UNIVERSITAT DE BARCELONA

Identificación y caracterización de alteraciones epigenéticas responsables de síndrome de Lynch y Lynch-like

Estela Dámaso Riquelme

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

IDENTIFICACIÓN Y CARACTERIZACIÓN DE ALTERACIONES EPIGENÉTICAS RESPONSABLES DE SÍNDROME DE LYNCH Y LYNCH-LIKE

Memoria presentada por **Estela Dámaso Riquelme** para optar al grado de Doctora por la Universidad de Barcelona

Estela Dámaso Riquelme

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:

Dra. Susanna Balcells Comás

Barcelona, 2018

ÍNDICE

Índice

ABREVIATURAS i
Abreviaturas generales iii
Genes v
INTRODUCCIÓN 1
1. Epidemiologia del cáncer 3
2. El cáncer colorrectal 5
2.1. La carcinogénesis colorrectal 5
2.2. Vias de carcinogénesis 6
La vía de la inestabilidad cromosómica 7
La vía del fenotipo metilador de islas CpG 7
La vía de la inestabilidad de microsatélites 8
2.3. Clasificación del cáncer colorrectal según subtipos moleculares 9
Clasificación TCGA: Caracterización molecular integrada en cáncer colorrectal . 9
Clasificación CMS: Perfiles de expresión genética en cáncer colorrectal 10
2.4. Clasificación del cáncer colorrectal según su patrón de herencia 13
Esporádico 13
Familiar 13
Hereditario 14
2.5. Síndromes de predisposición a cáncer colorrectal hereditario 15
Síndromes polipósicos adenomatosos 15
Síndromes polipósicos hamartomatosos 17
Síndromes no polipósicos 18
3. El síndrome de Lynch 20
3.1. Historia del síndrome de Lynch 20
3.2. Características genéticas 21
Vía de reparación de desapareamientos del DNA 21
Espectro mutacional del síndrome de Lynch 22
3.3. Epidemiologia 23
3.4. Características clínicas 24

3.5. Características moleculares de los tumores 26

3.6. Asesoramiento genético y recomendaciones de seguimiento 27

4. Diagnóstico molecular del síndrome de Lynch 28

4.1. Criterios de selección 28

4.2. El algoritmo diagnóstico 31

Cribado molecular de los tumores 32

Diagnóstico molecular germinal 38

4.3. Evaluación de la patogenicidad de variantes en genes reparadores 42

Tasa de detección de mutaciones 46

5. Individuos con sospecha de síndrome de Lynch sin mutación identificada o Lynch-like 47

5.1. Potenciales causas del síndrome Lynch-like 47

Causas germinales 47

Causas somáticas 48

5.2. Riesgo de cáncer 48

5.3. Recomendaciones de seguimiento 49

6. Epimutaciones en síndrome de Lynch 50

6.1. El concepto de epimutación 50

6.2. Epidemiologia 50

6.3. Clasificación 51

Epimutaciones secundarias 52

Epimutaciones primarias en MLH1 54

6.4. Asesoramiento genético 56

HIPÓTESIS 59

Hipótesis 1 61

Hipótesis 2 61

OBJETIVOS 63

Objetivo principal 1 65

Objetivo principal 2 65

RESULTADOS 67

ARTÍCULO 1 71

Elucidating the molecular basis of MSH2-deficient tumors in suspected Lynch Syndrome cases. 71
ARTÍCULO 2 103
Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals 103
ARTÍCULO 3 139
Primary constitutional MLH1 epimutations: a focal epigenetic event 139
ARTÍCULO 4: 167
Highly sensitive MLH1 methylation analysis in blood allows the identification of low-level epigenetic mosaicism 167
DISCUSIÓN 197
1. El papel de las alteraciones epigenéticas constitucionales en el síndrome Lynch-like 200
1.1. Refinamiento genético de candidatos 200
1.2-Identificación de epimutaciones constitucionales en MLH1 205
1.3- Las alteraciones constitucionales epigenéticas fuera de los genes reparadores no parecen jugar un papel en el síndrome Lynch-like 206
1.4-Alteraciones somáticas en tumores de individuos Lynch-like 207
2-La importancia de la identificación y caracterización de las alteraciones epigenéticas constitucionales en síndrome de Lynch 209
2.1- Las técnicas altamente sensibles permiten la identificación de mosaicismo epigenético 209
2.2- Las epimutaciones constitucionales en MLH1 son un evento focal 211
2.3- La caracterización exhaustiva de las epimutaciones constitucionales permite su clasificación 212
2.4- En búsqueda del mecanismo causal 213
Sumario 217
CONCLUSIONES 219
REFERENCIAS 223
ANEXO 263
INFORME DE LOS DIRECTORES SOBRE EL FACTOR DE IMPACTO DE LOS ARTÍCULOS PUBLICADOS 265

ABREVIATURAS

Abreviaturas generales

Abreviatura	Nomenclatura oficial en inglés	Nomenclatura oficial en castellano
ACMG	*American College of Medical Genetics and Genomics*	-
AMP	*Association for Molecular Pathology*	-
C	*Cytosine*	Citosina
CCR	-	Cáncer ColoRrectal
CGH	*Comparative Genomic Hybridization*	Hibridación Genómica Comparativa
CIMP	CpG Island Methylator Phenotype	Fenotipo metilador de islas CpG
CIN	Chromosomal INstability	Inestabilidad cromosómica
CMMRD	*Constitutional MisMatch Repair Deficiency*	Deficiencia constitucional de los genes reparadores de desapareamientos del DNA
CMS	*Consensus Molecular Subtype*	Subtipos moleculares consenso
CNV	*Copy Number Variation*	Alteración en el número de copias
COBRA	*COmbined Bisulfite Restriction Analysis*	Análisis combinado de restricción con bisulfito
CpG	*5'—Cytosine—phosphate—Guanine—3'*	5'-Citosina-fosfato-Guanina-3'
Ct	*Threshold Cycle*	Ciclo umbral
dATP	*Deoxyadenosine triphosphate*	Desoxiadenosina trifosfato
dGTP	*Deoxyguanosine triphosphate*	Desoxiguanosina trifosfato
DNA	DeoxyriboNucleic Acid	Ácido desoxiribonucleico
EMAST	*Elevated Microsatellite Alterations at Selected Tetranucleotide repeats*	Elevada inestabilidad de microsatélites en repeticiones de tetranucleótidos
G	*Guanine*	Guanina
IARC	*International Agency for Research on Cancer*	-
IHQ	-	Inmunohistoquímica
InSiGHT	*Society for Gastrointestinal Hereditary Tumours*	-
LLS	*Lynch-Like Syndrome*	Síndrome Lynch-like

Abreviatura	Nomenclatura oficial en inglés	Nomenclatura oficial en castellano
LOH	*Loss Of Heterozygosity*	Pérdida de heterocigosidad
MLPA	*Multiplex Ligation-dependent Probe Amplification*	Amplificación dependiente de la ligación de sondas multiplexadas
MMR	*MisMatch Repair*	Reparación de desapareamientos del DNA
mRNA	*Messenger RiboNucleic Acid*	RNA mensajero
MSI	MicroSatellite Instability	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
MSP	*Methylation-Specific PCR*	PCR específica de metilación
MSS	*MicroSatellite Stability*	Estabilidad de microsatélites
NCI	*National Cancer Institute*	-
NGS	*Next Generation Sequencing*	Secuenciación de nueva generación
OMS	-	Organización mundial de la salud
PCR	*Polimerase Chain Reaction*	-
qPCR	*quantitative polymerase chain reaction*	PCR cuantitativa
RNA	*RiboNucleic Acid*	Ácido ribonucleico
SEOM	-	Sociedad Española de Oncología Médica
TCGA	*The Cancer Genome Atlas*	-
VUS	*Variant of Unkown significance*	Variante de significado desconocido

Genes

Genes	Nomenclatura oficial en inglés
APC	*Adenomatous polyposis coli*
ATM	*ATM serine/threonine kinase*
BMPR1A	*Bone morphogenetic protein receptor, type IA*
BRAF	*B-Raf proto-oncogene, serine/threonine kinase*
BRCA1	*Breast Cancer 1, Early Onset*
BRCA2	*Breast cancer 2, early onset*
BUB1	*BUB1 mitotic checkpoint serine/threonine kinase*
BUB3	*BUB3 mitotic checkpoint protein*
CACNA1G	*Calcium channel, voltage-dependent, T type, alpha 1G subunit*
CDKN2A	*Cyclin-dependent kinase inhibitor 2A*
CRABP1	*Cellular retinoic acid binding protein 1*
CTNNB1	*Catenin Beta 1*
DAPK1	*Death Associated Protein Kinase 1*
DCC	*DCC netrin 1 receptor*
EPCAM	*Epithelial cell adhesion molecule*
EXO1	*Exonuclease 1*
FAN1	*FANCD2/FANCI-Associated Nuclease*
GREM1	*Gremlin 1, DAN family BMP antagonist*
IGF2	*Insulin-like growth factor 2*
KLLN	*Killin, P53 Regulated DNA Replication Inhibitor*
KRAS	*Kirsten rat sarcoma 2 viral oncogene homolog*
LKB1 (ahora STK11)	*Liver Kinase B1*
MCM9	*Minichromosome Maintenance 9 Homologous Recombination Repair Factor*
MLH1	*MutL homolog 1*
MMACHC	*Methylmalonic Aciduria (Cobalamin Deficiency) CblC Type, With Homocystinuria*
MSH2	*MutS homolog 2*
MSH3	*MutS homolog 3*
MSH6	*MutS homolog 6*
MUTYH	*MutY homolog*

Genes	Nomenclatura oficial en inglés
MYC	*MYC Proto-Oncogene, BHLH Transcription Factor*
NEUROG1	*Neurogenin 1*
NTHL1	*Nth Endonuclease III-Like 1*
PD-1	*Programmed cell death protein 1*
PD-L1	*Programmed Death-ligand 1*
PIF1	*PIF1 5'-To-3' DNA Helicase*
PIK3CA	*Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha*
PMS2	*PMS2 postmeiotic segregation increased 2*
POLD1	*Polymerase (DNA directed), delta 1, catalytic subunit*
POLE	*Polymerase (DNA directed), epsilon, catalytic subunit*
PTEN	*Phosphatase and tensin homolog*
PTPRJ	*Protein tyrosine phosphatase, receptor type, J*
RAD51C	*RAD51 Paralog C*
RB1	*Retinoblastoma 1*
RBL1	*RB Transcriptional Corepressor Like 1*
RNF43	*Ring Finger Protein 43*
RPS20	*Ribosomal protein S20*
RUNX3	*Runt-related transcription factor 3*
SCG5	*Secretogranin V*
SEMA4	*Sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4A*
SETD2	*SET Domain Containing 2, Protein-Lysine N-Methyltransferase SETD2*
SMAD2	*SMAD Family Member 2*
SMAD4	*SMAD family member 4*
SOCS1	*Suppressor of cytokine signaling 1*
STK11 (antes LKB1)	*Serine/threonine kinase 11*
TELO2	*Telomere Maintenance 2*
TGF-β	*Transforming growth factor, beta*
TP53	*Tumor protein p53*
XAF1	*XIAP Associated Factor 1*

INTRODUCCIÓN

1. Epidemiologia del cáncer

Según la Organización Mundial de la Salud (OMS), "cáncer" es un término genérico que designa un amplio grupo de enfermedades que pueden afectar a cualquier parte del organismo. El cáncer se produce por la transformación de células normales en células tumorales, cuya característica definitoria es la multiplicación rápida de estas células anormales que se extienden más allá de sus límites habituales y pueden invadir partes adyacentes del cuerpo o propagarse a otros órganos, un proceso que se denomina metástasis (OMS, 2018).

El cáncer es una de las principales causas de morbilidad y mortalidad en el mundo. Sólo en 2012, se registraron alrededor de 14 millones de nuevos casos (OMS, 2018).

Conocemos como "incidencia" al número de casos nuevos de una enfermedad en una población y en un periodo de tiempo determinados En España, el número de nuevos casos estimados en el año 2017, excluyendo los tumores cutáneos no melanoma, fue de 228.482 (GLOBOCAN, 2012; OMS, 2018; SEOM, 2018). Según la Sociedad Española de Oncología Médica (SEOM, 2018), los tumores con mayor incidencia en este periodo fueron los de colon y recto, próstata, pulmón, mama, vejiga y estómago. En este trabajo, nos centraremos en el tumor más incidente, el cáncer colorrectal.

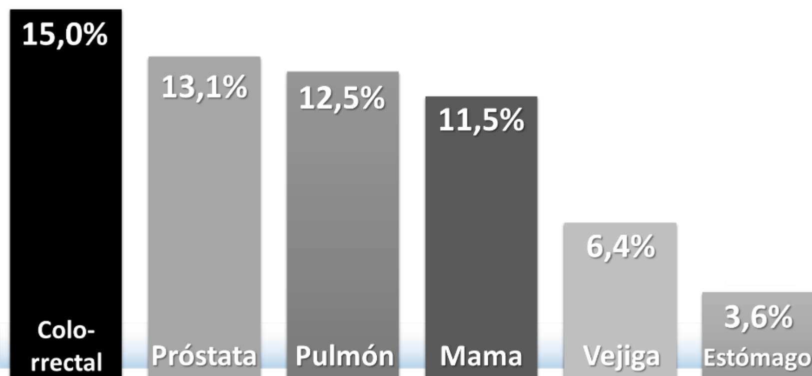


Figura 1: Incidencia estimada de los 6 tumores más frecuentes en España en el año 2017 para ambos sexos. Porcentaje total de los 228.482 casos estimados. Datos procedentes de GLOBOCAN 2012 y extrapolados a los datos de la población española para el año 2017 proporcionada por el INE.

La prevalencia es el número o la proporción de la población con una enfermedad determinada durante un periodo concreto. El concepto de prevalencia, por tanto, incluye tanto los pacientes con diagnóstico reciente como pacientes diagnosticados en el pasado y está fuertemente vinculada a la supervivencia de estos. De este modo, la prevalencia es más elevada en los tumores con mayor supervivencia; mientras que menor en tumores con una supervivencia más corta, (independientemente de que se puedan diagnosticar de forma más frecuente que los primeros, es decir, que tengan una incidencia mayor). El cáncer colorrectal es el tercer tipo de cáncer más prevalente en la población (solamente por detrás del cáncer de mama y de próstata), y se estima que es el cuarto tipo de tumor con más mortalidad, después de los tumores de pulmón, hígado y estómago. La mortalidad del cáncer colorrectal, sin embargo, ha experimentado un descenso de entre el 5,3 y el 6.7% desde 2012 gracias a las mejoras en el diagnóstico precoz y en los nuevos tratamientos (Malvezzi *et al*, 2017).

En líneas generales, se estima que uno de cada dos hombres y casi una de cada tres mujeres tendrá cáncer a lo largo de su vida. El riesgo de cáncer, entendido como la probabilidad de desarrollar un tumor, aumenta con la edad, tal y como se indica la figura 2.

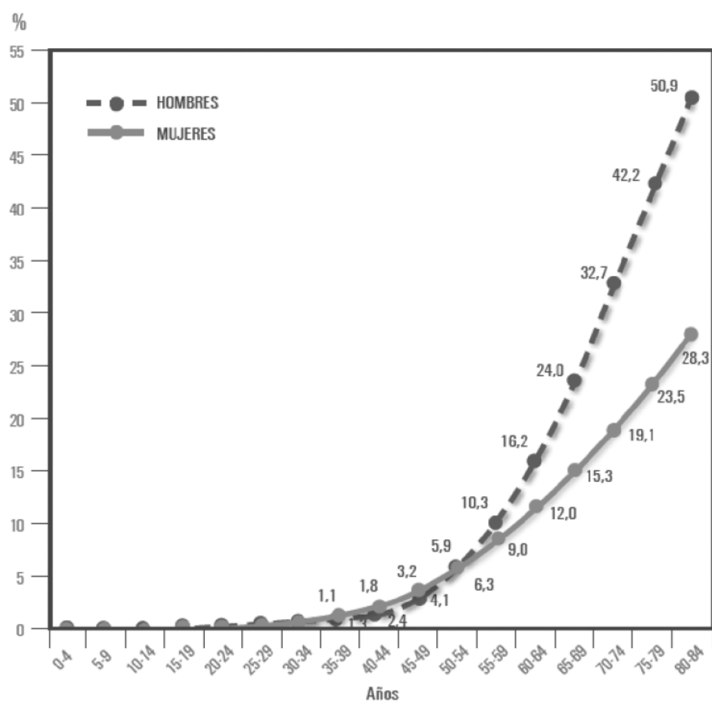


Figura 2: Riesgo acumulado de cáncer en España durante el periodo 2003-2007 por género. Figura procedente de SEOM 2018; Fuente REDECAN 2015.

2. El cáncer colorrectal

2.1. La carcinogénesis colorrectal

La aparición del cáncer colorrectal se produce por una transición del epitelio normal a neoplasia, que se inicia por la acumulación de alteraciones genéticas y epigenéticas en una célula (Hanahan & Weinberg, 2011).

El primer modelo genético de tumorogénesis para cáncer colorrectal fue propuesto por Fearon y Vogelstein en 1990. En él se definía la secuencia adenoma-carcinoma como la acumulación de 4 a 6 alteraciones genéticas, que podían activar oncogenes o silenciar genes supresores de tumores, y las cuales eran necesarias para la adquisición de una ventaja de crecimiento y de expansión clonal de las células tumorales (Fearon & Vogelstein, 1990).

En el año 2000, se definieron seis señas en fisiología celular que definen el comportamiento de las células neoplásicas: la autosuficiencia en cuanto a señales de crecimiento, la falta de sensibilidad hacia señales inhibidoras del crecimiento, la evasión de la apoptosis, el potencial ilimitado de crecimiento, la inducción de angiogénesis y la capacidad de invadir tejidos y metastatizar (Hanahan & Weinberg, 2000). Una década después, también se incluyeron como señas la capacidad de reprogramar su metabolismo energético y la de evadir su destrucción por el sistema inmune (Figura 3).

La adquisición de todas estas capacidades permite la supervivencia, proliferación y diseminación de las células cancerosas. Para que esta adquisición sea posible, las células tumorales necesitan dos características de habilitación: la primera y más prominente es el desarrollo de instabilidad genómica que genera mutaciones aleatorias, en las que se incluyen reordenamientos cromosómicos; y la segunda característica sería el estado inflamatorio del tumor provocado por las células del sistema inmune, que en algunos casos el tumor puede utilizar para promover su progresión (Hanahan & Weinberg, 2011).

La inestabilidad genómica se ha descrito ampliamente para el cáncer colorrectal donde se han propuesto tres vías de carcinogénesis principales: la de la inestabilidad cromosómica, la del fenotipo metilador de islas CpG y la de la inestabilidad de microsatélites (Figura 4).

Figura 3: Esquema de las señas del cáncer. Esta figura incluye tanto las 6 señas originales (Hanahan and Weinberg, 2000) como las propuestas posteriormente y las características de habilitación de las células neoplásicas (Hanahan and Weinberg, 2011).
Figura adatada de Hanahan and Weinberg, 2011

2.2. Vias de carcinogénesis

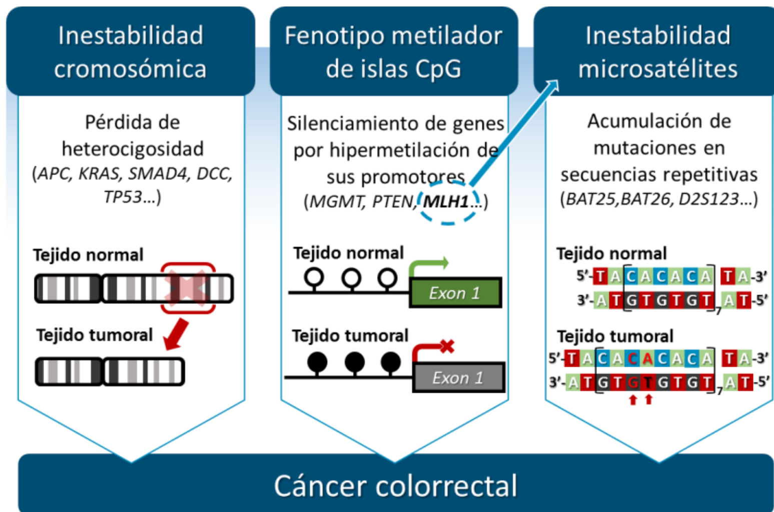


Figura 4: Esquema de las principales vías de carcinogénesis.

