chr3*	*37034730*	*LSvsMC*	*-0,465400466*	*1,21E-09*	*3,0181E-05*	*EPM2AIP1-MLH1*
cg23658326	*chr3*	*37034787*	*LSvsMC*	*-0,411265051*	*3,6414E-11*	*3,406E-06*	*EPM2AIP1-MLH1*
cg11600697	*chr3*	*37034814*	*LSvsMC*	*-0,445471307*	*7,6436E-10*	*2,0427E-05*	*EPM2AIP1-MLH1*
cg21490561	*chr3*	*37034825*	*LSvsMC*	*-0,516735633*	*2,052E-10*	*8,5305E-06*	*EPM2AIP1-MLH1*
cg00893636	*chr3*	*37034840*	*LSvsMC*	*-0,46385969*	*2,3849E-08*	*0,00042489*	*EPM2AIP1-MLH1*
cg06791151	*chr3*	*37034956*	*LSvsMC*	*-0,317113868*	*9,7243E-11*	*5,9011E-06*	*EPM2AIP1-MLH1*
cg24985459	*chr3*	*37035090*	*LSvsMC*	*-0,394538719*	*5,5892E-08*	*0,00095052*	*EPM2AIP1-MLH1*
cg12790037	*chr3*	*37035117*	*LSvsMC*	*-0,456526498*	*9,5236E-09*	*0,0002096*	*EPM2AIP1-MLH1*
cg17621259	*chr3*	*37035168*	*LSvsMC*	*-0,382172302*	*4,4043E-10*	*1,3018E-05*	*EPM2AIP1-MLH1*
cg05906740	*chr3*	*37035205*	*LSvsMC*	*-0,271068372*	*1,7629E-09*	*4,1224E-05*	*EPM2AIP1-MLH1*
cg27331401	*chr3*	*37035207*	*LSvsMC*	*-0,437689445*	*5,4097E-11*	*4,048E-06*	*EPM2AIP1-MLH1*
cg25837710	*chr3*	*37035220*	*LSvsMC*	*-0,307499026*	*2,8123E-13*	*5,261E-08*	*EPM2AIP1-MLH1*
cg12851504	*chr3*	*37035222*	*LSvsMC*	*-0,300720203*	*1,2618E-10*	*5,9011E-06*	*EPM2AIP1-MLH1*
cg06590608	*chr3*	*37035228*	*LSvsMC*	*-0,414447475*	*1,8062E-08*	*0,00037543*	*EPM2AIP1-MLH1*
cg11224603	*chr3*	*37035282*	*LSvsMC*	*-0,463126031*	*2,6519E-10*	*9,9217E-06*	*EPM2AIP1-MLH1*
cg19208331	*chr3*	*37035345*	*LSvsMC*	*-0,502994144*	*4,4572E-14*	*1,6676E-08*	*EPM2AIP1-MLH1*
cg14598950	*chr3*	*37035355*	*LSvsMC*	*-0,331178728*	*2,2839E-07*	*0,00341796*	*EPM2AIP1-MLH1*
cg13846866	*chr3*	*37035399*	*LSvsMC*	*-0,302244693*	*1,6035E-07*	*0,00249977*	*EPM2AIP1-MLH1*

Table S9: Differentially methylated CpG islands **(A)** and CpG sites **(B)** found by methylome analysis of CRC DNA from LLS cases against controls (FDR p-value<0.05).

A) Differentially methylated CpG islands

Chromosome	Coordinates	Comparison Group	Mean difference	p.value	FDR p.value	Num sites	Gene
chr3	37034229-37035356	LLS-MLH1vsMC	-0.544866628	4.49E-09	0.00011382	22	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LLSvsMC	-0.573192401	1.49E-15	3.79E-11	22	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LS-MLH1vsMC	-0.596405263	7.44E-13	1.89E-08	22	*EPM2AIP1-MLH1*
chr3	37034229-37035356	LSvcMC	-0.595567704	2.04E-18	5.17E-14	22	*EPM2AIP1-MLH1*