La vía de la inestabilidad cromosómica

La vía de la inestabilidad cromosómica (CIN, del inglés *Chromosomal INstability*) es la más común en el cáncer colorrectal, puesto que es la que se observa en el 75-85% de los casos. A pesar de que la vía de la inestabilidad cromosómica es muy frecuente, se desconoce su mecanismo causal, aunque se proponen mecanismos relacionados con la segregación cromosómica, la disfunción de los telómeros y la respuesta a daño del DNA (Pino & Chung, 2010). Este tipo de cáncer colorrectal se caracteriza por presencia de variaciones somáticas en el número de copia a través de todo el genoma, lo que resulta en tumores aneuploides y con pérdidas de heterocigosidad que, además, poseen una gran variabilidad cariotípica entre las diferentes células que lo forman. Junto con las aberraciones cariotípicas, se ha observado que existe una acumulación de mutaciones características en genes supresores de tumores (*APC, TP53, SMAD2, SMAD4* y *DCC*) y oncogenes (*KRAS, MYC, CTNNB1* y *PIK3CA*) que provocan la activación de vías oncogénicas, que son críticas para el desarrollo del cáncer colorrectal a través de esta vía (Pino & Chung, 2010; Vogelstein *et al*, 1988).

La vía del fenotipo metilador de islas CpG

La vía del fenotipo metilador de islas CpG (CIMP, del inglés *CpG Island Methylator Phenotype*) está presente en el 15-35% de los cánceres colorrectales. Se caracteriza por una hipermetilación global del genoma en unas regiones del DNA específicas, conocidas como islas CpG. Estas regiones son secuencias de una longitud mayor a 300-500 pares de bases y un contenido en dinucleótidos "CpG" en más del 50% de su secuencia (Gardiner-Garden & Frommer, 1987; Lao & Grady, 2011). Generalmente, las islas CpG se solapan con la región promotora del 60-70% de los genes, que tienden a estar protegidas de la metilación. Sin embargo, la inestabilidad epigenética de la vía CIMP provoca la hipermetilación aberrante de islas CpG/promotores de genes relacionados con el cáncer que, en el caso de los genes supresores de tumores, deriva en su silenciamiento transcripcional y la ausencia de expresión proteica (Puccini *et al*, 2017). La falta de consenso para la definición del estado CIMP de un tumor ha hecho que, históricamente, se hayan propuesto dos paneles de genes distintos: el de Ogino, que incluye 8 genes (*RUNX3, CACNA1G, IGF2, MLH1, NEUROG1, CRABP1, SOCS1* y *CDKN2A*) y el de Weisenberger, que incluye 5 (*CACNA1G, IGF2, NEUROG1, RUNX3* y *SOCS1*) (Ogino *et al*, 2007; Weisenberger *et al*, 2006). Sin embargo, la introducción de técnicas que permiten el estudio global de metiloma ha permitido la definición de 4 subgrupos dependiendo del estado CIMP de cada tumor mediante su modelización, basado en un análisis de agrupaciones (Hinoue *et al*, 2012). De estos 4 subgrupos, el de mayor relevancia para el presente estudio es el grupo 1, que está enriquecido en tumores CIMP positivos (según el panel de Weisenberger) con hipermetilación de *MLH1*. En este grupo se incluyen todos los tumores con mutación en *BRAF* y aproximadamente la mitad de los tumores con mutaciones en *KRAS*. Cuando la hipermetilación afecta a *MLH1,* la falta

de expresión de esta proteína induciría a su vez la vía de la inestabilidad de microsatélites.

La vía de la inestabilidad de microsatélites

La vía de la inestabilidad de microsatélites (MSI, del inglés *MicroSatellite Instability*) se observa en aproximadamente el 15% de los cánceres colorrectales y se utiliza como biomarcador de deficiencia de función reparadora de los genes reparadores de desapareamientos del DNA (MMR, del inglés *MisMatch Repair*). La causa más prevalente de la deficiencia MMR en casos de cáncer colorrectal es la hipermetilación aberrante del gen *MLH1* como consecuencia del CIMP, que supone el 80% de estos casos. En el 20% de casos restantes, la MSI está provocada por mutaciones en los genes MMR (*MLH1, MSH2, MSH6* y *PMS2*) (Inamura, 2018). La deficiencia reparadora provocada por la pérdida de función de los genes MMR, provoca a su vez una falta de corrección de errores producidos durante la replicación del DNA, que deriva en tumores "hipermutados". Estos errores no corregidos se acumulan sobre todo en microsatélites, que son secuencias repetitivas del DNA muy susceptibles a errores de replicación por las polimerasas, en particular en regiones de eucromatina e intrónicas, aunque no en secuencias de unión a nucleosomes (Kim & Park, 2014; Yamamoto & Imai, 2015). Otra característica de los tumores colorrectales hipermutados es la desregulación de la vía de WNT para favorecer la proliferación celular, que está provocada por la adquisición secundaria de mutaciones en *APC/CTNNB1* o *RNF43*, de forma excluyente (Yan *et al*, 2017).

2.3. Clasificación del cáncer colorrectal según subtipos moleculares

El cáncer colorrectal, por tanto, se puede interpretar como un grupo heterogéneo de enfermedades moleculares con diferentes manifestaciones clínicas. Con el fin de poder refinar este grupo y definir subtipos moleculares característicos que permitan la implementación de terapias personalizadas y el mejor manejo clínico de los pacientes, estos últimos años han tenido lugar diferentes estudios que han permitido una mejor integración de la patología molecular y las características clínicas del cáncer colorrectal. Esta integración de datos moleculares ha generado dos sistemas de clasificación: la clasificación TCGA y la CMS.

Clasificación TCGA: Caracterización molecular integrada en cáncer colorrectal

En 2012, la red de investigación del *The Cancer Genome Atlas* (TCGA) llevó a cabo una caracterización molecular integral de 224 casos con cáncer colorrectal, donde se analizaron exomas o genomas completos, variaciones en número de copia, metilación de promotores y patrones de expresión (The Cancer Genome Atlas, 2012).

La clasificación principal se realizó dependiendo de la tasa de mutación encontrada en los tumores, de modo que aquellos con más de 12 mutaciones por 10^6 bases se clasificaron como hipermutados y los que no superaban las 8,24 mutaciones por 10^6 bases se clasificaron como no-hipermutados.

Los tumores hipermutados representaron un 16% del total de casos examinados. Dentro de estos, el 75% de los casos presentaban MSI, metilación en *MLH1* y fenotipo metilador de islas CpG, mientras que el 25% restante presentaron mutaciones somáticas en genes MMR y en *POLE*. Los tumores con mutaciones en *POLE* mostraron tasas de mutación superiores a las 40 mutaciones por 10^6 bases, por lo que se denominaron “ultrahipermutados”. Los estudios de metilación agruparon la mayor parte de los tumores hipermutados dentro del fenotipo metilador de islas CpG así como la mayor parte de tumores con la mutación *BRAF* V600E.

Los tumores no-hipermutados se caracterizaron por presentar microsatélites estables (MSS, del inglés *MicroSatellite Stability*) y tener frecuentemente variaciones somáticas de números de copia. Este grupo, al contrario que el anterior, no presentaba alteraciones epigenéticas significativas relacionadas con el fenotipo metilador de islas CpG (Hao *et al*, 2012; Inamura, 2018; The Cancer Genome Atlas, 2012).

Clasificación CMS: Perfiles de expresión genética en cáncer colorrectal

En 2015, Guinney y colaboradores propusieron una categorización de los tumores según el perfil de expresión génica, puesto que esta está íntimamente asociada con el fenotipo celular y el comportamiento del tumor (Guinney *et al*, 2015). Este estudio integrativo con 4151 muestras, incluyendo las disponibles en TCGA, derivó en la categorización de los tumores en cuatro subtipos moleculares consenso (CMS, del inglés *Consensus Molecular Subtype*). Esta clasificación se correlacionó posteriormente con las características genéticas y epigéneticas de los tumores colorrectales, con el fin de representar mejor heterogeneidad del cáncer colorrectal (Figura 5).

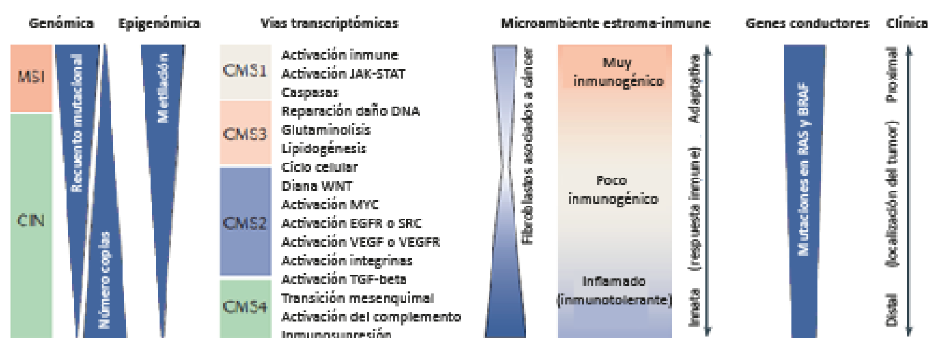


Figura 5: Representación esquemática de los subtipos de CRC y su integración con las diferentes clasificaciones moleculares.
Abreviaturas: EGFR=receptor del factor de crecimiento epidérmico; JAK=Janus quinasa; STAT=transductor de señal y activador de la transcripción; TGFβ=factor de crecimiento transformante-β; VEGF=factor de crecimiento endotelial vascular; VEGFR=receptor de VEGF.
(Dienstmann et al 2018)

El primer grupo, CMS1 contó con el 14% de las muestras y se definió como el subgrupo “MSI-inmune”, puesto que presentaba MSI y activación inmune. El CMS2 se definió, con el 37% de las muestras, como el subgrupo “canónico”. El CMS3 (13%) se nombró “metabólico” por el carácter epitelial y la desregulación metabólica de estos tumores. Y, por último, se definió el subtipo “mesenquimal”, CMS4, con un 23% de las muestras. Este subgrupo se caracterizó por la infiltración estromal, una transición epitelio-mesenquimal activa, remodelación de la matriz y angiogénesis. Las características moleculares de estos grupos se resumen en la figura 6. Sin embargo, un 13% de las muestras mostraron características de varios grupos que podrían ser consecuencia de fenotipo de transición o heterogeneidad intratumoral.

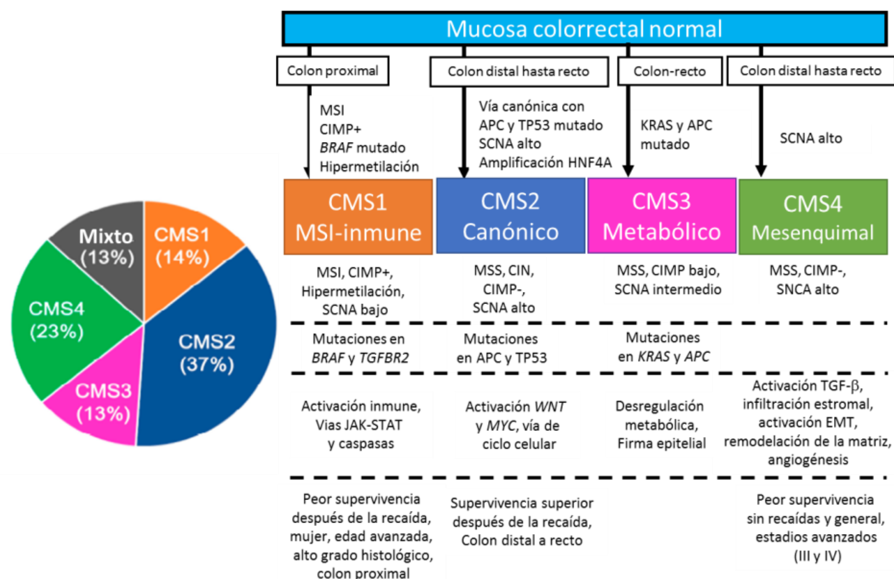


Figura 6: Clasificación de los distintos tipos de cáncer colorrectal según el modelo CMS, que refleja diferencias biológicas significativas en los subtipos moleculares basados en la expresión génica.

Abreviaturas: CIMP=fenotipo de metilador de isla CpG; MSI: inestabilidad de microsatélites; SCNA=alteraciones en el número de copias somáticas.

(Inamura 2018; Guinney et al 2015)

Ambos sistemas de clasificación coinciden en la identificación del grupo con MSI/deficiencia MMR (Figura 7). Recientemente se ha demostrado que el gran número de neoantígenos generados por la deficiencia MMR provoca que haya una buena de respuesta de este grupo de tumores al bloqueo de puntos de control inmunitario (anticuerpos contra PD-1/PD-L1), que activa ataques de células T citotóxicas en células tumorales y que podría resultar muy útil para el tratamiento de estos pacientes mediante inmunoterapia (Le *et al*, 2017; Lote *et al*, 2015; Lynch & Murphy, 2016; Xiao & Freeman, 2015).

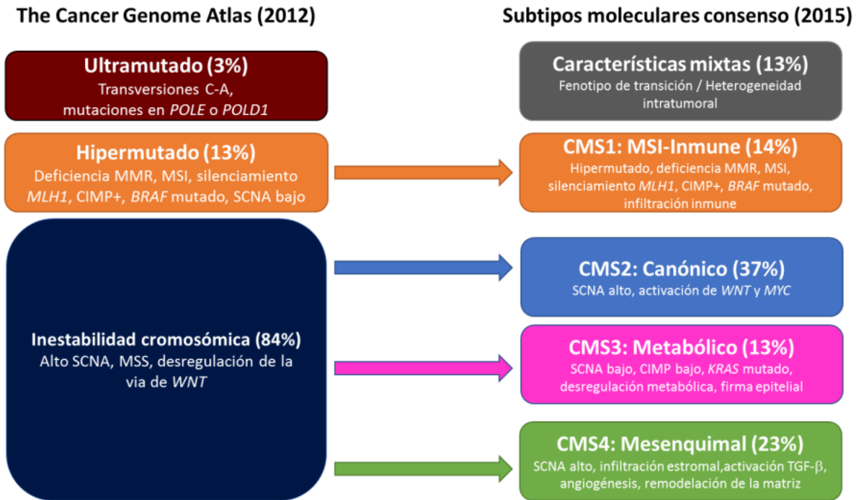


Figura 7: Comparación de las agrupaciones de tumores colorrectales por los principales sistemas de clasificación.

2.4. Clasificación del cáncer colorrectal según su patrón de herencia

En función del patrón de herencia, el cáncer colorrectal se puede clasificar como esporádico, familiar o hereditario (Figura 8).

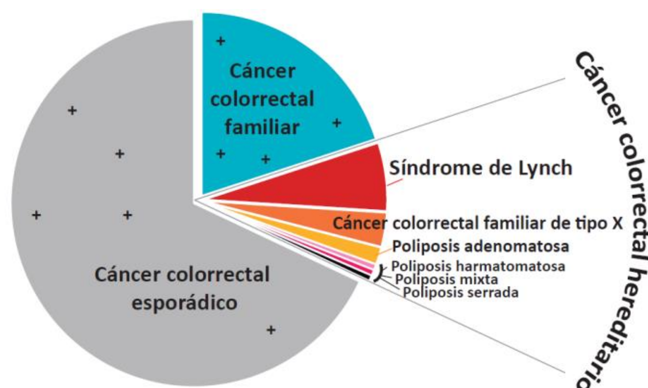


Figura 8: Gráfico circular que representa la marcada heterogeneidad genotípica y fenotípica en los síndromes de cáncer colorrectal hereditario. Las cruces indican posibles variantes aún no descubiertas de cánceres hereditarios.
Figura adatada de Lynch et al., 2009

Esporádico

La mayoría de los casos de cáncer colorrectal (70-80%) se consideran esporádicos, asociados al estilo de vida. Generalmente no tienen agregación familiar de cáncer y la edad media al diagnóstico suele ser tardía, alrededor de los 70-75 años (Kapitanović, 2017; Watson & Collins, 2011).

Familiar

Alrededor del 25% de los pacientes con cáncer colorrectal tienen antecedentes familiares de cáncer y un riesgo incrementado sobre el riesgo poblacional (entre 1.5 y 2.5 veces), aunque se desconoce su causa (Alonso Sánchez *et al*, 2008). Entre los mecanismos propuestos para explicar esta predisposición están: las alteraciones en genes menos penetrantes que los asociados a síndromes hereditarios, alteraciones en loci de susceptibilidad con efectos aditivos y/o modificadores ambientales (Jasperson *et al*, 2010; Kapitanović, 2017).

Hereditario

Aproximadamente el 5% de los casos de cáncer colorrectal están asociados a mutaciones o epimutaciones germinales en genes de susceptibilidad altamente penetrantes y/o a patrones de herencia mendelianos (Jasperson *et al*, 2010; Lynch *et al*, 2009).

Existen diferentes mecanismos genéticos de predisposición. El más común es la mutación puntual que incluye la sustitución, inserción o deleción de un único nucleótido. Otros mecanismos son las variaciones en el número de copia (que incluyen inserciones y deleciones de diferentes tamaños, pudiendo abarcar desde unos pocos nucleótidos a genes enteros) y los grandes reordenamientos (como inversiones y translocaciones). También existen mecanismos epigenéticos de predisposición, como es el caso de las epimutaciones, donde la metilación aberrante de un gen se puede encontrar en tejidos somáticos normales de un individuo (ver apartado 6.1. para más información).

Dependiendo de sus manifestaciones clínicas, el cáncer colorrectal hereditario se ha subdividido en dos tipos de síndromes: los polipósicos y los no polipósicos.

2.5. Síndromes de predisposición a cáncer colorrectal hereditario

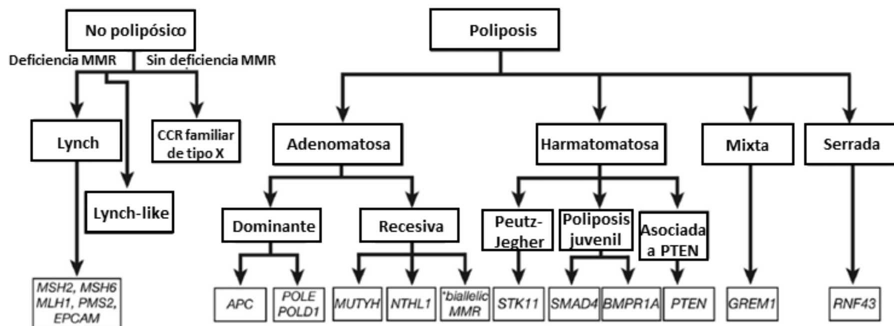


Figura 9: Clasificación de los síndromes de predisposición a cáncer colorrectal hereditario (Valle 2017).

Síndromes polipósicos adenomatosos

<u>-Dominantes</u>

La **poliposis adenomatosa familiar asociada a APC.** Está causada por mutaciones en el gen *APC*, que codifica para una gran proteína supresora tumoral con afecto en la adhesión y migración celular. Su función es la regulación de la beta-catenina dentro de la vía de señalización de WNT. Su patrón de herencia es autosómico dominante. Es el síndrome polipósico más común (1% del total de casos de cáncer colorrectal). El diagnóstico clínico de la poliposis adenomatosa familiar se basa en la presencia de al menos 100 adenomas colorrectales (**poliposis clásica**) o entre 10 y 100 adenomas que aparecen en cuarta y quinta década de la vida además de historia familiar de poliposis (**poliposis atenuada).** En la poliposis clásica los adenomas generalmente se desarrollan en la adolescencia y el riesgo de cáncer colorrectal es del 100% a los 40-50 años. Además de las manifestaciones colónicas, la poliposis adenomatosa familiar se asocia con múltiples manifestaciones extracolónicas: adenomas gastroduodenales, osteomas, hipertrofia congénita del pigmento epitelial de la retina, anomalías dentales, tumores desmoides y adenomas adrenocorticales (Byrne & Tsikitis, 2018).

La **poliposis asociada a las polimerasas con prueba de lectura, descrito más recientemente,** está causada por mutaciones en los dominios exonucleasa de las DNA polimerasas épsilon (*POLE*) o delta (*POLD1*) (Palles *et al*, 2012). Se trata de un síndrome

autosómico dominante altamente penetrante, caracterizado por poliposis colorrectal atenuada u oligoadenomatosa, cáncer colorrectal, adenomas gastroduodenales y tumores cerebrales en algunos casos. Las mujeres portadoras de mutaciones en *POLD1*, además, tienen un alto riesgo de cáncer de endometrio y un riesgo moderado de cáncer de mama. Los tumores desarrollados en estos pacientes muestran fenotipo (ultra)hipermutado, con una alta proporción de transversiones G:C>T:A y A:T>C:G (Bellido *et al*, 2015; The Cancer Genome Atlas, 2012; Valle, 2017).

-Recesivos

La **poliposis asociada a *MUTYH*** es un síndrome raro de poliposis adenomatosa autosómico recesivo con una alta penetrancia. Está causado por mutaciones bialélicas en el gen *MUTYH*, que codifica para una proteína que repara el daño oxidativo del DNA. Las mutaciones más comunes en *MUTYH* son NM_001128425.1: c.536A>G y c. 1187G>A, que están presentes en el 70% de pacientes con este tipo de poliposis (Achatz *et al*, 2017; Nielsen *et al*, 2009). La mayor parte de los pacientes presentan menos de 100 adenomas, aunque algunos presentan fenotipos severos. Generalmente estos pacientes se diagnostican a los 40-50 años (Balmaña *et al*, 2013; Sampson *et al*, 2003). Aunque se trata de un síndrome autosómico recesivo, el riesgo de cáncer colorectal en portadores monoalélicos es controvertido (Jones *et al*, 2009; Rosner *et al*, 2015; Theodoratou *et al*, 2010; Win *et al*, 2014). Estudios recientes han reportado que estos individuos presentan un ligero incremento de riesgo respecto a la población general, especialmente cuando tienen historia familiar de cáncer colorrectal (Win *et al*., 2014).

La **poliposis adenomatosa asociada a *NTHL1*** está causado por mutaciones en el gen *NTHL1*, cuya proteína también está involucrada en la reparación del daño oxidativo del DNA. Aunque se han descrito muy pocas familias con este tipo de poliposis, las evidencias encajan con un trastorno recesivo. La mutación más común es *NTHL1* NM_002528:c.268C>T (p.Gln90*) (Belhadj et al, 2017) Aunque aún no se ha definido el espectro tumoral de este síndrome, se ha observado que el fenotipo asociado a este síndrome abarca varios tipos de tumores, incluyendo el cáncer colorrectal (Rivera et al, 2015; Weren et al, 2015).

La **poliposis adenomatosa asociada a *MSH3*** está provocada por mutaciones bialélicas en *MSH3*, cuya proteína forma parte del sistema de reparación de desapareamientos de DNA. Su papel más importante es la reparación de los bucles de inserción/deleción tras la replicación. Debido a la pérdida de esta función, los tumores de estos pacientes presentan EMAST (elevated microsatellite alterations at selected tetranucleotide repeats) (Adam *et al*, 2016). El fenotipo observado en los 2 individuos identificados hasta el momento es muy similar al de la poliposis familiar atenuada ya que presentan pólipos colorrectales adenomatosos que son diagnosticados entre los 30-40 años de edad.

Además, estos casos también pueden presentar otras lesiones extracolónicas tanto benignas como malignas: adenomas duodenales, adenomas tiroideos, papilomas intraductales, cáncer gástrico y astrocitomas (Adam *et al*, 2016).

Síndromes polipósicos hamartomatosos

El **síndrome de Peutz-Jeghers** se caracteriza por la presencia pólipos hamartomatosos del músculo liso en el tracto gastrointestinal, principalmente en intestino delgado (78%) y colon (42%), así como de la pigmentación melanótica de piel y mucosas (Graziano *et al*, 2009). Los criterios de la OMS determinan que se deben confirmar al menos 3 pólipos harmatomatosos en casos sin historia familiar de la patología, aunque en casos familiares bastaría con confirmación de al menos un pólipo o de la pigmentación melanótica. Este síndrome está causado por mutaciones en el gen *STK11* (también conocido como *LKB1*) y se transmite de forma autosómica dominante. El riesgo de cáncer en cualquiera de las manifestaciones del síndrome se estima del 90% (Aretz *et al*, 2005; Volikos *et al*, 2006).

El **síndrome de poliposis juvenil** se caracteriza por la presencia de múltiples (5 o más) pólipos hamartomatosos en el tracto gastrointestinal (Byrne & Tsikitis, 2018). En un 15% de los casos también se observan anomalías congénitas (macrocefalia, paladar y labios hendidos, enfermedad cardíaca y telangiectasias, entre otros). Se asocia a mutaciones en los genes *SMAD4* y *BMPR1A* y presenta patrón de herencia autosómico dominante (Byrne & Tsikitis, 2018; Howe *et al*, 2004). El riesgo de desarrollar cáncer colorrectal en este síndrome es del 39% (Brosens *et al*, 2007).

El **síndrome de poliposis hamartomatosa asociada a *PTEN*** engloba dos condiciones hereditarias autosómicas dominantes: el síndrome de Cowden y el síndrome de Bannayan-Riley-Ruvalcaba. Ambos síndromes se caracterizan por la presencia pólipos harmatomatosos colorrectales y transtornos en el desarrollo, que en el caso del síndrome de Bannayan-Riley-Ruvalcaba conlleva retraso mental en el 50% de los casos (García-Palacios & Bautista-Casasnovas, 2014).

> El **síndrome de Cowden** es más común en adultos y mujeres, y cuenta con una alta incidencia de cáncer de mama y tiroides. Los pólipos gastrointestinales están presentes en el 40-70% de los casos, aunque el riesgo de cáncer colorrectal no está bien definido.

El **síndrome de Bannayan-Riley-Ruvalcaba** es más común en niños y hombres. Los hamartomas colorrectales están presentes en el 50% de los pacientes, y suelen presentarse en periodo lactante, aunque no se tienen evidencias de un incremento del riesgo de cáncer colorrectal.

La **poliposis hereditaria mixta** es un síndrome raro que se caracteriza por una mezcla de lesiones exclusivamente colorrectales que incluyen pólipos Peutz-Jeghers, pólipos juveniles, hiperplásicos, adenomas clásicos y cáncer colorrectal. Se ha asociado a alteraciones en la región 5' del gen *GREM1*. La más conocida es la duplicación de 40Kb que se extiende desde el intrón 2 de *SCG5* (gen previo a *GREM1)* hasta la isla CpG de *GREM1* y que se ha documentado exclusivamente en población judía Ashkenazi (Jaeger *et al*, 2012; Laitman *et al*, 2015; Venkatachalam *et al*, 2011). Sin embargo, en 2016, Rohlin y colaboradores describieron una familia no-Ashkenazi con poliposis mixta, en la cual relacionaron una duplicación de 16Kb de un dominio regulador de *GREM1* con la enfermedad (Rohlin *et al*, 2016). En 2014, se asoció la presencia de la variante rs16969681 en población general, que afecta a una región potenciadora de *GREM1*, con la susceptibilidad a cáncer colorrectal, confiriendo aproximadamente un 20% de riesgo diferencial (Lewis *et al*, 2014; Valle, 2017). Recientemente las mutaciones en el gen *MCM9* también se han asociado a poliposis mixta y cáncer colorrectal (Goldberg *et al*, 2015).

La **poliposis serrada** es una condición caracterizada por la abundancia de lesiones dentadas en el intestino grueso. Estas lesiones dentadas incluyen pólipos hiperplásicos y adenomas/pólipos sésiles serrados con o sin displasia, entre otros (Hui *et al*, 2014). El riesgo de cáncer colorrectal está muy relacionado con el incremento de pólipos serrados en estos pacientes. Aunque una gran parte de los tumores colorrectales en poliposis serradas se consideran esporádicos, debidos a la acumulación de mutaciones somáticas y CIMP, en un 38-50% de los casos se ha observado un claro componente familiar. Gala y colaboradores demostraron que había una clara asociación de la poliposis serrada familiar con mutaciones en genes reguladores de la senescencia, como son *RNF43, ATM, PIF1, TELO2, XAF1* y *RBL1* (Gala *et al*, 2014). El más destacable es el gen *RNF43*, donde se ha demostrado su naturaleza como gen supresor tumoral e implicación en el desarrollo de la patología en varias familias (Quintana *et al*, 2018; Taupin *et al*, 2015; Yan *et al*, 2017).

Síndromes no polipósicos

Principalmente se han descrito dos síndromes de cáncer colorrectal hereditario no polipósico: el cáncer colorrectal familiar de tipo X y el síndrome de Lynch.

El **cáncer colorrectal familiar de tipo X** es un síndrome autosómico dominante cuya casa genética es desconocida. Estas familias se caracterizan por tener una fuerte agregación familiar de tumores colorrectales con estabilidad de microsatélites y representan el 40% de las familias que cumplen criterios Ámsterdam 1 con una edad de inicio elevada (Nejadtaghi *et al*, 2017). Estudios recientes han descrito la correlación entre el cáncer colorrectal familiar tipo X y variantes en genes como *BRCA2* (Esteban-Jurado *et al*, 2016; Garre *et al*, 2015; Yurgelun *et al*, 2015), *SEMA4* (Kinnersley *et al*, 2016; Schulz *et al*, 2014), *BMPR1A* (Nieminen et al, 2011; Sjöblom et al, 2006), *RPS20* (Nieminen *et al*, 2014), *FAN1* (Seguí *et al*, 2015; Smith *et al*, 2016) y *BUB1* y *BUB3* (de Voer *et al*, 2013), aunque esta correlación se ha observado en estudios aislados.

El **síndrome de Lynch** es el síndrome de cáncer colorrectal hereditario más frecuente. Puesto que se trata del grupo de estudio del presente trabajo, se desarrolla en profundidad en el siguiente apartado.

3. El síndrome de Lynch

3.1. Historia del síndrome de Lynch

La historia del síndrome de Lynch comienza en 1895 con Aldred Warthin, un patólogo que documentó la historia familiar de cáncer de su costurera, la cual padecía depresión debido a la gran cantidad de muertes en su familia por cáncer colorrectal, de estómago y útero. Warthin publicó este pedigrí, al que llamó “Familia G” (Figura 10), junto con otros dos de similares características en 1913, donde ya advirtió que el patrón de herencia de la enfermedad era consistente con un patrón mendeliano autosómico dominante (Warthin, 1985).

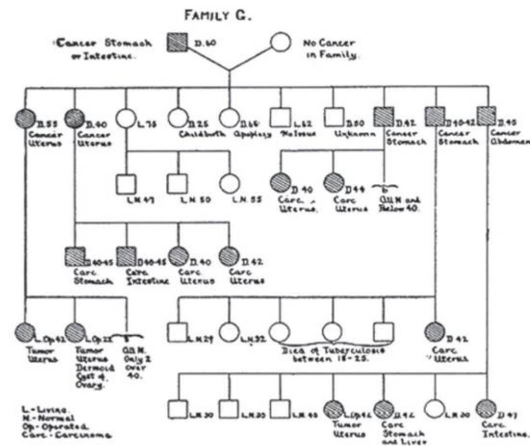


Figura 10: Pedigrí de la “Familia G” reportada por Warthin en 1913. Los informes sobre la “Familia G” fueron uno de los primeros registros de agrupación familiar de cáncer, que proporcionaron la base para el descubrimiento de la herencia genética la de la predisposición a la enfermedad

Setenta años después, Henry Lynch publicó dos familias adicionales, de características muy similares a las reportadas por Warthin, a las que denomino familias N y M (Lynch *et al*, 1966). En esta publicación, Lynch describió que el cáncer en estas familias se reducía solo a ciertos órganos y que la edad de presentación del cáncer era inusualmente temprana. Por primera vez se consideró esta patología como un síndrome, el cual recibió el nombre de “síndrome de cáncer familiar”. Este nombre, sin embargo, fue evolucionando a lo largo del tiempo en un intento de refinar su significado. Así pues, en 1984 se utilizó por primera vez el término “síndrome de Lynch” para remarcar que el riesgo en estas familias estaba restringido a ciertos órganos (Boland & Troncale, 1984) y, en 1985, se propuso el nombre “Cáncer colorrectal hereditario no polipósico” para enfatizar que esta enfermedad era distinta de la poliposis adenomatosa familiar (Lynch *et al*, 1985). Debido a que este último nombre solo hacía referencia al riesgo de cáncer colorrectal y no abarcaba todo el espectro de tumores observado en estas familias, este término ha quedado casi en desuso para hacer referencia al síndrome y en la actualidad se sigue refiriendo a él con el nombre de síndrome de Lynch (Lynch & Murphy, 2016; Lynch *et al*, 2015).

3.2. Características genéticas

El síndrome de Lynch es una condición autosómica dominante que está causada por mutaciones y/o epimutaciones germinales en los genes *MLH1, MSH2, MSH6* o *PMS2*. Estos genes codifican para proteínas cuya función principal es la reparación de desapareamientos del DNA, aunque también están involucradas en la modulación de la recombinación del DNA, la señalización de daño en el DNA y la regulación de la apoptosis.

Vía de reparación de desapareamientos del DNA

Durante la replicación del DNA, la polimerasa puede introducir errores como desapareamientos base-base o pequeños bucles de inserción/deleción (especialmente en secuencias repetitivas como los microsatélites). La proteína MSH2 forma un heterodímero con MSH6 que se desliza por el DNA hasta identificar el error. Una vez posicionado, se produce el reclutamiento del heterodímero formado por MLH1 y PMS2. El tetrámero formado tiene actividad endonucleasa y es capaz de crear incisiones en la hebra de DNA recién sintetizada. Estas incisiones permiten la entrada de EXO1 para reparar el error. Una vez reparado, el complejo se disocia y se reanuda la síntesis de DNA (Figura 11) (Gruber, 2006).

La vía de reparación de desapareamientos del DNA es una vía muy conservada en la evolución, puesto que es fundamental para mantener la integridad del genoma. Si esta vía falla y los errores no son reparados, quedan fijados en el DNA tras la siguiente replicación de la hebra alterada.

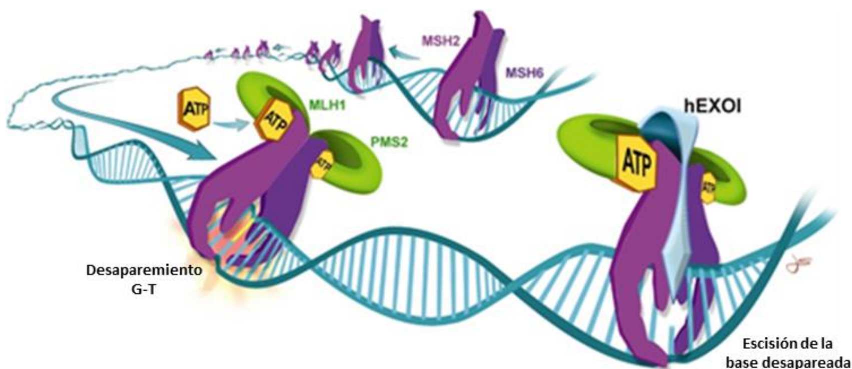


Figura 11: Modelo de reparación de desapareamiento humano.
Ilustración de Jerry Schoendorf (Gruber 2006)

Espectro mutacional del síndrome de Lynch

Según datos de la *Society for Gastrointestinal Hereditary Tumours (InSiGHT) database*, las variantes en los genes reparadores se distribuyen de la siguiente forma: las variantes en *MLH1* representan el 40% de las reportadas, las de *MSH2* alcanzan el 34%, las de *MSH6* el 18% y, por último, las variantes de *PMS2* representan el 8% (Figura 12A) (Peltomäki, 2016).

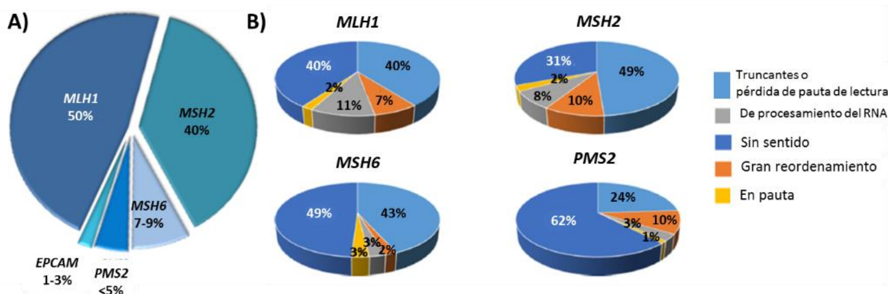


Figura 12: Distribución de variantes en genes MMR.
A) Proporcion de mutaciones patogénicas encontradas en pacientes con síndrome de Lynch. Datos extraídos de Kohlmann and Gruber 2014.
B) Distribuciones de los tipos de variantes germinales no sinónimas en cada gen MMR según los datos depositados en la InSiGHT database. El número total de variantes por gen es de 1104 para *MLH1*, 883 para *MSH2*, 414 para *MSH6* y 197 para *PMS2* (Peltomäki 2016).

La mayoría de las variantes de *MLH1, MSH2* y *MSH6* son truncantes, ya sea por mutaciones que introducen un codón de parada prematuro o por mutaciones que provocan cambios en la pauta de lectura (Figura 12B), aunque la proporción de variantes missense también es elevada (31-62%). La proporción de grandes reordenamientos es variable según el gen, representando hasta el 10% de las variantes en *MSH2* y *PMS2*.

3.3. Epidemiologia

La prevalencia de las mutaciones germinales en los genes MMR se ha calculado tradicionalmente en el contexto de pacientes diagnosticados de cáncer colorrectal o de endometrio. En estas cohortes, el síndrome de Lynch abarca entre el 2 y el 6% de los casos (Buchanan *et al*, 2014; Hampel *et al*, 2005; Moreira *et al*, 2012; Ward *et al*, 2013b; Yurgelun *et al*, 2015). Los datos epidemiológicos actuales, sin embargo, indican que el síndrome de Lynch es más común y menos penetrante de lo que se había pensado hasta el momento. Asimismo, estudios basados en registros poblacionales de Estados Unidos, Canadá y Australia, han permitido estimar la frecuencia de portadores de mutaciones en los genes MMR dentro de la población general, alcanzando el 0,359% (1: 279). La frecuencia de portadores de mutaciones patogénicas subdivididas por genes sería la siguiente: 0.051% (1:1946) para las mutaciones *MLH1*, 0.035% (1:2841) para las mutaciones *MSH2*, 0.132% (1:758) para mutaciones de *MSH6*, y 0.140% (1:714) para *PMS2* (Boland *et al*, 2018; Win *et al*, 2017). Estos datos recientes indican que el síndrome de Lynch es más frecuente en población general de lo que se había estimado previamente, puesto que los genes con variantes patogénicas más prevalentes en la población general (*MSH6* y *PMS2*) confieren un riesgo más moderado de cáncer, especialmente *PMS2* que confiere un riesgo mucho más moderado, y no todos los individuos portadores desarrollan la patología.

En algunas poblaciones aisladas el síndrome de Lynch llega a ser incluso más prevalente debido al efecto de mutaciones fundadoras, como es el caso de la población islandesa (Haraldsdottir *et al*, 2017), la población franco-canadiense (Castellsagué *et al*, 2015) o la población judía Ashkenazi (Foulkes *et al*, 2002; Raskin *et al*, 2011), entre otros. Las mutaciones fundadoras, además, producen un aumento de la probabilidad de individuos con deficiencia constitucional de los genes MMR (CMMRD, del inglés *Constitutional MisMatch Repair Deficiency*) debido a la herencia bialélica de variantes patogénicas en el mismo gen MMR (Baris *et al*, 2016; Durno *et al*, 2015). Aunque se trate de un síndrome raro, el fenotipo en estos individuos es muy agresivo, puesto que la mayor parte de ellos son diagnosticados de cáncer en edades pediátricas. Destacan sobre todo el desarrollo de tumores hematopoyéticos y tumores cerebrales o del sistema nervioso central, además de los tumores del espectro de síndrome de Lynch (gastrointestinales y ginecológicos) (Wimmer *et al*, 2014).