B) Differentially methylated CpG sites

CG ID	Chromosome	Start	Comparison Group	Mean difference	p.value	FDR p.value	Gene
cg13846866	chr3	37035399	LLS-MLH1vsMC	-0.56036905807712	2,22E+06	0.0010338986	*EPM2AIP1-MLH1*
cg14598950	chr3	37035355	LLS-MLH1vsMC	-0.56434037131189	2,66E+07	0.0062199455	*EPM2AIP1-MLH1*
cg19208331	chr3	37035345	LLS-MLH1vsMC	-0.60633077150807	3,11E+06	0.0068378534	*EPM2AIP1-MLH1*
cg11224603	chr3	37035282	LLS-MLH1vsMC	-0.55825741324599	1,02E+07	0.0029392003	*EPM2AIP1-MLH1*
cg06590608	chr3	37035228	LLS-MLH1vsMC	-0.52885621236221	1,10E+07	0.0029487223	*EPM2AIP1-MLH1*
cg12851504	chr3	37035222	LLS-MLH1vsMC	-0.41940902799020	3,51E+06	0.0013127681	*EPM2AIP1-MLH1*
cg25837710	chr3	37035220	LLS-MLH1vsMC	-0.42213775928667	6,41E+06	0.0020226251	*EPM2AIP1-MLH1*
cg27331401	chr3	37035207	LLS-MLH1vsMC	-0.49710631752900	2,49E+06	0.0010338986	*EPM2AIP1-MLH1*
cg05906740	chr3	37035205	LLS-MLH1vsMC	-0.49395504029405	5,72E+04	7,58E+09	*EPM2AIP1-MLH1*
cg17621259	chr3	37035168	LLS-MLH1vsMC	-0.52771211626585	2,07E+06	0.0010338986	*EPM2AIP1-MLH1*
cg12790037	chr3	37035117	LLS-MLH1vsMC	-0.58060634759330	1,85E+08	0.0383864993	*EPM2AIP1-MLH1*
cg24985459	chr3	37035090	LLS-MLH1vsMC	-0.58965342675621	2,09E+06	0.0010338986	*EPM2AIP1-MLH1*
cg06791151	chr3	37034956	LLS-MLH1vsMC	-0.55404321152687	6,08E+04	7,58E+09	*EPM2AIP1-MLH1*
cg00893636	chr3	37034840	LLS-MLH1vsMC	-0.63868390803355	1,06E+06	0.0009892793	*EPM2AIP1-MLH1*
cg21490561	chr3	37034825	LLS-MLH1vsMC	-0.66996051258746	1,95E+06	0.0010338986	*EPM2AIP1-MLH1*
cg11600697	chr3	37034814	LLS-MLH1vsMC	-0.60737658649223	5,61E+04	7,58E+09	*EPM2AIP1-MLH1*
cg23658326	chr3	37034787	LLS-MLH1vsMC	-0.51452026784820	6,49E+06	0.0020226251	*EPM2AIP1-MLH1*
cg10769891	chr3	37034730	LLS-MLH1vsMC	-0.63002059444829	2,59E+07	0.0062199455	*EPM2AIP1-MLH1*
cg13846866	chr3	37035399	LLSvsMC	-0.56799652756749	1,05E+02	2,48E+04	*EPM2AIP1-MLH1*
cg14598950	chr3	37035355	LLSvsMC	-0.58056597963197	2,22E+03	4,62E+07	*EPM2AIP1-MLH1*
cg19208331	chr3	37035345	LLSvsMC	-0.64883626883094	3,18E+01	9,91E+05	*EPM2AIP1-MLH1*
cg11224603	chr3	37035282	LLSvsMC	-0.59546060302655	4,02E+01	1,16E+06	*EPM2AIP1-MLH1*
cg06590608	chr3	37035228	LLSvsMC	-0.56040881286050	7,81E+01	2,09E+06	*EPM2AIP1-MLH1*
cg12851504	chr3	37035222	LLSvsMC	-0.44026210071641	1,79E+00	9,58E+04	*EPM2AIP1-MLH1*
cg25837710	chr3	37035220	LLSvsMC	-0.43985263049873	2,39E+00	1,12E+05	*EPM2AIP1-MLH1*
cg27331401	chr3	37035207	LLSvsMC	-0.52204541374176	1,78E+00	9,58E+04	*EPM2AIP1-MLH1*
cg05906740	chr3	37035205	LLSvsMC	-0.50240226975958	4,15E-03	7,77E+02	*EPM2AIP1-MLH1*
cg17621259	chr3	37035168	LLSvsMC	-0.55297668758404	1,30E-01	9,58E+04	*EPM2AIP1-MLH1*
cg12790037	chr3	37035117	LLSvsMC	-0.62100428471847	4,86E+02	1,07E+06	*EPM2AIP1-MLH1*
cg24985459	chr3	37035090	LLSvsMC	-0.60948659188461	2,13E+01	7,23E+05	*EPM2AIP1-MLH1*