3.4. Características clínicas

Una de las características clínicas más importantes del síndrome de Lynch es la temprana edad a la que estos pacientes desarrollan cáncer. La edad media de diagnóstico en síndrome de Lynch para cáncer colorrectal es de 45 años, 23 antes que la media de la población general (Hampel *et al*, 2005). Además, estos pacientes presentan una carcinogénesis acelerada: un pequeño adenoma colónico puede formar un cáncer colorrectal en tan solo 2-3 años, mientras que para el cáncer colorrectal esporádico su tiempo de evolución es de 6 a 10 años (Lynch *et al*, 2015).

Los tumores colorrectales en el síndrome de Lynch se localizan preferentemente en el colon derecho y es frecuente el hallazgo de múltiples neoplasias sincrónicas (diagnosticadas al mismo tiempo) o metacrónicas (diagnosticadas más de 6 meses después de la resección tumoral). Estos tumores presentan unas características clínico-patológicas específicas que incluyen poca diferenciación, exceso de células mucinosas, presencia de células en anillo de sello, alto grado de infiltración linfocitaria acompañado de una intensa reacción de tipo Crohn y un patrón de crecimiento medular (Lynch *et al*, 2015; Risio *et al*, 1996).

Tabla 1: Incidencia relativa acumulada de cáncer a los 75 años en portadores de variantes patogénicas en genes MMR estratificad por genes. Los intervalos de confianza están escritos entre paréntesis. Los incrementos significativos de la incendecia acumulada (p<0.05) están señalados en negrita.

(Møller et al 2018)

IDC9	Órgano	Incidencia poblacional (%)	Incidencia relativa acumulada (95% IC)			
			pat_MLH1	***pat_MSH2***	***pat_MSH6***	***pat_PMS2***
Cualquier cancer		24,4	3,1 (2,8 a 3,4)	3,3 (2,9 a 3,7)	2,5 (1,7 a 3,2)	2,1 (0 a 4,1)
Por órganos aislados						
152	Duodeno	0,1	**64,7 (27,4 a 102,1)**	20,1 (0,6 a 39,6)	0	0
182	Endometrio	1,6	**26,7 (20,7 a 32,7)**	**35,5 (26,1 a 44,8)**	**28,9 (17,1 a 40,6)**	16,5 (0,5 a 32,4
153	Colon	2,1	**22,3 (18,7 a 25,9)**	**20,2 (15,6 a 24,7)**	**6,8 (1,5 a 12,1)**	0
156	Vesícula y conducto biliar	0,2	**18,7 (6,3 a 31,1)**	8,6 (0 a 25,4)	0	0
183	Ovario	1	**10,1 (4,8 a 15,4)**	**16,9 (5,7 a 28,0)**	13,1 (0 a 31,2)	0
189	Ureter y riñón	1,3	**3,5 (1,2 a 5,9)**	**13,7 (8,2 a 19,2)**	2,3 (0 a 5,4)	0
154	Sigmoide y recto	1,4	**8,4 (5,2 a 11,7)**	**13,0 (7,8 a 18,3)**	3,3 (0 a 6,9)	0
191	Cerebro	0,5	1,9 (0 a 4,8)	10,5 (0,4 a 20,6)	2,9 (0 a 8,4)	0
151	Estómago	0,8	**8,9 (4,4 a 13,4)**	**9,7 (2,3 a 17,0)**	6,6 (0 a 16,4)	0
188	Vejiga	1	**4,1 (1,5 a 6,7)**	**8,1 (2,8 a 13,3)**	8,2 (0 a 16,9)	0
157	Páncreas	0,8	**7,8 (3,3 a 12,3)**	0,6 (0 a 1,9)	1,8 (0 a 5,2)	0
185	Próstata	10	1,7 (0,9 a 2,7)	**3,2 (1,2 a 5,1)**	1,8 (0 a 4,4)	3,8 (0 a 9,6)
174	Mama	9,4	1,3 (0,7 a 1,8)	1,2 (05 a 2,0)	1,4 (0,2 a 2,6)	6,0 (0 a 10,6)
Por regiones anatómicas						
	Ginecologico	2,6	**19,1 (15,6 a 22,7)**	**25,3 (20,1 a 30,4)**	**20,8 (13,3 a 28,2)**	10,1 (0,3 a 20)
	Colorrectal	3,8	**12,1 (10 a 14,2)**	**11,3 (8,7 a 13,9)**	3,9 (0,9 a 7,0)	0
	Tracto gastrointestinal alto	1,9	11,2 (8,2 a 14,3)	**5,4 (2,1 a 8,6)**	3,5 (0 a 7,8)	0
	Tracto urinario	2,3	**3,5 (1,9 a 5,1)**	**10,8 (7,2 a 14,4)**	4,8 (0,7 a 8,8)	0

El síndrome de Lynch también incrementa el riesgo de desarrollar tumores extracolónicos. Dentro del espectro del síndrome se incluirían los tumores endometriales, de ovario, gástricos, de intestino delgado, de tracto biliar, pancreáticos y en vías urinarias (Vasen *et al*, 2013). En la tabla 1, se resumen las incidencias acumuladas de cada tipo de neoplasia por cada gen reparador (Møller *et al*, 2018).

3.5. Características moleculares de los tumores

Las mutaciones germinales en genes MMR confieren a los individuos portadores una predisposición al cáncer con patrón de herencia dominante. Sin embargo, a nivel celular se necesita inactivar ambos alelos del gen para el desarrollo de un tumor (Knudson, 2001). Esta inactivación puede darse por pérdida de heterocigosidad (LOH, del inglés *Loss Of Heterozygosity*), mutaciones somáticas o muy raramente hipermetilación de sus promotores. La inactivación de ambas copias del gen provoca la pérdida de expresión de la proteína MMR y, por tanto, la pérdida de la función reparadora, que conlleva a la acumulación de errores en estas secuencias, provocando inestabilidad de microsatélites (Chung & Rustgi, 2003). La tasa de mutación se incrementa entre 100 y 1000 veces bajo estas condiciones, lo que a su vez aumenta la probabilidad de que otros genes supresores de tumores y oncogenes se vean afectados y se promueva la tumorogénesis (Lynch *et al*, 2010).

La deficiencia reparadora, entendida como la pérdida de expresión de alguna de las proteínas MMR y/o la inestabilidad de microsatélites, es un rasgo típico de los tumores Lynch, aunque no es exclusivo. Como se ha descrito en el apartado 2.2 la deficiencia reparadora está presente en el 10-15% de los tumores colorrectales principalemente debido a metilación somática de *MLH1* como consecuencia de la vía carcinogénica del fenotipo metilador de islas CpG. En el 50-68% de estos casos se ha visto asociado a la mutación V600E de *BRAF* (Gausachs *et al*, 2012; Yamamoto & Imai, 2015). En la figura 13 se esquematizan las dos vías moleculares que pueden dar lugar a tumores con deficiencia reparadora.

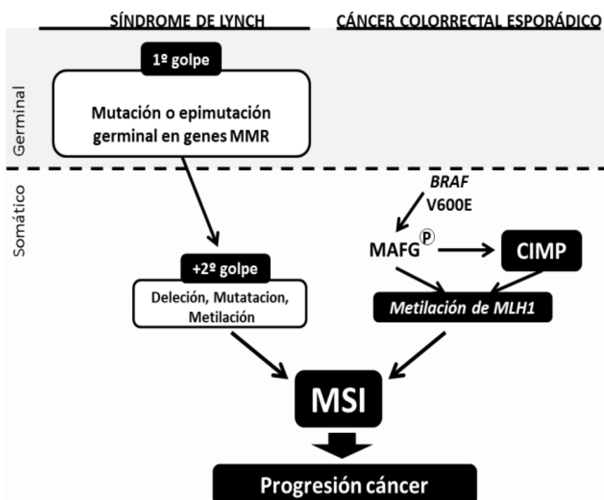


Figura 13: Comparación de los modelos de progresión del cáncer colorrectal con inestabilidad de microsatélites en el síndrome de Lynch y el cáncer colorrectal esporádico.
Adaptado de Yamamoto and Imai 2015.

3.6. Asesoramiento genético y recomendaciones de seguimiento

El asesoramiento genético es un proceso para atender las necesidades y las preocupaciones de personas y de familias en relación con la posibilidad de desarrollar y transmitir una enfermedad genética, como es el caso del síndrome de Lynch. En el contexto del proceso de asesoramiento genético de síndromes de predisposición a cáncerlos asesores genéticos ayudan a los pacientes a entender tanto su diagnóstico y la posible evolución de la enfermedad, como el patrón de herencia de ésta y la posible necesidad de realizar estudios genéticos predictivos a otros miembros de la familia. También se informa a los pacientes sobre sus opciones para reducir el riesgo de aparición del cáncer, medidas para el diagnóstico precoz, el tratamiento y posibles líneas de investigación en el campo.

Puesto que los tumores en los pacientes con síndrome de Lynch se desarrollan en edades tempranas y presentan una carcinogénesis acelerada, las guías de vigilancia publicadas recomiendan un seguimiento estricto en individuos portadores. La mayor parte de las organizaciones proponen un seguimiento por colonoscopia cada 1-2 años a partir de los 20-30 años del individuo, control ginecológico anual y la consideración de histerectomía y la salpingooforectomía bilateral profiláctica en mujeres una vez cumplidos los deseos genésicos (Giardiello *et al*, 2014; Syngal *et al*, 2015; Vasen *et al*, 2013). En los últimos años, además, se recomendado el uso quimiopreventivo de aspirina en pacientes con síndrome de Lynch, debido a los estudios que han documentado una reducción del riesgo de cáncer colorrectal tras su administración a largo plazo (Burn *et al*, 2011). Estas opciones de vigilancia y prevención han demostrado ser muy eficaces para reducir la morbilidad y la mortalidad de los cánceres asociados con el síndrome de Lynch (Hampel, 2016)

4. Diagnóstico molecular del síndrome de Lynch

4.1. Criterios de selección

Debido al riesgo intrínseco de desarrollar cáncer que tienen los individuos con síndrome de Lynch, la identificación de individuos y familias afectas es clave a la hora de aplicar un protocolo de vigilancia preventivo (Järvinen *et al*, 2009; Palomaki *et al*, 2009; Stupart *et al*, 2009). Sin embargo, el principal desafío en el manejo del síndrome de Lynch es su identificación.

Ante esta dificultad, en 1990 el *International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer* desarrolló los criterios clínicos de Ámsterdam para facilitar su detección en base a la historia personal y familiar de cáncer colorrectal del paciente (Tabla 2) (Vasen *et al*, 1991). Ocho años más tarde, se hizo una revisión de estos criterios de forma que contemplasen también la aparición de tumores extracolónicos, definiéndose así los criterios de Ámsterdam II (Tabla 2) (Vasen *et al*, 1999). Ambos criterios de Ámsterdam resultan muy específicos para la detección de familias con síndrome de Lynch, sin embargo, estos resultaron ser demasiado restrictivos y poco sensibles.

Tabla 2: Criterios Amsterdam I y II

Amsterdam I

	Se deben cumplir todos los siguientes criterios:
1	Al menos 3 familiares deben haber tendio cáncer colorectal, 1 de ellos debe ser familiar de primer grado de los otros 2.
2	Al menos 2 generaciones sucesivas deben estar afectas
3	En uno de los familiares, el cáncer colorrectal debe haber sido diagnosticado antes de los 50 años
4	La poliposis familiar adenomatosa debe ser excluída

Amsterdam II

	Se deben cumplir todos los siguientes criterios:
1	Debe haber al menos 3 familiares con un cáncer de los asociados a cáncer colorrectal hereditario no polipósico (cáncer colorrectal, cáncer de endometrio, intestino delgado, uréter o pelvis renal)
2	Uno debe ser familiar de primer grado de los otros 2
3	Al menos 2 generaciones sucesivas deben estar afectadas
4	Al menos 1 debe haber sido diagnosticado antes de los 50 años
5	La poliposis familiar adenomatosa debe ser excluída en los casos de cáncer colorrectal (si los hubiese)
6	Los tumores deben ser verificados mediante un estudio patológico

El pequeño tamaño de las familias y la ausencia en ocasiones de historia familiar hizo que en 1997 se definiesen los criterios clínicos de Bethesda (Tabla 3). Estos criterios clínicos tenían en cuenta la correlación entre síndrome y los tumores con inestabilidad de microsatélites, por lo que el análisis de la inestabilidad de microsatélites pasó a ser la técnica de referencia para el cribado de pacientes. En 2004, se modificaron los criterios de Bethesda por los criterios Bethesda "revisados" (Tabla 3) con el fin de

maximizar su especificidad sin perder sensibilidad. Esta mejora fue posible gracias a la consideración durante el proceso de cribado del patrón de inmunohistoquímica referido a la pérdida de las proteínas MMR en el tumor (Laghi *et al*, 2004; Umar *et al*, 2004).

Tabla 3: Criterios Bethesda y criterios Bethesda modificados.

Bethesda

Se debe cumplir al menos uno de los siguientes criterios:	
1	Individuos con cáncer que cumplen criterios Amsterdam
2	Individuos con 2 tumores de los asociados a cáncer colorrectal hereditario no polipósico, inlcuyendo cáncer colorrectal sincrónico o metacrónicos u otros tumores extracolónicos.
3	Individuos con cáncer colorrectal y un familiar de primer grado con cáncer colorrectal y/o tumor extracolónico asociado a cáncer colorrectal hereditario no polipósico y/o un adenoma colorrectal, 1 de los cánceres diagnosticado antes de los 45 años o de los 40 para el adenoma.
4	Individuos con cáncer colorrecal o de endometrio deiagnosticados antes de los 45 años
5	Individuos con cáncer de colon derecho con un patrón histopatológico no diferenciado
6	Individuos con cáncer colorrectal con células en anillo de sello
7	Individuos con adenomas diagnosticados antes de los 40 años.

Bethesda revisados

Se debe cumplir al menos uno de los siguientes criterios:	
1	Individuos diagnosticados de cáncer colorrectal antes de los 50 años
2	Presencia de tumores colorrectales sincrónicos o metacrónicos u otros tipos de tumores asociados a cáncer colorrectal hereditario no polipósico independientemente de la edad.
3	Cáncer colorrectal con histología de inestabilidad de microsatélites alta (presencia de limfocitos infiltrantes, reacción de Crohn positiva, mucinoso/células en anillo de sello o patrón de crecimiento medular) en paciente menor de 60 años
4	Cáncer colorrectal en uno o más familiares de primer grado con tumores asociados a cáncer colorrectal hereditario no polipósico, con uno de los tumores diagnosticado antes de los 50 años
5	Cáncer colorrectal diagnosticado en 2 o más familiares de primer o segundo grado con tumores asociados a cáncer colorrectal hereditario no polipósico, independientemente de la edad

En 2009, el *Evaluation of Genomic Applications in Practice and Prevention Working Group* propuso el “cribado universal” con el fin de aumentar al 100% la tasa de detección de individuos con síndrome de Lynch. El cribado universal propone que, de forma prospectiva, todos los tumores colorrectales y de endometrio sean cribados por análisis de inestabilidad de microsatélites e inmunohistoquímica de las proteínas MMR en el tumor (Batte *et al*, 2014; Hampel, 2010; Hampel & de la Chapelle, 2011; Hampel *et al*, 2008). Aunque muchos grupos han apoyado el cribado universal debido a su impacto sobre la bajada de la morbilidad y mortalidad de las familias con síndrome de Lynch, este tipo de cribado supone un alto coste. Por este motivo, el mismo año se propusieron las “recomendaciones de Jerusalén” que sugieren la implementación un punto de corte a los 70 años, según el cual, solo se recomienda el cribado de pacientes menores a dicha edad (Boland & Shike, 2010).

En 2012, Moreira y colaboradores realizaron un análisis de datos agrupado de carácter internacional para determinar la estrategia más sensible y eficiente para la identificación de pacientes con síndrome de Lynch entre los pacientes diagnosticados de cáncer colorrectal. Determinaron que la estrategia más coste-efectiva consiste en la aplicación de unos criterios combinados o “estrategia selectiva” para cribar los pacientes según

cumplan los criterios de Jerusalén o al menos un criterio revisado de Bethesda (Figura 14). Aunque el porcentaje de pacientes no identificados con esta estrategia es del 5%, su aplicación conlleva una reducción del 34,8% en el número de pacientes en los que debe efectuarse el análisis del sistema de reparación del tumor, y del 28,6% en el de aquellos en los debe realizarse el estudio en línea germinal (Moreira *et al*, 2012). Aunque la estrategia selectiva tiene una gran solidez debido a su base poblacional, sus limitaciones residen en la falta de contemplación en su algoritmo de los tumores extracolónicos.

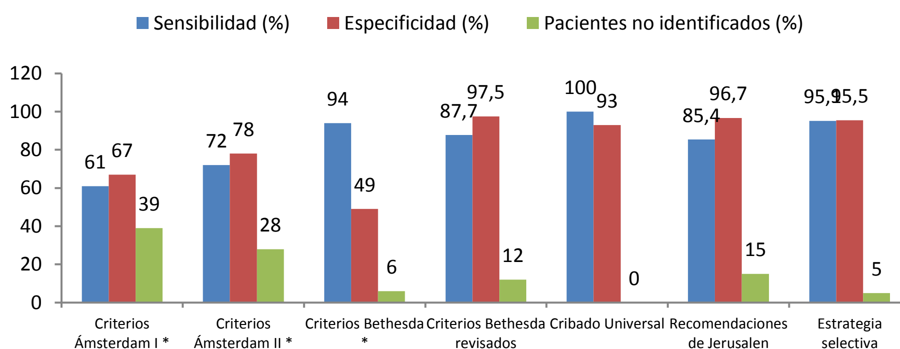


Figura 14: Comparación de la sensibilidad, especificidad y porcentaje de pacientes no identificados seguún los criterios/estratègies propuestas hasta el momento.

**Los datos de los criterios Amsterdam I y II y de los criterios Bethesda hacen referencia exclusivamente a la identificación de portadores de mutación en MLH1 y MSH2.*

4.2. El algoritmo diagnóstico

Teniendo en cuenta las anteriores premisas, el algoritmo diagnóstico para la identificación de pacientes con síndrome de Lynch en pacientes que cumplen criterios clínicos, seleccionados por cribado poblaciónal o con criterios de Jerusalen, se inicia con el estudio de inestabilidad de microsatélites y/o inmunohistoquímica de las proteínas MMR en tejido tumoral, para la selección de pacientes candidatos a un estudio genético.

En el caso de que por inmunohistoquímica se detecte la pérdida de MLH1 se realiza un análisis de hipermetilación de *MLH1* o de la mutación *BRAF* V600E en el tumor, para descartar la posibilidad de que se trate de un tumor esporádico con fenotipo metilador de islas CpG. Si el resultado es negativo, se procede a su estudio genético. Sin embargo, si el estudio es positivo, puede ser necesario descartar la presencia de epimutación constitucional en *MLH1* (Hitchins & Ward, 2009). En nuestro centro, se realiza estudio de metilación en sangre a los pacientes altamente sugestivos, es decir, con tumores metilados a una edad menor de 50 años o con múltiples tumores metilados (el primero antes de los 60).