cg06791151	chr3	37034956	LLSvsMC	-0.56009746539956	1,43E-03	5,35E+02	*EPM2AIP1-MLH1*
cg00893636	chr3	37034840	LLSvsMC	-0.65754670739289	6,50E-01	6,08E+04	*EPM2AIP1-MLH1*
cg21490561	chr3	37034825	LLSvsMC	-0.69505633296746	4,81E+00	1,80E+05	*EPM2AIP1-MLH1*
cg11600697	chr3	37034814	LLSvsMC	-0.62309586772255	1,50E-02	1,87E+03	*EPM2AIP1-MLH1*
cg23658326	chr3	37034787	LLSvsMC	-0.53998183302662	4,12E+00	1,71E+05	*EPM2AIP1-MLH1*
cg10769891	chr3	37034730	LLSvsMC	-0.66092747374574	1,06E+02	2,48E+04	*EPM2AIP1-MLH1*
cg16433211	chr3	37034693	LLSvsMC	-0.39311455628655	6,07E+04	1,20E+09	*EPM2AIP1-MLH1*
cg07101782	chr3	37034495	LLSvsMC	-0.62443734519103	3,24E+06	0.0006067209	*EPM2AIP1-MLH1*
cg17641046	chr3	37034473	LLSvsMC	-0.62330636324278	1,35E+06	0.0024093632	*EPM2AIP1-MLH1*
cg16764580	chr3	37034346	LLSvsMC	-0.52566742785887	3,05E+08	0.0496714456	*EPM2AIP1-MLH1*
cg14751544	chr3	37034166	LLSvsMC	-0.55670143153729	7,99E+07	0.0135911220	*EPM2AIP1-MLH1*
cg13846866	chr3	37035399	LS-MLH1vsMC	-0.57700115009987	4,04E+05	8,40E+09	*EPM2AIP1-MLH1*
cg14598950	chr3	37035355	LS-MLH1vsMC	-0.61003028779275	2,45E+05	5,39E+09	*EPM2AIP1-MLH1*
cg19208331	chr3	37035345	LS-MLH1vsMC	-0.67960589476524	2,79E+02	2,09E+07	*EPM2AIP1-MLH1*
cg11224603	chr3	37035282	LS-MLH1vsMC	-0.62089916329509	1,38E+03	4,70E+07	*EPM2AIP1-MLH1*
cg06590608	chr3	37035228	LS-MLH1vsMC	-0.58825974479024	1,26E+03	4,70E+07	*EPM2AIP1-MLH1*
cg12851504	chr3	37035222	LS-MLH1vsMC	-0.45419833419882	5,70E+02	2,38E+07	*EPM2AIP1-MLH1*
cg25837710	chr3	37035220	LS-MLH1vsMC	-0.45221217020111	3,55E+00	1,33E+06	*EPM2AIP1-MLH1*
cg27331401	chr3	37035207	LS-MLH1vsMC	-0.53478931539881	4,61E+02	2,38E+07	*EPM2AIP1-MLH1*
cg05906740	chr3	37035205	LS-MLH1vsMC	-0.50986714466067	5,72E+02	2,38E+07	*EPM2AIP1-MLH1*
cg17621259	chr3	37035168	LS-MLH1vsMC	-0.57182956564362	3,37E-01	6,30E+06	*EPM2AIP1-MLH1*
cg12790037	chr3	37035117	LS-MLH1vsMC	-0.64900073589012	2,37E+05	5,39E+09	*EPM2AIP1-MLH1*
cg24985459	chr3	37035090	LS-MLH1vsMC	-0.62727544877050	2,51E+04	6,25E+08	*EPM2AIP1-MLH1*
cg06791151	chr3	37034956	LS-MLH1vsMC	-0.56659900215975	1,09E+04	2,91E+08	*EPM2AIP1-MLH1*
cg00893636	chr3	37034840	LS-MLH1vsMC	-0.67408658851140	3,37E+02	9,69E+07	*EPM2AIP1-MLH1*
cg21490561	chr3	37034825	LS-MLH1vsMC	-0.7173998676445	3,30E+03	9,69E+07	*EPM2AIP1-MLH1*
cg11600697	chr3	37034814	LS-MLH1vsMC	-0.63921226386939	2,06E+02	1,93E+07	*EPM2AIP1-MLH1*
cg23658326	chr3	37034787	LS-MLH1vsMC	-0.55739374796194	1,47E+01	1,83E+07	*EPM2AIP1-MLH1*
cg10769891	chr3	37034730	LS-MLH1vsMC	-0.68544924182801	4,00E+02	2,38E+07	*EPM2AIP1-MLH1*
cg07101782	chr3	37034495	LS-MLH1vsMC	-0.67651300654494	1,10E+06	0.0002161294	*EPM2AIP1-MLH1*
cg17641046	chr3	37034473	LS-MLH1vsMC	-0.66346585410099	3,98E+07	0.0074500818	*EPM2AIP1-MLH1*
cg16764580	chr3	37034346	LS-MLH1vsMC	-0.56890278218231	2,48E+08	0.0441428625	*EPM2AIP1-MLH1*

Table S9 -B (Cont.)

CG ID	Chromosome	Start	Comparison Group	Mean difference	p.value	FDR p.value	Gene
cg13846866	chr3	37035399	LSvcMC	-0.57865965032189	3,19E+00	7,46E+04	*EPM2AIP1-MLH1*
cg14598950	chr3	37035355	LSvcMC	-0.60621350776258	4,34E+00	9,55E+04	*EPM2AIP1-MLH1*
cg19208331	chr3	37035345	LSvcMC	-0.67881958916790	1,85E-06	4,70E-01	*EPM2AIP1-MLH1*
cg11224603	chr3	37035282	LSvcMC	-0.61899896493323	2,32E-01	7,89E+03	*EPM2AIP1-MLH1*
cg06590608	chr3	37035228	LSvcMC	-0.58142585005356	2,57E+00	6,41E+04	*EPM2AIP1-MLH1*
cg12851504	chr3	37035222	LSvcMC	-0.45158105007256	2,46E-02	1,15E+03	*EPM2AIP1-MLH1*
cg25837710	chr3	37035220	LSvcMC	-0.44944229982936	1,12E-04	1,05E+01	*EPM2AIP1-MLH1*
cg27331401	chr3	37035207	LSvcMC	-0.52497501009898	1,44E+00	3,85E+04	*EPM2AIP1-MLH1*
cg05906740	chr3	37035205	LSvcMC	-0.50813922666674	8,70E-04	5,43E+01	*EPM2AIP1-MLH1*
cg17621259	chr3	37035168	LSvcMC	-0.56556264163912	2,95E-02	1,23E+03	*EPM2AIP1-MLH1*
cg12790037	chr3	37035117	LSvcMC	-0.64064174589883	7,19E+00	1,49E+05	*EPM2AIP1-MLH1*
cg24985459	chr3	37035090	LSvcMC	-0.61958860871428	1,25E+00	3,60E+04	*EPM2AIP1-MLH1*
cg06791151	chr3	37034956	LSvcMC	-0.56314040525707	1,08E-02	5,76E+02	*EPM2AIP1-MLH1*
cg00893636	chr3	37034840	LSvcMC	-0.66680772362344	5,95E-02	2,23E+03	*EPM2AIP1-MLH1*
cg21490561	chr3	37034825	LSvcMC	-0.71065571097132	7,37E-02	2,30E+04	*EPM2AIP1-MLH1*
cg11600697	chr3	37034814	LSvcMC	-0.63661758065464	8,80E-05	1,05E+01	*EPM2AIP1-MLH1*
cg23658326	chr3	37034787	LSvcMC	-0.55702752746880	2,51E-06	4,70E-01	*EPM2AIP1-MLH1*
cg10769891	chr3	37034730	LSvcMC	-0.68440776443326	1,43E-04	1,07E+00	*EPM2AIP1-MLH1*
cg16433211	chr3	37034693	LSvcMC	-0.39174064770101	1,59E+05	2,83E+09	*EPM2AIP1-MLH1*
cg07101782	chr3	37034495	LSvcMC	-0.67979912802424	9,08E+02	1,79E+07	*EPM2AIP1-MLH1*
cg17641046	chr3	37034473	LSvcMC	-0.68069607072177	1,34E+05	2,50E+08	*EPM2AIP1-MLH1*
cg01302270	chr3	37034441	LSvcMC	-0.71430032968409	1,99E+06	0.0003386039	*EPM2AIP1-MLH1*
cg16764580	chr3	37034346	LSvcMC	-0.57190811279986	1,54E+07	0.0024380940	*EPM2AIP1-MLH1*
cg14751544	chr3	37034166	LSvcMC	-0.59316781865047	1,56E+07	0.0024380940	*EPM2AIP1-MLH1*