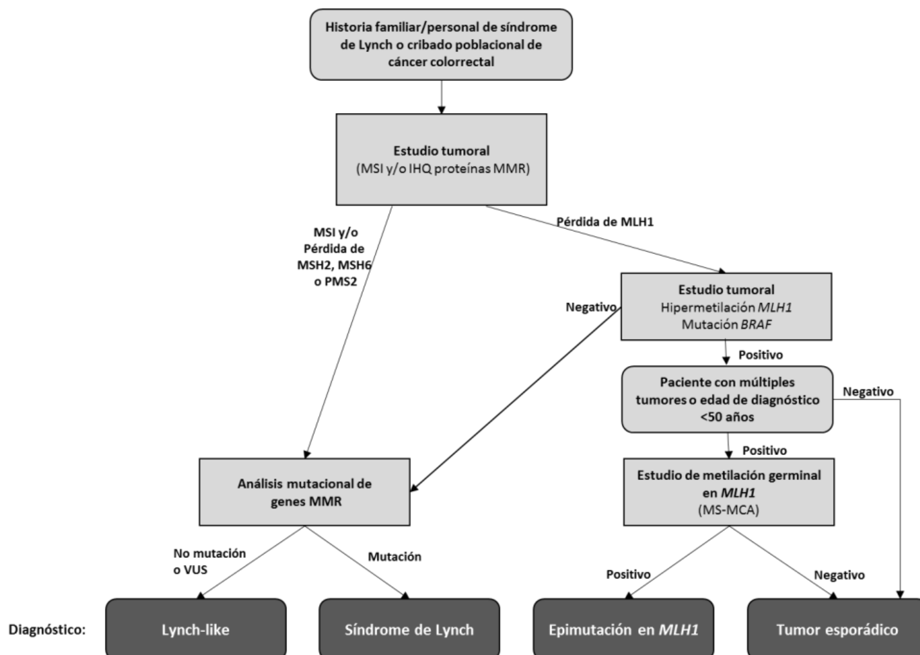


Figura 15: Algoritmo diagnóstico propuesto para la identificación del síndrome de Lynch.

Cribado molecular de los tumores

Análisis de inestabilidad de microsatélites

Una de las principales características de deficiencia MMR es la inestabilidad de microsatélites. Entendemos como inestabilidad a la diferencia en el patrón de repeticiones de un microsatélite cuando se compara el patrón obtenido de la amplificación de DNA tumoral frente al proveniente de DNA de tejido normal apareado del mismo paciente (Figura 16) (Gruber, 2006).

En 1997, el *National Cancer Institute* de Estados Unidos recomendó el uso del panel de Bethesda para el análisis de la inestabilidad de microsatélites. Este panel está compuesto por 5 microsatélites: 2 mononucleótidos (BAT26 y BAT25) y 3 dinocleótidos (D2S123, D5S346 y D17S250) (Boland *et al*, 1998). Según la cantidad de marcadores inestables, los tumores se clasifican en: "tumores con alta inestabilidad" cuando dos o más marcadores están alterados, "tumores con baja inestabilidad" cuando solo uno de los marcadores es inestable o "tumores estables", si ningún marcador aparece alterado.

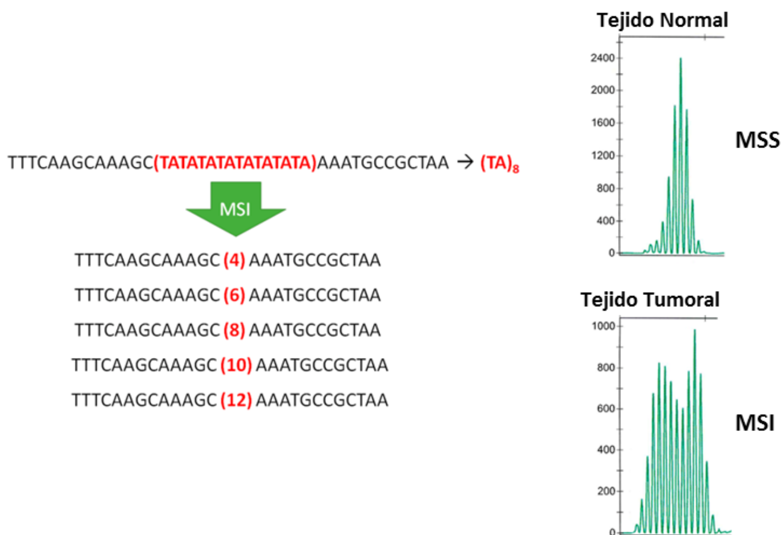


Figura 16: Esquema sobre la inestabilidad de microsatélites (MSI). A la izquierda, se ilustra la variación de un microsatélite compuesto por dinucleótidos de timina y adenina como consecuencia de la inestabilidad de microsatétlites. A la derecha, la visualización de este microsatélite mediante electroferograma: el DNA de tejido normal muestra un patrón de estabilidad de microsatélites (MSS) mientras que en la parte superior se muestra que el tejido tumoral emparejado muestra un patrón de inestabilidad de microsatélites (MSI). (Lázaro et al 2018)

Sin embargo, poco después se observó que los marcadores de dinucleótidos muestran una sensibilidad y especificidad menores que los de mononucleótidos (Pedroni *et al*, 2007; Perucho, 1999; Suraweera *et al*, 2002). Por esta razón, se propuso un panel cuasimonomórfico que consta de cinco marcadores mononucleótidos: BAT26, BAT25, NR21, NR22, y NR24. Este panel es más sensible para detectar inestabilidad de microsatélites, especialmente en tumores deficientes en MSH6 y, además, permite el estudio de tumores sin tejido normal aperado (Buhard *et al*, 2004; Ebinger *et al*, 2006; Hegde *et al*, 2013; Suraweera *et al*, 2002; Umar *et al*, 2004; You *et al*, 2010).

Inmunohistoquimica

La inmunohistoquímica (IHQ) de proteínas reparadoras MMR es la técnica más utilizada para analizar deficiencia reparadora en el tumor, debido a que está considerada como una técnica muy sensible (80.8-100%) y específica (80.5-91.9%) (Snowsill *et al*, 2017). Desde una perspectiva clínica, la identificación de la pérdida de expresión de alguna de las proteínas o heterodímeros MMR servirá de guía en los posteriores estudios genéticos cuyo fin es encontrar la mutación causal responsable del síndrome de Lynch en el paciente. Esto es posible debido a que las mutaciones en cada uno de los genes MMR suelen asociarse a un patrón de pérdida de expresión específico. Las proteínas MMR se asocian formando 2 heterodímeros, MLH1-PMS2 y MSH2-MSH6, que estabilizan las proteínas que los componen y evitan su degradación. Sin embargo, este sistema tiene cierta redundancia puesto que tanto MLH1 como MSH2 pueden unirse a otras proteínas para realizar funciones complementarias, por lo que su expresión sería estable en ausencia de su proteína heterodimérica principal (PMS2 o MSH6, respectivamente) (Gruber, 2006). De este modo, mutaciones en *PMS2* o *MSH6* provocan una pérdida aislada de la proteína afecta, mientras que mutaciones o epimutaciones en *MLH1* o *MSH2* provocan la pérdida de la proteína afecta y de su proteína heterodimérica (Figura 17).

Sin embargo, cuando se obtiene un resultado negativo para el dímero MLH1-PMS2, debe considerarse la posibilidad de que se trate de un tumor esporádico con hipermetilación somática de *MLH1* debido al fenotipo metilador de islas CpG. Para ello, las pruebas de detección de mutaciones en *BRAF* y de metilación de *MLH1* se han propuesto como pruebas de refinamiento de candidatos previas a la secuenciación en línea germinal de *MLH1*.

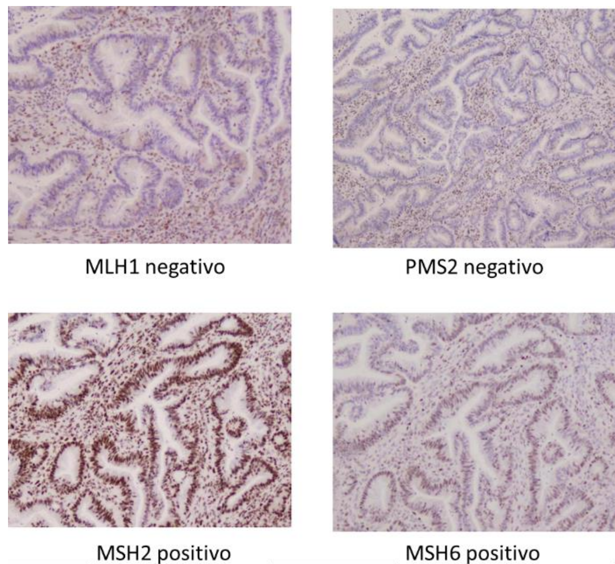


Figura 17: Análisis de expresión de las proteínas reparadoras de desapareamientos del DNA mediante inmunohistoquímica. Este ejemplo es de un cáncer colorrectal con inactivación de MLH1: se puede observar la tinción nuclear en las proteínas MSH2 y MSH6 pero no es MLH1 y PMS2.
(Gruber 2006)

Detección de mutaciones somáticas en *BRAF*

BRAF es un proto-oncogen que codifica una serina/treonina proteín quinasa. La mutación somática más prevalente es la c.1799T>A, que provoca un cambio de la valina 600 a un ácido glutámico (V600E), y que puede encontrarse en el 15% de los tumores colorrectales. Dentro de los tumores colorrectales esporádicos, la mutación V600E de *BRAF* se detecta con mayor frecuencia en los tumores con inestabilidad de microsatélites debido a hipermetilación somática de *MLH1* por el fenotipo metilador de islas CpG (Domingo *et al*, 2004; Oliveira *et al*, 2003; Rajagopalan *et al*, 2002; Yamamoto & Imai, 2015). La presencia de la mutación *BRAF* V600E puede ofrecer un criterio de exclusión para la realización de análisis mutacional de genes MMR en línea germinal, ya que en este tipo de pacientes su prevalencia es muy baja, aunque se han documentado algunos casos Lynch con tumores *BRAF* mutados (Gausachs *et al*, 2012; Kim *et al*, 2008; Parsons *et al*, 2012; Walsh *et al*, 2009)

Las mutaciones somáticas en *BRAF* se pueden detectar usando diferentes técnicas, tales como secuenciación directa, extensión alelo-específica, digestión con enzimas de restricción o PCR en tiempo real (Hegde *et al*, 2013; Pineda *et al*, 2010). La sensibilidad

de estos análisis varía del 96-100% y su especificidad, referida a la identificación de tumores Lynch con pérdida de MLH1, es del 34% (Gausachs *et al*, 2012; Pérez-Carbonell *et al*, 2010).

Técnicas de análisis de metilación en el promotor de *MLH1*

El análisis de metilación del promotor de *MLH1* también se puede usar para discriminar entre tumores asociados a síndrome de Lynch y esporádicos. Los estudios comparativos de la coste-efectividad del estudio mutacional de *BRAF* y el estudio de hipermetilación somática de *MLH1* determinan que, el estudio de hipermetilación, muestra mayor especificidad en la selección de pacientes que son candidatos para el análisis de la línea germinal de *MLH1* (Cenin *et al*, 2018; Gausachs *et al*, 2012). Sin embargo, diferentes rangos de especificidad han sido reportados según la técnica utilizada, el criterio utilizado para la selección de pacientes y el umbral definido (Gausachs *et al*, 2012; Moreira *et al*, 2015; Newton *et al*, 2014; Pérez-Carbonell *et al*, 2010).

La metilación en la región C de Deng (Figura 18) es la que mejor correlaciona con el silenciamiento transcripcional del gen, seguidamente de la región D. Es por este motivo los análisis de metilación en *MLH1* actuales se centran en el estudio de estas regiones (Deng *et al*, 1999; Gausachs *et al*, 2012; Pérez-Carbonell *et al*, 2010; Pineda *et al*, 2012).

Figura 18: Localización de las regiones definidas por Deng et al 1999 para el estudio del promotor de *MLH1*. Los números hacen referencia a la posición de los nucleótidos señalados con respecto al codón iniciador de *MLH1*.

Existen diversas técnicas para el análisis de la metilación. En la mayoría de ellas, el DNA requiere un tratamiento previo con bisulfito sódico. El bisulfito sódico provoca la desaminación de las citosinas no metiladas, convirtiéndolas en uracilos. Sin embargo, en las citosinas metiladas, la presencia de un grupo metilo en el carbono 5 no permite esta conversión (Clark *et al*, 1994; Frommer *et al*, 1992). Durante el proceso de amplificación, en las posiciones donde ha habido conversión se insertan timinas mientras que, en las que no se han convertido por estar metiladas, se conservarán las citosinas. De este modo, las citosinas metiladas y no metiladas se pueden distinguir de acuerdo con los cambios de secuencia.

Las propiedades otorgadas por estos cambios de secuencia se han utilizado para el diseño de diferentes técnicas de detección de metilación dirigidas al estudio de *MLH1*. Entre las técnicas más utilizadas en los laboratorios de diagnóstico molecular se encuentran:

La **pirosecuenciación** se basa en la detección de la liberación de pirofosfato cuando se incorporan, de forma secuencial, los nucleótidos de la secuencia analizada. En un ensayo de metilación mediante pirosecuenciación, el dGTP y dATP son incorporados para interrogar el estado epigenético de una misma posición. La intensidad lumínica obtenida del uso del pirofosfato por la luciferasa en la incorporación de cada uno de los dos nucleótidos permite cálculo del porcentaje de metilación en cada sitio CpG. La sensibilidad de esta técnica se sitúa en un 5% de metilación, por lo que es muy utilizada para análisis de metilación en tumores, debido a la elevada heterogeneidad de este tipo de muestras (Kurdyukov & Bullock, 2016).

El **análisis de la curva de fusión específica de metilación** (o MS-MCA, del inglés *Methylation Specific – Melting Curve Analysis*), está basado en la diferencia de temperatura de desnaturalización de las hebras de DNA. En las moléculas que mantienen las citosinas tras la conversión por bisulfito (metiladas), la temperatura de fusión será más elevada, puesto su asociación con la guanina de la hebra complementaria genera tres puentes de hidrógeno. En el caso de las citosinas convertidas a timina, la temperatura de fusión es más baja puesto que solo se generan dos puentes de hidrógeno para su unión con la adenina complementaria. Dependiendo del porcentaje de metilación de la muestra, se generan unos perfiles específicos de fusión que permiten su estudio de forma semicuantitativa.

El DNA convertido con bisulfito también puede utilizarse para secuenciar la región de interés tras su amplificación. Dependiendo del diseño de los cebadores y de su especificidad o no para amplificar del DNA metilado, distinguimos entre varias técnicas:

Secuenciación con bisulfito. El diseño de los cebadores permite la amplificación tanto del DNA metilado como no metilado, al no incluir ningún sitio CpG en su diseño o incluir nucleótidos degenerados. La interpretación de los resultados obtenidos de la secuenciación directa de estos productos de amplificación es a menudo problemática, sobre todo en poblaciones mixtas de células con diferente metilación donde las señales de metilación pueden diluirse (Kurdyukov & Bullock, 2016). Para solucionar este problema, se han desarrollado modificaciones de la técnica que permiten la discriminación de los alelos metilados de los no metilados.

> La **secuenciación clonal con bisulfito** incluye el clonaje y transformación en una célula competente del producto amplificado previamente a su secuenciación. Esta técnica permite conocer el estado epigenético de cada sitio CpG dentro de una única molécula de DNA.

El **análisis combinado de restricción con bisulfito** (COBRA, del inglés *COmbined Bisulfite Restriction Analysis*), se basa en la pérdida o detección de nuevos sitios de restricción generados en la secuencia debido a su conversión con bisulfito (Eads & Laird, 2002). La separación por electroforesis en gel permite la visualización de las bandas digeridas y sin digerir informando de la presencia o ausencia de metilación en la muestra. También se puede realizar una cuantificación relativa con el cálculo del producto escindido frente a la cantidad total del producto de PCR.

En la **PCR específica de metilación** (MSP, del inglés Methylation-Specific PCR) se tienen en cuenta las modificaciones en la secuencia creadas por el bisulfito. De este modo, se diseñan dos pares de cebadores: un par que favorece la amplificación del DNA metilado y otro de DNA no metilado. Dependiendo del tipo de estudio que se quiere realizar, se pueden obtener resultados cualitativos de metilación, al correr los productos de PCR en un gel de agarosa (donde se observaría la presencia/ausencia de bandas), o cuantitativos, cuando el análisis se realiza por qPCR, donde la metilación relativa se calcula en función de la diferencia de valores Ct entre ambas reacciones (Kurdyukov & Bullock, 2016).

Sin embargo, también existen técnicas que no necesitan una conversión del DNA con bisulfito sódico. Entre estas técnicas, destaca la **amplificación dependiente de la ligación de sondas multiplexadas específicas de metilación** (MS-MLPA, del inglés *Methylation-Specific Multiplex Ligation-dependent Probe Amplification*). Esta técnica usa una endonucleasa de restricción (HhaI) sensible a metilación, que actúa tras la ligación de las sondas. De este modo, si el sitio CpG incluido en una sonda concreta está metilado, se impide la digestión. Las sondas no digeridas se amplifican por PCR y se puede observar un pico en el electroferograma. La cuantificación de la metilación en una muestra es relativa al patrón de picos obtenidos de las muestras control analizadas (Zhang & Jeltsch, 2010). En la actualidad existen varios kits comerciales para detectar metilación, aunque el más relevante en el presente estudio es el SALSA MS-MLPA ME011 MMR (MRC Holland) permite la identificación de la metilación en 6 promotores de genes MMR (*MLH1, MSH2, MSH6, MSH3, MLH3, PMS2*) al mismo tiempo. La sensibilidad de detección de metilación en este kit se sitúa alrededor del 10% (Gausachs *et al*, 2012) y es, junto con la pirosecuenciación, una de las técnicas más utilizadas en la rutina diagnóstica para los estudios de metilación en *MLH1*.

Aunque las técnicas descritas son las más comúnmente utilizadas en los laboratorios de diagnóstico molecular, existen muchas otras técnicas para los estudios de metilación. En la figura 19, se sintetizan los métodos actualmente disponibles a través de un algoritmo de selección para la técnica más adecuada según las necesidades del usuario.

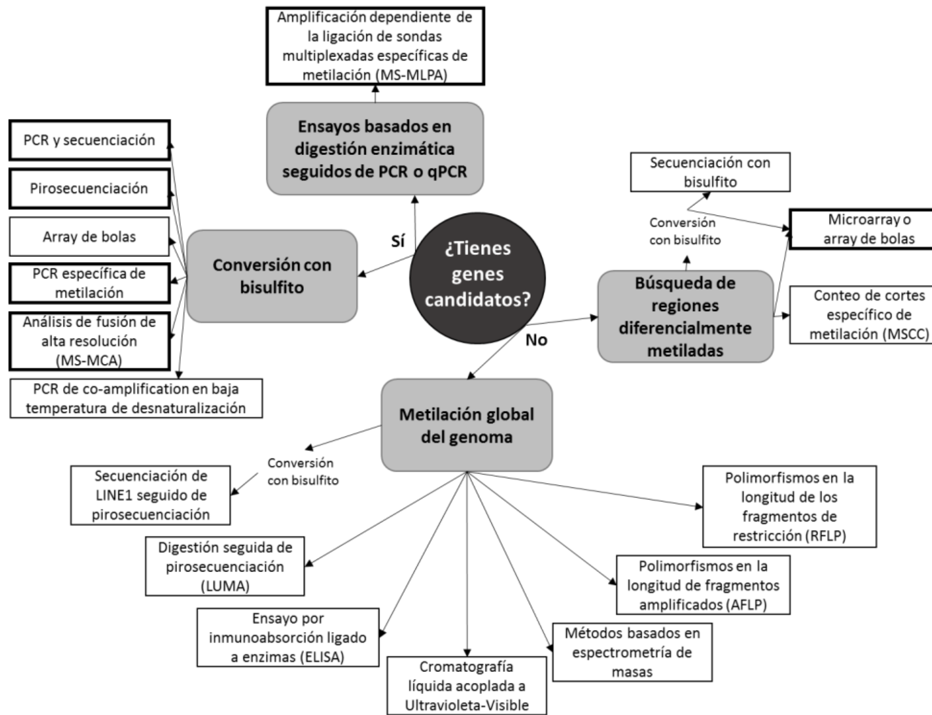


Figura 19: Algoritmo para selección de análisis de la metilación del DNA. Las técnicas destacadas en negrita son algunas de las técnicas utilizadas en el presente trabajo.

Modificado de Kurdyukov and Bullock 2016.