ANEXO II. INFORME DE LOS DIRECTORES

INFORME DE LOS DIRECTORES SOBRE EL FACTOR DE IMPACTO DE LOS ARTÍCULOS PUBLICADOS Y CONTRIBUCIÓN DE LA DOCTORANDA A CADA UNO DE ELLOS

Gabriel Capella y Marta Pineda, directores de la tesis doctoral de María Isabel González (titulada como "*Desarrollo de nuevas aproximaciones para el diagnóstico molecular de los síndromes de predisposición hereditaria al cáncer asociados a deficiencia del sistema de reparación de apareamientos erróneos*"), hacen constar que la doctoranda ha participado activamente en el diseño y realización experimental de los trabajos que se incluyen en esta tesis, en el análisis de los resultados, su discusión, obtención de conclusiones y en la redacción y preparación de los manuscritos finales. La contribución concreta a cada trabajo, así como el factor de impacto de los artículos científicos publicados, se especifican a continuación.

ARTÍCULO 1: Elucidating the clinical significance of two PMS2 missense variants coexisting in a family fulfilling hereditary cancer criteria.

Maribel González-Acosta, Jesús del Valle, Matilde Navarro, Bryony A. Thompson, Sílvia Iglesias, Xavier Sanjuan, María José Paúles, Natàlia Padilla, Anna Fernández, Raquel Cuesta, Àlex Teulé, Guido Plotz, Juan Cadiñanos, Xavier de la Cruz, Francesc Balaguer, Conxi Lázaro, Marta Pineda*, Gabriel Capellá*.

* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Famial Cancer: 2017 Oct;16(4):501-507.

Factor de impacto (2017 JCR Science Edition): 1,943

Contribución de la doctoranda: Recopilación de datos clínicos y anatomopatológicos de la familia portadora de las variantes, predicciones *in silico* del impacto de las variantes a nivel de RNA y proteína y diseño y análisis del estudio para determinar la fase alélica de las dos

variantes en identificadas en un mismo paciente. Estudio de cosegregación, coordinación del análisis multifactorial de probabilidad, cultivo de linfocitos con y sin puromicina, diseño y puesta a punto del análisis del procesamiento y estabilidad del RNA, modificación del plásmido de expresión específico para cada variante y su transfección, extracción de las proteínas y estudio de la actividad reparadora y expresión proteica de las variantes. Estudio de la gMSI y realización del análisis estadístico. Interpretación y discusión de resultados. Preparación de tablas y figuras. Finalmente, preparación y escritura del manuscrito.

ARTÍCULO 2: Validation of an in vitro mismatch repair assay used in the functional characterization of mismatch repair variants.

Maribel González-Acosta, Inga Hinrichsen, Anna Fernández, Conxi Lázaro, Marta Pineda*, Guido Plotz*, Gabriel Capellá*.

* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Manuscrito enviado al *Journal of Molecular Diagnostics* y pendiente de revisión.

Contribución de la doctoranda: Puesta a punto del ensayo *in vitro* de actividad reparadora en el laboratorio. Diseño y coordinación de la estrategia para evaluar la variabilidad intra- e inter-experimental y la reproducibilidad entre centros del ensayo. Modificación del plásmido de expresión específico para cada variante, transfección, extracción de las proteínas, estudio de la actividad reparadora y análisis de la expresión proteica de las variantes. Análisis estadístico. Recopilación, interpretación y discusión de resultados. Preparación de tablas y figuras. Finalmente, preparación y escritura del manuscrito.

ARTÍCULO 3: High-sensitivity microsatellite instability assessment for the detection of mismatch repair defects in normal tissue of biallelic germline mismatch repair mutation carriers.

Maribel González-Acosta*, Fàtima Marín*, Benjamin Puliafito*, Nuria Bonifaci, Anna Fernández, Matilde Navarro, Héctor Salvador, Francesc Balaguer, Sílvia Iglesias, Àngela Velasco, Èlia Grau, Víctor Moreno, Luis Ignacio Gonzalez-Granado, Pilar Guerra-García, Rosa Ayala9 Benoît Florkin, Christian P. Kratz, Tim Ripperger, Thorsten Rosenbaum, Danuta Januszkiewicz-Lewandowska, Amedeo A. Azizi, Iman Ragab, Michaela Nathrath, Hans-Jürgen Pander, Stephan Lobitz, Manon Suerink, Karin Dahan, Thomas Imschweiler, Ugur Demirsoy, Joan Brunet, Conxi Lázaro, Daniel Rueda, Katharina Wimmer, Gabriel Capellá‡, Marta Pineda‡.

* Ambos autores han contribuido en igual medida a este trabajo y comparten autoría.

‡ Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Journal of Medical Genetics: Epub 2019 Sep 7 (pii: jmedgenet-2019-106272).

Factor de impacto (2018* JCR Science Edition): 5,899

* El factor de impacto de 2019 aún no ha sido publicado, por lo que referenciamos el del 2018.

Contribución de la doctoranda: Investigación de la tecnología NGS más adecuada para el estudio y supervisión del diseño de las sondas incluidas en el panel NGS. Diseño de la estrategia experimental y recopilación de datos clínicos y anatomopatológicos de todos los individuos estudiados. Extracción de DNA de sangre periférica y mucosa bucal. Enriquecimiento y preparación de las librerías de DNA y cuantificación de las mismas mediante Qubit y/o BioAnalyzer de Agilent. Soporte en el análisis bioinformático. Análisis de los marcadores microsatélite dinucleótidos mediante contage de *reads* en la plataforma IGV v.2.4.10 y estudio de la gMSI. Análisis, interpretación y discusión de resultados. Preparación de tablas y figuras. Finalmente, preparación y escritura del manuscrito.

ARTÍCULO 4: Comprehensive characterization of MLH1 p.D41H and p.N710D variants coexisting in a Lynch syndrome family with conserved MLH1 expression tumors.

Marta Pineda*, **Maribel González-Acosta***, Bryony A. Thompson, Ricardo Sánchez, Carolina Gómez, Joaquín Martínez-López, José Perea, Pilar Garre, Trinidad Caldés, Yolanda Rodríguez, Stefania Landolfi, Judith Balmaña, Conxi Lázaro, Luis Robles, Gabriel Capellá‡, Daniel Rueda‡.

* Ambos autores han contribuido en igual medida a este trabajo y comparten autoría.

‡ Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Clinical Genetics: 2015 Jun;87(6):543-8. Epub 2014 Sep 16.