Diagnóstico molecular germinal

Tras el cribado molecular de los tumores, aquellos pacientes con tumores con deficiencia reparadora, incluyendo los casos con pérdida de *MLH1* donde se haya descartado previamente la presencia de hipermetilación somática y/o la mutación de *BRAF*, deben someterse a análisis mutacional de genes MMR, como son su secuenciación y el análisis de grandes reordenamientos. De forma tradicional, el estudio de los genes MMR puede estar guiado por el patrón de inmunohistoquímica obtenido del estudio del tumor. En los últimos años el desarrollo de estrategias basadas en la secuenciación de nueva generación ofrece el análisis completo de genes MMR o incluso otros genes de predisposición a cáncer de forma simultánea.

Las técnicas principales utilizadas para el análisis de genes reparadores en línea germinal incluyen:

Secuenciación Sanger

La secuenciación Sanger está considerada como la técnica de referencia para el diagnóstico molecular germinal del síndrome de Lynch. Con fines diagnósticos, se considera necesario el estudio de toda la región codificante del gen y de sus límites intrón-exón. También se recomienda la secuenciación de los promotores, ya que se han descrito mutaciones patogénicas en los promotores de genes reparadores (Liu *et al*, 2016b).

Secuenciación de nueva generación

La secuenciación de nueva generación (NGS, del inglés *Next Generation Sequencing*) está siendo incorporada paulatinamente en los laboratorios de diagnóstico molecular gracias a que su precio cada vez es más competitivo (Pritchard & Grady, 2011; Pritchard *et al*, 2012; Stoffel *et al*, 2018).

La NGS ofrece tres enfoques distintos según las regiones que se deseen abarcar. De este modo distinguimos entre el enriquecimiento dirigido de un conjunto de genes (también llamado múltiplex o panel de genes), secuenciación del exoma completo y secuenciación del genoma completo.

En la actualidad, el análisis de paneles es el más utilizado en la práctica clínica (Figura 20)(Lázaro *et al*, 2018). Existen diversos métodos de enriquecimiento de los genes de interés, que pueden estar basados en PCR, hibridación o circularización. Este último método de enriquecimiento está pensado especialmente para regiones pequeñas o medianas y las moléculas diana se pueden seleccionar y circularizar en una sola reacción para posteriormente amplificarse por PCR o por hibridación (Moorthie *et al*, 2011).

Para la secuenciación pueden utilizarse plataformas de lectura largas o cortas dependiendo de la longitud de los fragmentos que se quieran analizar (Hegde *et al*, 2013; Rehm *et al*, 2013). La secuenciación por secuenciadores de alto rendimiento ofrece la posibilidad de escoger entre secuenciación en un solo sentido o en dos sentidos de forma emparejada. La secuenciación en doble sentido tiene la ventaja de que aumenta la cobertura y la rigurosidad de la secuenciación, particularmente en regiones repetitivas. Además, permiten la detección de variantes estructurales en algunos casos (Rehm *et al*, 2013).

El proceso de secuenciación en NGS es masivo, de tal forma que se pueden secuenciar a la vez entre cientos de miles y miles de millones de focos en cada reacción. Las reacciones de secuenciación son graduales y consisten en: la adicción de nucleótidos, la identificación del nucleótido incorporado en cada fragmento en proceso de secuenciación y un lavado, donde se eliminan todos los nucleótidos marcados no

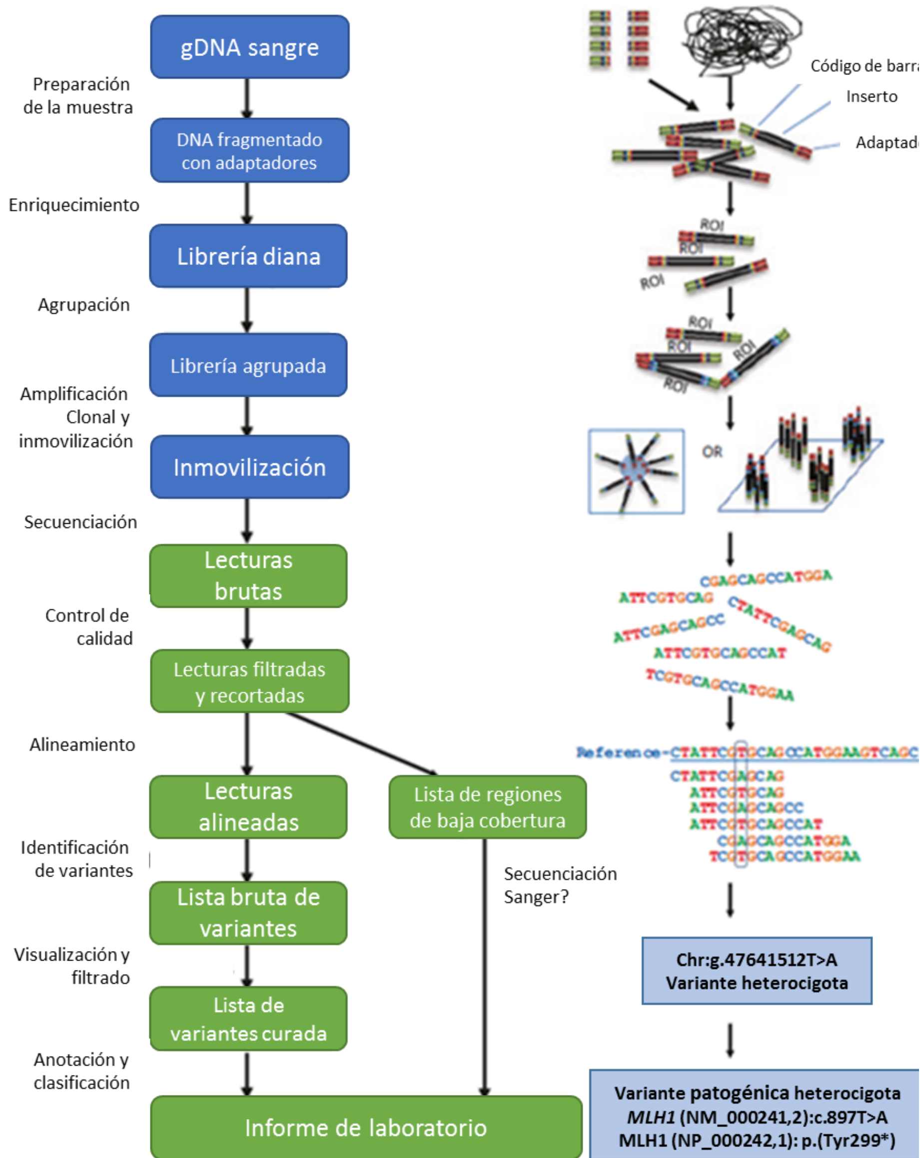


Figura 20: Esquema del protocolo básico de NGS para el diagnóstico. A la izquierda, los pasos principales de NGS desde la obtención de la muestra hasta la elaboración del informe (azul para el trabajo de laboratorio, verde para el trabajo bioinformática). A la derecha, representación de las muestras y productos en cada paso. (Lázaro et al 2018)

incorporados y los agentes químicos de la reacción antes de continuar con la siguiente (Mardis, 2013).

Entre las limitaciones de la técnica se encuentran la necesidad de validar los hallazgos mediante secuenciación Sanger y la necesidad de utilizar otras tecnologías en regiones con baja cobertura y en casos de deleciones y/o duplicaciones de exones, puesto que no pueden ser detectados (Hegde *et al*, 2013).

Detección de grandes reordenamientos

Para la identificación de deleciones o duplicaciones únicas o de múltiples de exones la técnica más sencilla y robusta es la amplificación dependiente de la ligación de sondas multiplexadas (MLPA, del inglés *Multiplex Ligation-dependent Probe Amplification*), en la cual la amplificación de las diferentes sondas hibridadas en el DNA permite su visualización en un electroferograma y su cuantificación relativa. Otras técnicas de cuantificación son la PCR cuantitativa en tiempo real de la región de interés y la PCR múltiple cuantitativa de fragmentos fluorescentes cortos, que también se ha demostrado que son útiles para la detección de variaciones en el número de copia.

Cribado de epimutaciones constitucionales

Para la identificación de pacientes portadores de epimutación es necesaria la identificación de metilación en tejido normal. Cuando se trata de pacientes con sospecha de epimutación en *MLH1,* el estudio de metilación se podrá realizar en DNA procedente de sangre periférica. En pacientes sospechosos de epimutación en *MSH2*, puesto que la metilación muestra mosaicismo tisular, la presencia de metilación solo es validable en tejido normal epitelial. Las técnicas de análisis de metilación más utilizadas son las anteriormente descritas para el cribado de metilación en tumores. El estudio de variaciones en el número de copia de la región 3’ de *EPCAM* mediante MLPA es útil para la identificación de portadores de epimutación en *MSH2*, debido a la asociación de este tipo de deleciones con la hipermetilación del promotor de *MSH2* en tejidos que expresan *EPCAM*.

4.3. Evaluación de la patogenicidad de variantes en genes reparadores

El diagnóstico de síndrome de Lynch depende de la detección en línea germinal de una mutación patogénica en uno de los genes reparadores (Vasen *et al*, 2013). Sin embargo, el hallazgo de variantes de significado desconocido, que representa el 30% de todas las variantes identificadas en individuos con sospecha de síndrome de Lynch (Figura 21), no permite dar un diagnóstico a estos pacientes (Peltomäki, 2016; Thompson *et al*, 2013). La inclusión de las tecnologías de nueva generación en la rutina diagnóstica ha provocado que el número de este tipo de variantes sea cada vez mayor (Howarth *et al*, 2015; Rohlin *et al*, 2017; Susswein *et al*, 2015; Yurgelun *et al*, 2015)

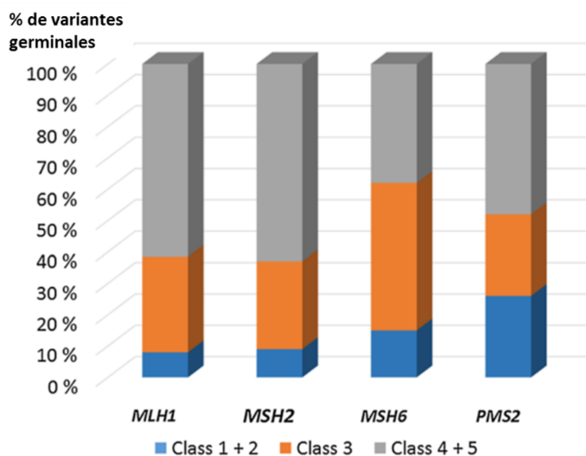


Figura 21: Distribución de los tipos de variantes reportadas en la INSIGHT database ordenador por genes MMR. El análisis incluye 932 variantes genéticas de *MLH1*, 842 para *MSH2*, 449 para *MSH6* y 137 para *PMS2*. (Peltomäki 2016)

Se han propuesto varios esquemas para la clasificación de variantes genéticas en el entorno clínico. Todos ellos están destinados a diferenciar las variantes de alta penetrancia de las de baja o neutra penetrancia, pero no consideran la existencia de variantes de riesgo intermedio. El sistema de clasificación de la *International Agency for Research on Cancer* (IARC), facilita la categorización estandarizada de las variantes en 5 clases (classe 5-patogénicas, clase 4-probablemente patogénicas, clase 3-variantes de significado desconocido o VUS, clase 2-probablemente neutras y clase 1-neutras), que pueden vincularse a datos cuantitativos validados y / o la interpretación de datos cualitativo, y que están vinculadas a recomendaciones clínicas de manejo y vigilancia. En el año 2014, el *InSiGHT Variant Interpretation Committee* desarrolló un esquema de clasificación estandarizado a partir del sistema de clasificación IARC para la clasificación variantes de genes MMR basado en múltiples líneas de evidencia, que incluyen datos clínicos y funcionales (Figura 22) (Nykamp *et al*, 2017; Richards *et al*, 2015; Thompson *et al*, 2013)y que sirvió para reclasificar dos tercios de las variantes informadas hasta el momento.

Clase 5 (patogénica)	Clase 4 (probablemente patogénica)	Clase 3 (variante de significado desconocido)	Clase 2 (probablemente neutra)	Clase 1 (neutra)
Truncante o pérdida de pauta de lectura o Gran deleción o Gran duplicación con confirmación de codificar para un tránscrito con pérdida en la pauta de lectura o Inactivación completa del alelo con la variante por aberraciones de empalme o PP>0,99 o 4 puntos de evidencia: Pérdida de función o CMMRD o en haplotipos diferentes Cosegregación con la enfermedad Al menos 2 tumores con fenotipo molecular de síndrome de Lynch No reportada en el proyecto 1000 Genomes	Sitio canónico de empalme pero sin estudio o PP= 0,95-0,99 o 2 puntos de evidencia: Pérdida de función o CMMRD o en haplotipos diferentes Cosegregación con la enfermedad Al menos 2 tumores con fenotipo molecular de síndrome de Lynch	PP= 0,05-0,949 o Evidencias insuficientes para clasificar	AF mayor o igual al 1% en grupos ancestrales específicos o Variante sinónima o intrónica sin aberración del mRNA o PP= 0,001-0,049 o 2 puntos de evidencia si no hay pérdida de función. Si la hay, 3 puntos de evidencia: Sin pérdida de función o co-ocurrencia sin CMMRD AF= 0,01-1% No hay cosegregación con la enfermedad Al menos 3 tumores colorrectales con estabilidad de microsatélites o con IHQ inconsistente Probabilidad <4 en estudios caso-control con un intervalo de confianza mayor del 95%	AF mayor o igual al 1% en grupos control de referencia o PP= <0,001 o 3 puntos de evidencia si no hay pérdida de función. Si la hay, 4 puntos de evidencia: Sin pérdida de función o co-ocurrencia sin CMMRD AF= 0,01-1% No hay cosegregación con la enfermedad Al menos 3 tumores colorrectales con estabilidad de microsatélites o con IHQ inconsistente Probabilidad <4 en estudios caso-control con un intervalo de confianza mayor del 95%

Figura 22: Descripción general de las pautas de clasificación InSiGHT en 5 niveles y los tipos de evidencias requeridas para clasificar una variante en cada una de las clases. (Thompson et al, 2014).

En 2015, el *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 afecciones mendelianas, asumiendo también el sistema de caracterización en 5 niveles pero proponían 28 criterios basados en datos poblacionales, *in-silico,* funcionales y de cosegregación. Recientemente, las pautas de clasificación de variantes ACMG-AMP se han mejorado gracias al sistema Sherloc, que se basa en un sólido marco de 33 reglas e introduce 108 refinamientos, proporcionando un enfoque más consistente y transparente para la clasificación de variante (Figura 23).

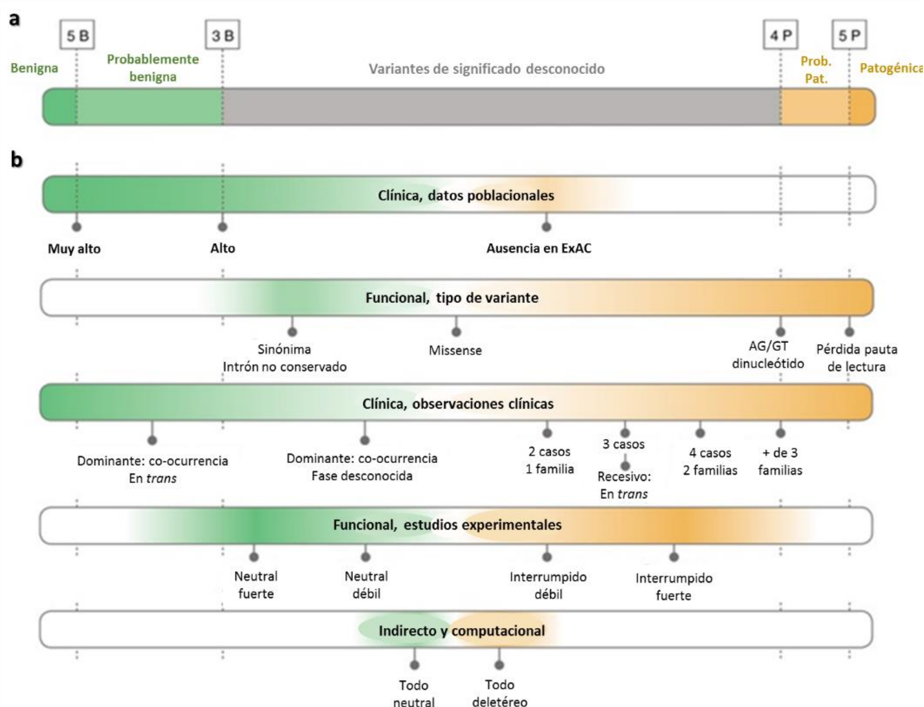


Figura 23: Umbrales de puntuación de clasificación Sherloc y categorías de evidencia. (a) Umbrales de puntuación para la clasificación de variantes patogénicas (P), probablemente patogénicas, variantes de significado desconocido, probablemente benigna y benigna (B). Las evidencias de las variantes patogénicas y benignas se califican por separado. (b) Las cinco categorías de evidencia en el orden de evaluación y con el valor en puntos de los criterios de selección indicados. Los criterios clínicos incluyen datos de población y hallazgos clínicos. Los criterios funcionales incluyen observaciones de secuencia, estudios moleculares e información indirecta y computacional.
(Nykamp et al. 2017)

Todos las guias de clasificación anteriormente mencionadas se basan en la interpretación de líneas de evidencia como las reportadas en bases de datos específicas de locus, evidencias moleculares y evidencias funcionales.

- **Evidencias reportadas en bases de datos específicas de locus:**

Las bases de datos específicas de locus ofrecen datos agregados, distintos enfoques metodológicos y un conocimiento profundo de los fenotipos clínicos. Su experiencia en

la clasificación de variantes proporciona una opinión consensuada sobre su patogenicidad antes de hacer pública su clasificación (Greenblatt *et al*, 2008). La transparencia con respecto a los criterios utilizados y la necesidad de que se argumente la información que soporta la clasificación sugerida, hace que los usuarios puedan considerar la información para su propia aplicación en la investigación y el entorno clínico.

- **Evidencias moleculares y funcionales:**

a) Información molecular sobre tumores: Para definir que un tumor tiene deficiencia MMR se utiliza el estudio de la inestabilidad de microsatélites y la pérdida de expresión de proteínas MMR en tejidos tumorales. También se valora la presencia de la mutación somática BRAF V600E y los estudios de MLH1 para descartar la naturaleza esporádica del tumor (Alexandrov & Stratton, 2014; Thompson *et al*, 2013). Eventualmente puede utilizarse la firma mutacional de un tumor o un grupo de tumores para la clasificación de variantes, puesto que estas reflejan su proceso neoplásico subyacente.

b) Datos funcionales: Los análisis funcionales de RNA y proteína son claves para la evaluación de la patogenicidad de una variante (Heinen & Rasmussen, 2012). Entre los estudios más extendidos se encuentran:

- Los estudios de procesamiento del mRNA y/o su estabilidad: Las variantes de significado desconocido pueden tener efectos sobre el procesamiento del RNA y/o la estabilidad del transcrito. Por lo general, las predicciones *in silico* se utilizan para identificar variantes que pueden alterar su procesamiento. Los resultados obtenidos pueden validarse con ensayos que incluyen el uso de linfocitos de portadores o minigenes (Gaildrat *et al*, 2010; Morak *et al*, 2011; Soukarieh *et al*, 2016; Tournier *et al*, 2008).
- Estabilidad de la proteína: Para el estudio de la estabilidad de una proteína se utiliza el Western blot. Para ello, la proteína portadora de la variante de significado desconocido se expresa ectópicamente. Sin embargo, la interpretación de los datos obtenidos mediante esta técnica no siempre es fácil. Así pues, la sobreexpresión de la proteína puede alterar la función normal o incluso enmascarar una pérdida parcial de actividad y su subexpresión puede dar como resultado resultados falsos negativos, puesto que la dosis de proteína también puede afectar la función (Cejka *et al*, 2003; Hinrichsen *et al*, 2013)
- Ubicación subcelular de la proteína: una localización celular adecuada es un requisito para el correcto funcionamiento de una proteína. Los ensayos celulares a menudo se usan para evaluar si una variante de significado incierto afecta la localización de su proteína (Borràs *et al*, 2012; Borràs *et al*, 2013).
- Funciones específicas de proteínas: para abordar el impacto de una variante de significado desconocido en la funcionalidad, se han desarrollado ensayos

específicos. Para las variantes de MMR, la mayoría de las herramientas analizan la actividad de MMR mediante el uso de ensayos de levadura in vivo, ensayos basados en células de mamíferos o ensayos MMR libres de células in vitro (los más utilizados) (Heinen & Rasmussen, 2012). Con respecto a las variantes de *MUTYH*, las herramientas incluyen ensayos para medir la actividad endonucleolítica (excisión de bases) de proteínas purificadas (D'Agostino *et al*, 2010). Las mutaciones que afectan a la revisión de la polimerasa generalmente se evalúan en sistemas de levadura, donde se analiza la capacidad de corregir mutaciones específicas (Murphy *et al*, 2006).

Tasa de detección de mutaciones

La tasa general de detección de mutaciones en pacientes preseleccionados oscila entre 30 y 78%, dependiendo de los criterios de inclusión aplicados (Lipton *et al*, 2004; Lynch *et al*, 2007; Mangold *et al*, 2005; Moreira *et al*, 2012; Syngal *et al*, 1999). Sin embargo, en series altamente seleccionadas de familias de Amsterdam con inestabilidad de microsatélites, el porcentaje de detección de mutaciones puede llegar al 95% (Mueller et al., 2009). Pese a que la tasa de detección del síndrome de Lynch se considera buena, en el 59% (95% CI 41-62%) de los casos con cáncer colorrectal a partir de cribado poblacional y en el 52% (95% CI 41-62%) de los casos de cáncer de endometrio con sospecha de síndrome de Lynch no se detecta mutación germinal en genes reparadores (Buchanan *et al*, 2014). Estos pacientes son los llamados individuos Lynch-like.

5. Individuos con sospecha de síndrome de Lynch sin mutación identificada o Lynch-like

Los individuos con tumores del espectro de síndrome de Lynch (con deficiencia reparadora y ausencia de metilación de *MLH1*) en los que no se ha identificado ninguna mutación germinal patogénica se denominan individuos con síndrome de Lynch-like (LLS, del inglés *Lynch-Like Syndrome*) (Rodríguez–Soler *et al*, 2013). Como se ha mencionado en el apartado anterior los individuso Lynch-like representan más del 50% de los individuos con tumores de colon y endometrio deficientes en reparación procedentes de cribado poblacional (Buchanan *et al*, 2014).

5.1. Potenciales causas del síndrome Lynch-like

Las posibles causas de predisposición de los individuos Lynch-like se pueden dividir entre causas germinales y somáticas.

Causas germinales

En algunos casos, los pacientes Lynch-like son portadores de variantes patogénicas no identificadas en alguno de los genes MMR. Generalmente, estas mutaciones suelen asociarse a la presencia de reordenamientos y variaciones en el número de copia que no siempre son fácilmente identificables mediante técnicas de análisis actuales. Algunos ejemplos son las inversiones de exones *MSH2* (Liu *et al*, 2016a; Mork *et al*, 2016; Rhees *et al*, 2014; Wagner *et al*, 2002) y la fusión genética del gen *MLH1* con otros genes circundantes como consecuencia de una inversión (Morak *et al*, 2011) o una deleción (Meyer *et al*, 2009) en el locus. En otros casos, las mutaciones pueden localizarse en regiones reguladoras de genes MMR, que raramente se criban en el análisis mutacional. En este sentido, se han reportados diferentes variantes de las regiones promotoras de genes MMR asociadas con el silenciamiento transcripcional del alelo o con la reducción de su actividad promotora (Liu *et al*, 2016b; Vargas-Parra *et al*, 2018).

Otro motivo de infradiagnóstico es la presencia de variantes de significado desconocido, que están presentes en aproximadamente el 30% de estos casos. Entre ellas, existen variantes que podrían ser patogénicas pero que por falta de evidencias clínicas, moleculares o funcionales no pueden ser clasificadas como tales. Es por este motivo que las familias portadoras son consideradas Lynch-like hasta que dichas variantes son reclasificadas.

El solapamiento de fenotipos con otros síndromes de cáncer colorrectal hereditario contribuye a las dificultades de diagnóstico de los pacientes Lynch-like. Se ha reportado que entre el 1 y el 3% de estos pacientes son realmente portadores bialélicos de mutaciones en *MUTYH*, gen asociado con poliposis adenomatosa familiar atenuada

(Castillejo *et al*, 2014; Morak *et al*, 2014). Del mismo modo, los tumores de pacientes con poliposis asociados a la corrección de la polimerasa también podrían conducir a la deficiencia de la MMR por la hipermutabilidad dada por las mutaciones en los genes *POLE* o *POLD1* (Bellido *et al*, 2015; Elsayed *et al*, 2014; Palles *et al*, 2012; Rohlin, 2014; Spier *et al*, 2015; Valle *et al*, 2014).

Por otro lado, el estudio de series de pacientes Lynch-like con estrategias de NGS ha permitido la identificación de variantes raras y potencialmente patogénicas en genes como *MCM9, FAN1, BUB1* y *SETD2*, que actualmente están emergiendo como genes candidatos responsables de síndrome Lynch-like (de Voer *et al*, 2013; Goldberg *et al*, 2015; Seguí *et al*, 2015; Smith *et al*, 2016; Vargas-Parra *et al*, 2017).

Causas somáticas

Finalmente, se han detectado dobles mutaciones somáticas de los genes de reparadores en una proporción variable de tumores de individuos Lynch-like (30-82%), lo que explicaría la deficiencia reparadora observada en estos (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). Sin embargo, y debido a su naturaleza somática, su implicación en una predisposición genética al cáncer es poco probable.

5.2. Riesgo de cáncer

Se considera que los pacientes Lynch-like y sus familiares de primer grado tienen un riesgo intermedio de desarrollar cáncer colorrectal (Chika *et al*, 2017; Hampel *et al*, 2005; Mas-Moya *et al*, 2015; Overbeek *et al*, 2007; Rodríguez–Soler *et al*, 2013; Vargas-Parra *et al*, 2017; Win *et al*, 2015). La edad media de diagnóstico es similar a la observada en individuos con síndrome de Lynch con mutación identificada o bien es intermedia entre individuos con síndrome de Lynch y pacientes con cáncer colorrectal esporádico (Overbeek *et al*, 2007; Rodríguez–Soler *et al*, 2013; Win *et al*, 2015). Al igual que ocurre con los individuos con síndrome de Lynch de mutación identificada, los tumores colorrectales de individuos Lynch-like se localizan predominantemente en el colon proximal, suelen tener histología mucinosa y pueden presentar tumores sincrónicos o metacrónicos del espectro del síndrome de Lynch.

La incidencia del cáncer colorrectal es menor en familias Lynch-like que en las familias con síndrome de Lynch, sin embargo es mayor a la incidencia de familias con cáncer colorrectal esporádico (Rodríguez–Soler *et al*, 2013; Win *et al*, 2015). La incidencia de tumores extracolónicos es de aproximadamente 11% en pacientes con Lynch-like, aunque varía ampliamente entre series (Chika *et al*, 2017; Hampel *et al*, 2005; Mas-Moya *et al*, 2015; Rodríguez–Soler *et al*, 2013; Vargas-Parra *et al*, 2017; Win *et al*, 2015). Hasta la fecha, no hay datos publicados sobre el riesgo asociado a familiares de primer

grado para los tumores del espectro del síndrome de Lynch. Estos resultados confirman la necesidad de estrategias especiales de detección y vigilancia para estos pacientes y sus familiares.

5.3. Recomendaciones de seguimiento

El grupo Lynch-like podría ser considerado como un grupo heterogéneo de diferentes bases moleculares, por lo que no se puede generalizar un cribado óptimo hasta que se realice un diagnóstico específico. La incapacidad de definir pautas de detección y de manejo para casos Lynch-like impide establecer unas recomendaciones de seguimiento claras para este tipo de pacientes. Debido a que el riesgo de cáncer es muy variable dentro de estas familias, algunas familias Lynch-like podrían estar siendo sometidas a un exceso de pruebas analíticas (en ocasiones innecesarias) que, además, pueden ser motivo de ansiedad en estos individuos, mientras que otras carecerían de un seguimiento adecuado a su riesgo (Geurts-Giele et al 2014). Este problema, además de afectar a estas familias, es un problema que también afecta al sistema de atención médica.

Actualmente y debido al riesgo intermedio de cáncer colorrectal observado en familias Lynch-like, se recomienda un seguimiento intermedio (Rodriguez-Soler et al 2013), aunque que aconseja tener siempre en consideración los antecedentes familiares de estos individuos para personalizar, en la medida de lo posible, su seguimiento.

6. Epimutaciones en síndrome de Lynch

6.1. El concepto de epimutación

La epigenética es el conjunto de elementos funcionales que regulan la expresión génica de una célula sin alterar la secuencia de DNA. Uno de los mecanismos epigenéticos más importantes es la metilación del DNA, que es la adición de un grupo metilo ($-CH_3$) en las citosinas (C) que están situadas contiguas a una guanina (G). Esta combinación de dinucleótidos recibe el nombre de sitio CpG.

Denominamos islas CpG a las secuencias de DNA con una alta densidad de sitios CpG (>50% de su secuencia). Cuando estas islas CpG se sitúan en el promotor de un gen, su metilación regula el silenciamiento génico permitiendo la plasticidad fenotípica del gen.

Llamamos epimutación a la metilación aberrante de un gen en tejidos somáticos normales de un individuo. Las epimutaciones pueden provocar por un lado el silenciamiento transcripcional del gen afecto de forma constitucional por una hipermetilación de la región o, por otro lado, una activación de transcripción también de forma constitucional por hipometilación. El estado epigenético de los promotores de los genes MMR es no metilado, por lo que las epimutaciones en síndrome de Lynch están relacionadas con la hipermetilación constitucional de sus secuencias y su silenciamiento transcripcional.

Las epimutaciones aparecen de forma hemialélica, es decir, solo afectan específicamente a una de las copias del gen. En el síndrome de Lynch afecta a uno de los alelos y es por esto por lo que representan un mecanismo etiológico alternativo a la mutación genética.

6.2. Epidemiologia

En 2002, Gazzoli y colaboradores reportaron el primer de epimutación en síndrome de Lynch, siendo el gen afecto *MLH1* (Gazzoli *et al*, 2002). En la actualidad, solo 87 casos de epimutación en *MLH1* han sido reportados. Aunque no se ha determinado la incidencia de las epimutaciones y parece un defecto poco frecuente, la frecuencia de identificación de estos casos depende en gran medida del cribado utilizado en los grupos de estudio. La tasa de detección se sitúa entre el 0 y el 1.6% en grupos donde solo se ha cribado a los pacientes por no presentar mutaciones en los genes MMR. La tasa de

detección asciende al 3-9% en grupos donde la selección se ha basado en pacientes con pérdida de MLH1 en tumor e inestabilidad de microsatélites y alcanza el 14% de casos cuando además se incluye como factor de selección la hipermetilación de *MLH1* en tumor (Hitchins, 2013)

En 2009 se reportó el primer caso de epimutación de *MSH2* (Ligtenberg *et al*, 2009). En este caso, la epimutación estaba causada por una deleción en el extremo 3' de *EPCAM*, el gen previo a *MSH2*. Desde entonces se han reportado alrededor de 50 familias diferentes con deleciones en *EPCAM* (Dymerska *et al*, 2017). La frecuencia de epimutaciones en *MSH2* en individuos con pérdida de MSH2 en tumor y sin mutaciones en línea germinal en este gen varía entre el 0 y el 9%, aunque esta frecuencia puede aumentar hasta el 40% en poblaciones con mutaciones fundadoras en *EPCAM* (Ligtenberg *et al*, 2013; Peltomäki, 2014).

Hasta el momento no se han documentado epimutaciones en *MSH6* y *PMS2*.

6.3. Clasificación

Dependiendo de la causa subyacente, las epimutaciones pueden presentar distintos patrones de herencia. En base a este hecho las epimutaciones se han clasificado en 2 grupos: las epimutaciones secundarias (asociada a una variante genética en *cis*) y las primarias (independientes del contexto genético). Como puede observarse en la figura 24, ambos tipos de epimutaciones han sido descritos para *MLH1*, mientras que en *MSH2* solo hay constancia de epimutaciones secundarias.

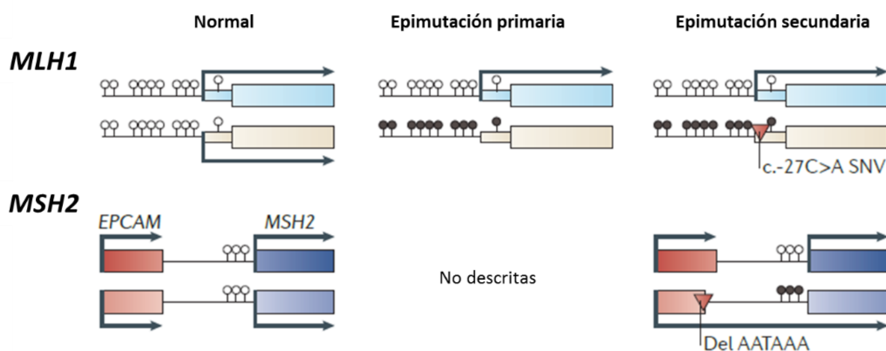


Figura 24: Esquema de los diferentes tipos de epimutaciones en síndrome de Lynch. A la izquierda, se representa la situación normal. En el centro, las epimutaciones primarias, que son

independientes de cualquier alteración genética. A la derecha, las epimutaciones secundarias, que están asociadas a mutaciones genéticas (c.-27C>A o Del AATAAA en el ejemplo). Los círculos en blanco representan sitios CpG del promotor que no están metilados, mientras que los negros representan sitios metilados. La presencia de actividad transcripcional se indica mediante una flecha.
(Hitchins 2015)

Epimutaciones secundarias

Las epimutaciones secundarias o "genéticamente facilitadas" son aquellas que están causadas por cambios genéticos en *cis* que alteran el estado epigenético del gen. Puesto que están ligadas a una alteración genética y la metilación es secundaria a ésta, siguen un patrón de herencia Mendeliano autosómico dominante (Figura 25) (Hesson *et al*, 2010).

Epimutaciones secundarias en MLH1

En *MLH1* se han identificado diferentes alteraciones genéticas en *cis* asociadas a su epimutación. Concretamente se han identificado una duplicación completa de *MLH1* (Morak *et al*, 2011; Morak *et al*, 2008), diferentes deleciones del extremo 5' del gen (Cini *et al*, 2015; Gylling *et al*, 2009; Renkonen *et al*, 2003) y la variante compleja c.-63_-58delins18 en su región promotora (Morak *et al*, 2018)

Además de estas alteraciones, se ha visto que la epimutación de *MLH1* también cosegrega de forma autosómica dominante en ciertos haplotipos. En cinco familias diferentes de ascendencia europea, la metilación se asocia al haplotipo c.[-27C>A; c.85G>T] (Hitchins *et al*, 2011; Kwok *et al*, 2014). En una familia independiente, la epimutación cosegrega con un determinado haplotipo sin variantes genéticas identificadas en el locus, aunque en este caso, la epimutación podría considerarse "posiblemente secundaria" por la falta de una variante genética asociada (Crépin *et al*, 2012).

Herencia autosómica dominante:

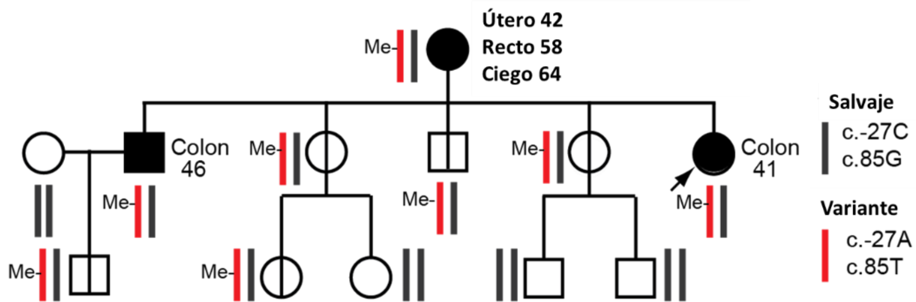


Figura 25: Familia con epimutación secundaria en *MLH1*. El alelo asociado a metilación (rojo) cosegrega de forma autosómica dominante. Los haplotipos no informativos se esquematizan con una línea vertical gris. Cuadrado=hombre, círculo=mujer, flecha=probando; símbolo negro=individuo afecto de cáncer; símbolo blanco=individuo no afecto, símbolo con línea vertical=individuo portador no afecto.
(Hitchins 2013)

De momento se desconoce el mecanismo causal por el que las variantes reportadas provocan la epimutación del gen, aunque todas ellas afectan o están localizadas dentro de la isla CpG del promotor de *MLH1*.

Epimutaciones secundarias en MSH2

En el caso de *MSH2*, su epimutación está causada por la deleción en el extremo 3' de *EPCAM*, el gen previo a *MSH2.* Se han reportado más de más de 20 grandes deleciones de diferentes tamaños (Dymerska *et al*, 2017). Los análisis de los puntos de corte de estas deleciones indican que su origen es debido mayoritariamente a eventos de recombinación entre secuencias Alu, las cuales tienen una gran densidad en este locus (Ligtenberg *et al*, 2013).

La eliminación de los exones finales de *EPCAM*, que comprende la señal de terminación de transcripción, provoca la expresión de un transcrito de fusión *EPCAM-MSH2* que probablemente induce la metilación del promotor de *MSH2* del alelo situado en *cis* con la deleción. Este fenómeno solo se observa en tejidos que expresan *EPCAM*, principalmente epiteliales, por lo que los portadores de deleciones en *EPCAM* presentan mosaicismo en lo que respecta a la epimutación de *MSH2*. Es por este motivo que la epimutación no es detectable cuando se analiza la metilación del promotor de *MSH2* en DNA sanguíneo (Ligtenberg *et al*, 2009; Peltomäki, 2014).

En el caso de los pacientes con epimutaciones secundarias de *MSH2* se ha observado que la extensión de la deleción del extremo 3' de *EPCAM* condiciona su riesgo (Figura 26). Todas las deleciones confieren un elevado riesgo de cáncer colorrectal a sus portadores, mientras que solo las deleciones que se extienden hasta el promotor de *MSH2* tienen incrementado el riesgo de cáncer de endometrio (Kempers *et al*, 2011).

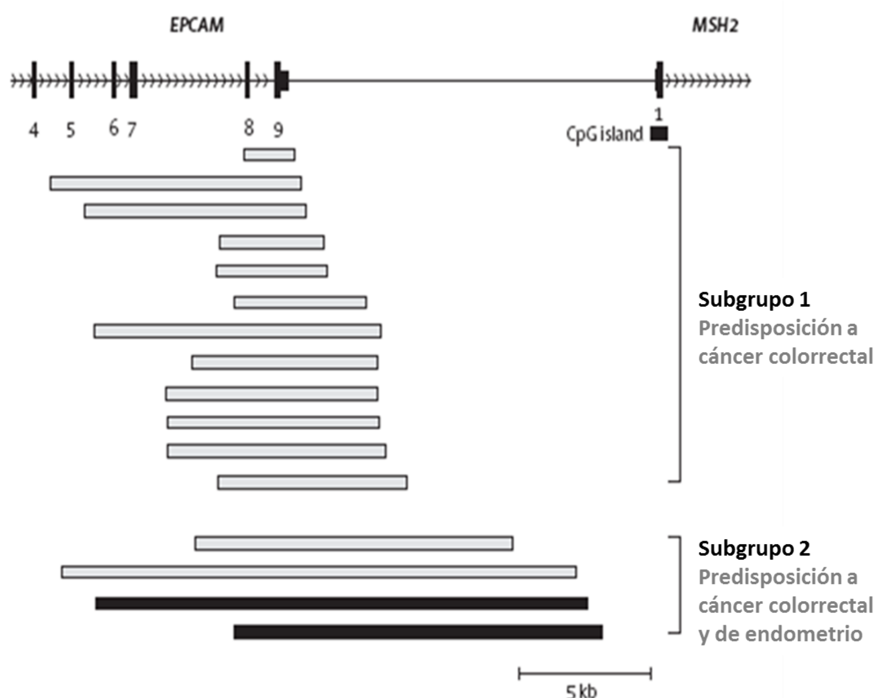


Figura 26: Clasificación de las deleciones en *EPCAM* según el riesgo de cáncer asociado en portadores. Representación esquemática del tamaño de cada deleción (barras) y su posición relativa con respecto a la isla CpG del promotor de MSH2 (cuadrado negro). Las barras negras muestran deleciones observadas pacientes con cáncer de endometrio.
(Kempers et al 2011)

Epimutaciones primarias en *MLH1*

Las epimutaciones primarias en *MLH1* no están asociadas a una causa genética conocida. Aparecen *de novo* el individuo y pueden ser reversibles entre generaciones. Los individuos portadores de epimutaciones constitucionales en *MLH1* desarrollan una fuerte historia personal de cáncer con tumores del espectro del síndrome de Lynch, aunque es muy infrecuente que tengan historia familiar de cáncer.