Factor de impacto (2014 JCR Science Edition): 3,931

Contribución de la doctoranda: Recopilación de datos clínicos y anatomopatológicos de la familia portadora de las variantes, predicciones *in silico* del impacto de las variantes a nivel de RNA y proteína, estudio de cosegregación, coordinación del análisis multifactorial de probabilidad, cultivo de linfocitos con y sin puromicina, diseño y puesta a punto del análisis del procesamiento y estabilidad del RNA y estudio de la expresión alélica diferencial. Modificación del plásmido de expresión específico para cada variante y transfección, extracción de las proteínas y estudio de la actividad reparadora y expresión proteica de las variantes. Análisis estadístico. Interpretación y discusión de resultados. Preparación de tablas y figuras. Finalmente, preparación y escritura del manuscrito.

ARTÍCULO 5: Elucidating the molecular basis of MSH2-deficient tumors by combined germline and somatic analysis.

Gardenia Vargas, **Maribel González-Acosta**, Bryony A. Thompson, Carolina Gómez, Anna Fernández, Estela Dámaso, Tirso Pons, Monika Morak, Jesús del Valle, Silvia Iglesias, Àngela Velasco, Ares Solanes, Xavier Sanjuan, Natàlia Padilla, Xavier de la Cruz, Alfonso Valencia,

Elke Holinki-Feder, Joan Brunet, Lídia Feliubadaló, Conxi Lázaro, Matilde Navarro, Marta Pineda* and Gabriel Capellá*.

* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

International Journal of Cancer: 2017 141: 1365-1380

Factor de impacto (2017 JCR Science Edition): 7,360

Contribución de la doctoranda: Recopilación de datos clínicos y anatomopatológicos de las familias portadoras de variantes de significado desconocido en MSH2, predicciones *in silico* del impacto de las variantes a nivel de RNA y proteína, predicción de las variantes en región promotora, estudios de cosegregación y coordinación del análisis multifactorial de probabilidad. Revisión de los resultados funcionales a nivel de RNA (*mRNA splicing analysis y allele-specific expression analysis*) y reclasificación de las variantes estudiadas mediante las evidencias generadas en el estudio. Preparación de tablas, participación en la escritura del manuscrito y revisión de la versión final.

ARTÍCULO 6: Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals.

Estela Dámaso, **Maribel González-Acosta**, Gardenia Vargas-Parra, Matilde Navarro, Judith Balmaña, Teresa Ramon y Cajal, Noemí Tuset, Fátima Marín, Anna Fernández, Carolina Gómez, Àngela Velasco, Ares Solanes, Sílvia Iglesias, Gisela Urgell, Consol López, Jesús del Valle, Olga Campos, Maria Santacana, Xavier Matias-Guiu, Conxi Lázaro, Laura Valle, Joan Brunet, Marta Pineda*, Gabriel Capellá*.

* Ambos autores han contribuido en igual medida a este trabajo y comparten la última posición.

Manuscrito en preparación.

Contribución de la doctoranda: Recopilación de datos clínicos y anatomopatológicos de las familias portadoras de variantes de significado desconocido en *MSH2* y *MSH6*, predicciones *in silico* del impacto de las variantes a nivel de RNA y proteína, estudios de cosegregación y coordinación del análisis multifactorial de probabilidad. Revisión de los resultados funcionales a nivel de RNA (*mRNA splicing analysis y allele-specific expression analysis*) y reclasificación de las variantes estudiadas mediante las evidencias generadas en el estudio. Preparación de tablas y figuras, participación en la escritura del manuscrito y revisión de la versión final.

Gabriel Capellá Munar, MD, PhD.
Director del Programa de Cáncer Hereditario
Laboratori de Recerca Translacional (LRT)
Instituto Catalán de Oncología (ICO) –
Instituto de Investigación Biomédica de
Bellvitge (IDIBELL)
gcapella@iconcologia.net

Marta Pineda Riu, PhD.
Unidad de Diagnóstico Molecular (UDM)
Laboratori de Recerca Translacional (LRT)
Instituto Catalán de Oncología (ICO) –
Instituto de Investigación Biomédica de
Bellvitge (IDIBELL)
mpineda@iconcologia.net