Un rasgo frecuente de las epimutaciones primarias es el borrado intergeneracional (Dámaso *et al*, 2018; Hitchins *et al*, 2007; Morak *et al*, 2008; Pineda *et al*, 2012; Suter *et al*, 2004a). Determinamos que ha habido borrado intergeneracional cuando un

individuo portador de epimutación transmite a su descendencia el alelo asociado a metilación pero, pese a ser el mismo alelo genético, en este el promotor de *MLH1* se encuentra en estado no metilado. Sin embargo, también se han documentado casos con patrones de herencia no mendelianos (Figura 27) (Hitchins *et al*, 2007; Sloane *et al*, 2015; Ward *et al*, 2013a).

En la mayor parte de los casos, la epimutación es constante en todas las células somáticas normales, lo que supone una metilación de aproximadamente el 50% y un silenciamiento completo del alelo (Dámaso *et al*, 2018). Sin embargo, en otros casos se ha observado diversos grados de mosaicismo epigenético.

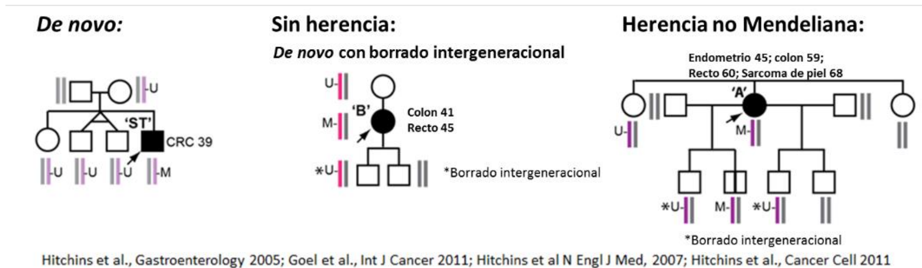


Figura 27: Familias con epimutación primaria de *MLH1*. El alelo asociado a metilación se esquematiza con una línea vertical en color, mientras que los haplotipos no informativos están esquematizados con líneas verticales grises. El estado epigenético del alelo asociado a metilación se indica con una "U" para el estado no metilado y con una "M" para el metilado (epimutación). El asterisco señala los casos con borrado intergeneracional.
Cuadrado=hombre, círculo=mujer, flecha=probando; símbolo negro=individuo afecto de cáncer; símbolo blanco=individuo no afecto, símbolo con línea vertical=individuo portador no afecto.
(Hitchins 2013)

En el presente trabajo, nos centramos en la identificación y caracterización de los pacientes con epimutación primaria de *MLH1*.

6.4. Asesoramiento genético

El asesoramiento genético de estos pacientes y sus familias depende en gran medida del tipo de epimutación encontrado.

Las familias con epimutaciones secundarias (tanto de *MLH1* como de *MSH2*) se consideran familias con síndrome de Lynch. Una vez identificada la causa genética de la epimutación, se procede al cribado de los individuos portadores dentro de la familia. A los individuos portadores se les recomienda un seguimiento estricto y se considera que los individuos no portadores tienen un riesgo poblacional, por lo que no precisan de un seguimiento específico.

Los individuos con epimutación primaria de *MLH1* fenocopian a los individuos con síndrome de Lynch en cuanto a su espectro de tumores. Sin embargo, se ha advertido que la edad de diagnóstico de los pacientes con epimutación es 5 años más temprana que en los casos con mutaciones en la secuencia de *MLH1*, siendo 40 años la edad media de diagnóstico de los pacientes con epimutación y 45 la de pacientes con mutación en *MLH1* (Hitchins, 2013). Los pacientes con epimutación primaria de *MLH1,* por tanto, también requieren de un seguimiento estricto.

Sin embargo, la estimación del riesgo y el asesoramiento genético de los familiares de primer grado de un paciente con epimutación primaria supone un desafío en las Unidades de Consejo Genético. Cuando se diagnostica un caso de epimutación primaria, es importante ofrecer el estudio epigenético a sus descendientes si los hubiera. Si la epimutación no ha sido transmitida, los descendientes pueden ser asesorados como familiares de un caso síndrome de Lynch donde las pruebas genéticas directas no han sido informativas. En este contexto, se supone que la falta de metilación en el alelo heredado no descarta que exista un estado de mosaico en el paciente o que una alteración genética no detectada que predispone a una adquisición tardía de metilación esté presente en esta familia (Pineda *et al*, 2012).

HIPÓTESIS

Hipótesis 1

La deficiencia reparadora es el marcador molecular de los tumores de pacientes con síndrome de Lynch con mutaciones o epimutaciones germinales en los genes MMR. En individuos sin mutación germinal identificada en genes reparadores, la presencia de metilación somática en *MLH1* y mutaciones somáticas bialélicas en genes reparadores se han descrito como responsables de la deficiencia reparadora en un porcentaje de estos casos. Sin embargo, la falta de detección de mutaciones germinales patogénicas en genes reparadores no excluye que estos individuos tengan una predisposición hereditaria al cáncer.

Hipotetizamos que en individuos Lynch-like la presencia de epimutaciones germinales podría ser responsable de la predisposición a desarrollar tumores con deficiencia reparadora.

Hipótesis 2

Los pacientes portadores de epimutaciones constitucionales en *MLH1* presentan fenotipo clásico de síndrome de Lynch. Mientras que las epimutaciones secundarias se asocian a cambios genéticos en *cis* y patrón de herencia autosómico dominante, las epimutaciones primarias aparecen habitualmente *de novo*, sin cambios genéticos aparentemente asociados y presentan herencia nula o no mendeliana.

Hipotetizamos que una mejor identificación y la caracterización exhaustiva de las epimutaciones constitucionales en *MLH1* permitirá mejorar el manejo clínico de estos pacientes y sus familiares, así como contribuir a elucidar su mecanismo causal.

OBJETIVOS

Objetivo principal 1

Buscar la presencia de alteraciones epigenéticas constitucionales responsables de síndrome de Lynch en individuos con sospecha de este síndrome en los cuales no se han detectado mutaciones germinales en genes reparadores (o individuos Lynch-like).

Subobjetivos:

1. Refinamiento de pacientes candidatos a ser estudiados mediante una caracterización molecular exhaustiva en línea germinal
2. Caracterización de las alteracions epigenéticas presentes en leucocitos de sangre periférica y mucosa colónica en individuos Lynch-like

Para la realización de este objetivo hemos analizado una serie de 121 individuos Lynch-like procedentes de diferentes hospitales de Cataluña.

Objetivo principal 2

Identificar y caracterizar molecularmente epimutaciones constitucionales en MLH1

Subobjetivos:

1. Búsqueda de alteraciones genéticas responsables de las epimutaciones constitucionales que expliquen el fenotipo de los pacientes afectos
2. Caracterización de las alteraciones epigenéticas presentes en pacientes portadores de epimutaciones constitucionales en *MLH1*
3. Cribado de epimutaciones constitucionales mediante el uso de técnicas de alta sensibilidad

Para la realización de este objetivo hemos analizado una serie de 22 individuos con tumores metilados en *MLH1*, así como 12 portadores de epimutaciones en *MLH1* previamente identificadas en nuestro centro o centros colaboradores.

RESULTADOS

Los resultados del trabajo experimental de esta tesis están incluidos en cuatro artículos científicos: uno de ellos ya publicado, otro aceptado por la revista y dos preparados para enviar a publicar. Siguiendo los objetivos de este trabajo, los artículos se presentan en el siguiente orden:

- **Artículos derivados del objetivo 1:**

  ARTÍCULO 1

  Elucidating the molecular basis of MSH2-deficient tumors in suspected Lynch Syndrome cases. International Journal of Cancer, 141: 1365-1380

  ARTÍCULO 2

  Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals. Manuscrito en preparación.

- **Artículos derivados del objetivo 2:**

  ARTÍCULO 3

  Primary constitutional *MLH1* epimutations: a focal epigenetic event. British Journal of Cancer, *in press.*

  ARTÍCULO 4

  Highly sensitive *MLH1* methylation analysis in blood allows the identification of low-level epigenetic mosaicism. Manuscrito en preparación.

Antes de cada artículo, se incluye un breve resumen de los resultados obtenidos de cada uno de los trabajos y la contribución del doctorando.

ARTÍCULO 1

Elucidating the molecular basis of MSH2-deficient tumors in suspected Lynch Syndrome cases.

Hipótesis: El análisis exhaustivo, tanto a nivel germinal como somático, de las alteraciones genéticas en los genes de reparación de desapareamientos del DNA (MMR) y otros genes de cáncer colorrectal (CRC) puede ser útil para dilucidar las bases moleculares de los casos con sospecha de síndrome de Lynch o pacientes Lynch-like (LLS)

Objetivo: Estudiar la eficacia del algoritmo diagnóstico para el 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.

Resumen de los resultados obtenidos: 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 proveniente de sangre periférica de los pacientes LLS identificó una variante patogénica previamente no identificada que provoca un transcrito aberrante y variantes predichas 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 genes reparadores en tumores de 2 de los 5 pacientes estudiados. En los casos restantes, se hallaron mutaciones heterocigotas complejas de genes MMR (*MSH6, PMS2, MLH3*) y / o polimerasas de con corrección de pruebas (*POLD1* o *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 pacientes Lynch-like, 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 predisponentes a CRC. Esta estrategia podría ayudar a dilucidar las bases moleculares de LLS.

Contribución del doctorando: Cultivo de linfocitos, extracción de RNA y retrotranscripción. Gestión de los bloques de parafina, corte y tinción de laminillas, macrodisección y extracción del DNA del tejido parafinado. Diseño y validación de los resultados obtenidos mediante el panel de NGS. Revisión del manuscrito.

Un borrador previo de este artículo fue incluido en la tesis doctoral de la Dra. Gardenia María Vargas Parra (Programa de doctorado en Genética, Universidad de Barcelona, 2015)

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

 © 2017 UICC

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 ≤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× in PBL samples and $\geq$10× 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): MSH2/EPCAM	MSH6	Identified variants, cDNA change^	Identified variant (predicted protein change)	Initial MMR variant classification (September 2015)	VUS assessment: Splicing and multifactorial analysis	MMR variant reclassification (this study)	RNA analysis: whole MSH2 transcript	Promoter analysis (Sanger): MSH2	MSH6	3'UTR (Sanger): EPCAM	MMR mutational Variants identified in: MSH2/MSH6 ^	CRC associated genes	Tumor analysis: Methylation MSH2	Methylation MSH6	MSH2 and MSH6 hits / LOH	Final classification^
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: 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.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.

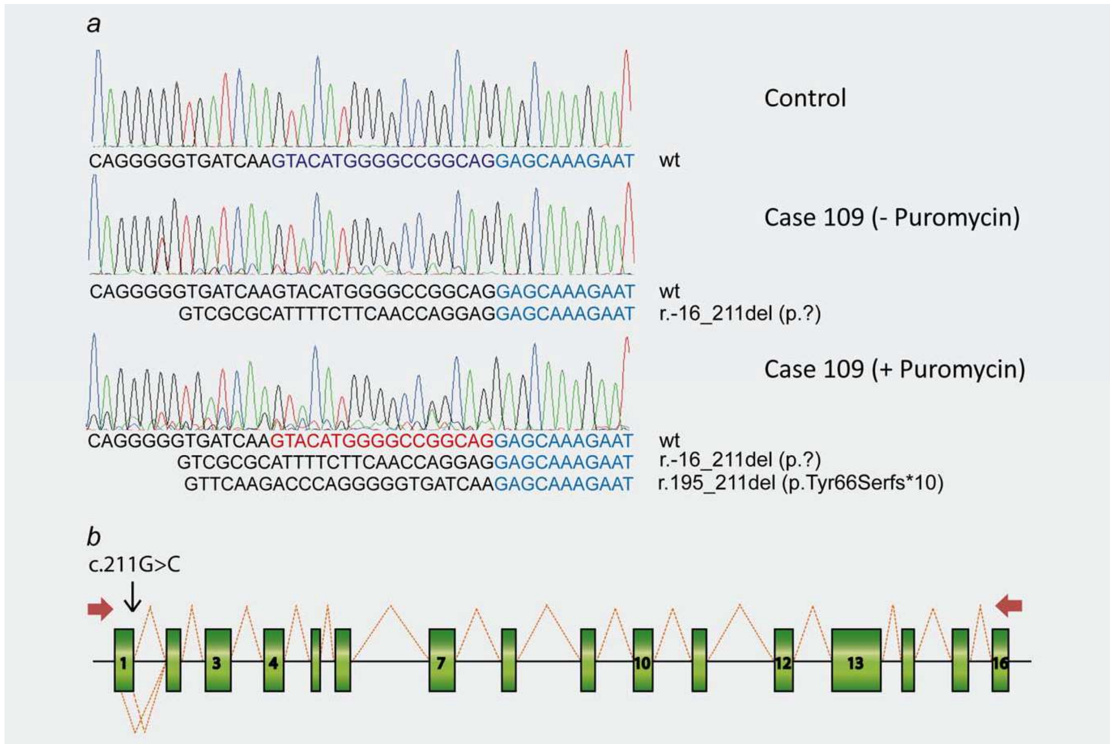


Figure 1. Results obtained in the analysis of the whole *MSH2* transcript of Case 109. (*a*) Result of the direct sequencing analysis of the RT-PCR product at the exon–intron boundary of exon 1 from a control sample and Case 109, harboring the germline *MSH2* c.211G > C variant, in presence and absence 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

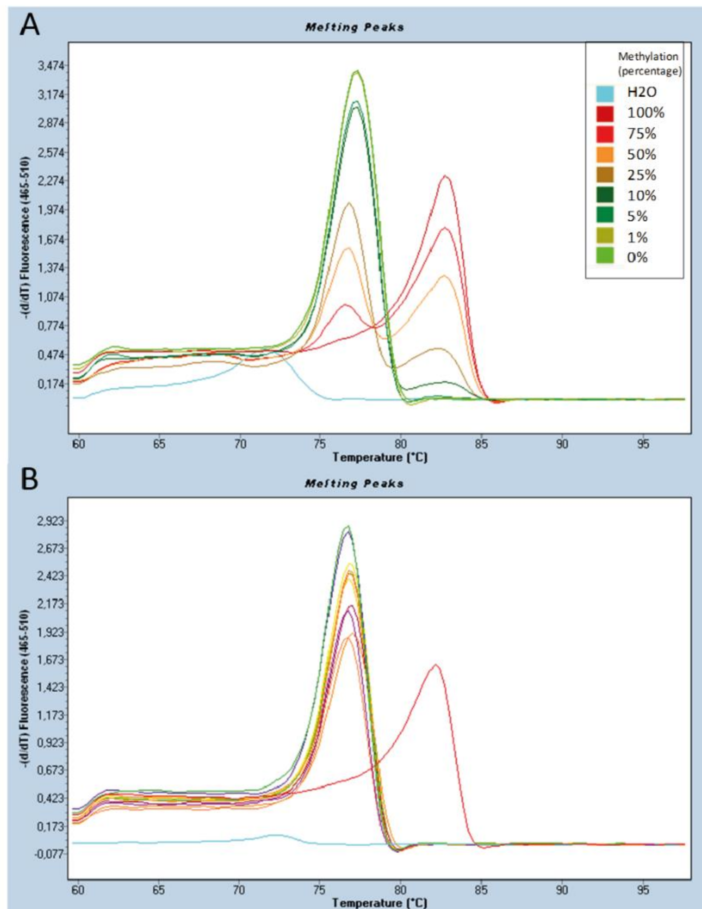


Figure S1. Methylation Specific-Melting Curve Analysis (MS-MCA) of *MSH2* promoter. **A.** Analytical sensitivity of MS-MCA for the detection of *MSH2* promoter methylation. Serial dilutions of a methylated control (CpG Methylated Jurkat Genomic DNA, New England Biolabs) were made with an unmethylated 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.

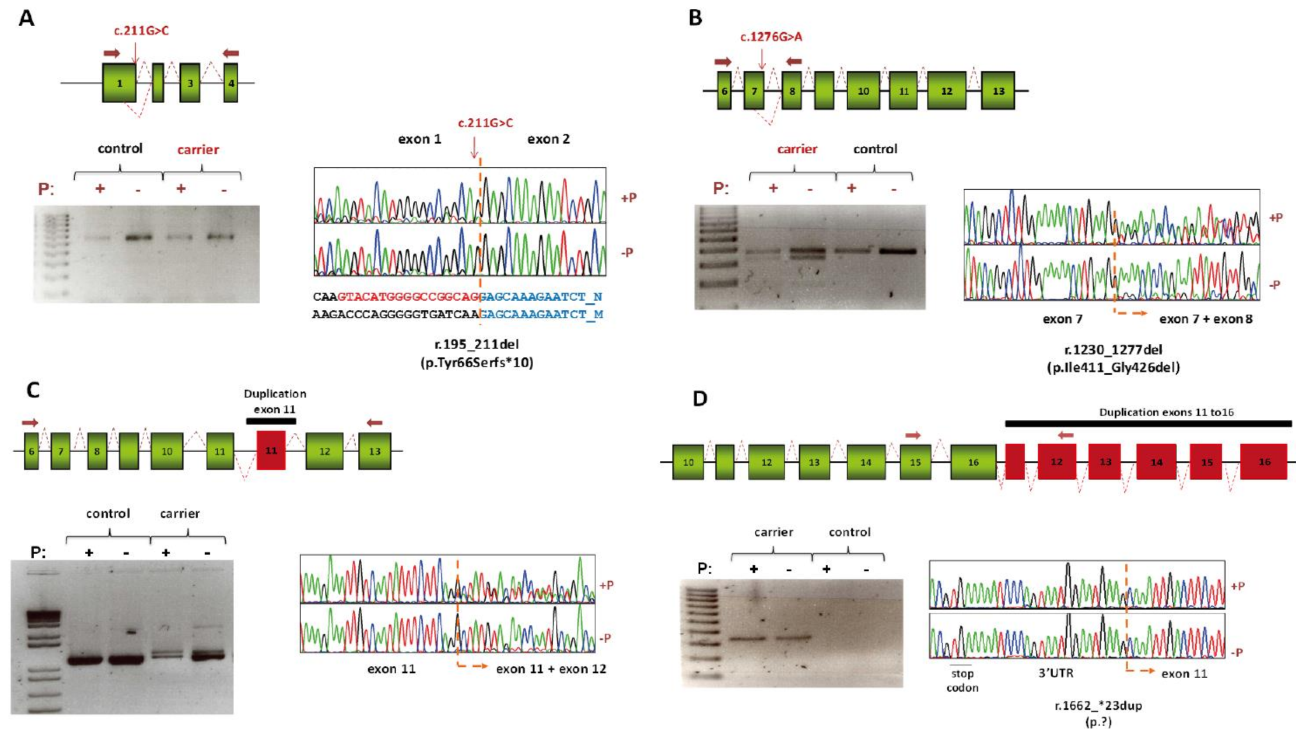


Figure S2. cDNA characterization of the *MSH2* c.211G>C (A), c.1276G>A (B), duplication of exon 11 (C), duplication of exons 11 – 16 (D). On the top, a schematic representation of normal transcripts (upper dotted lines) and aberrant transcripts (lower dotted lines) caused by the variants is shown. Green 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.

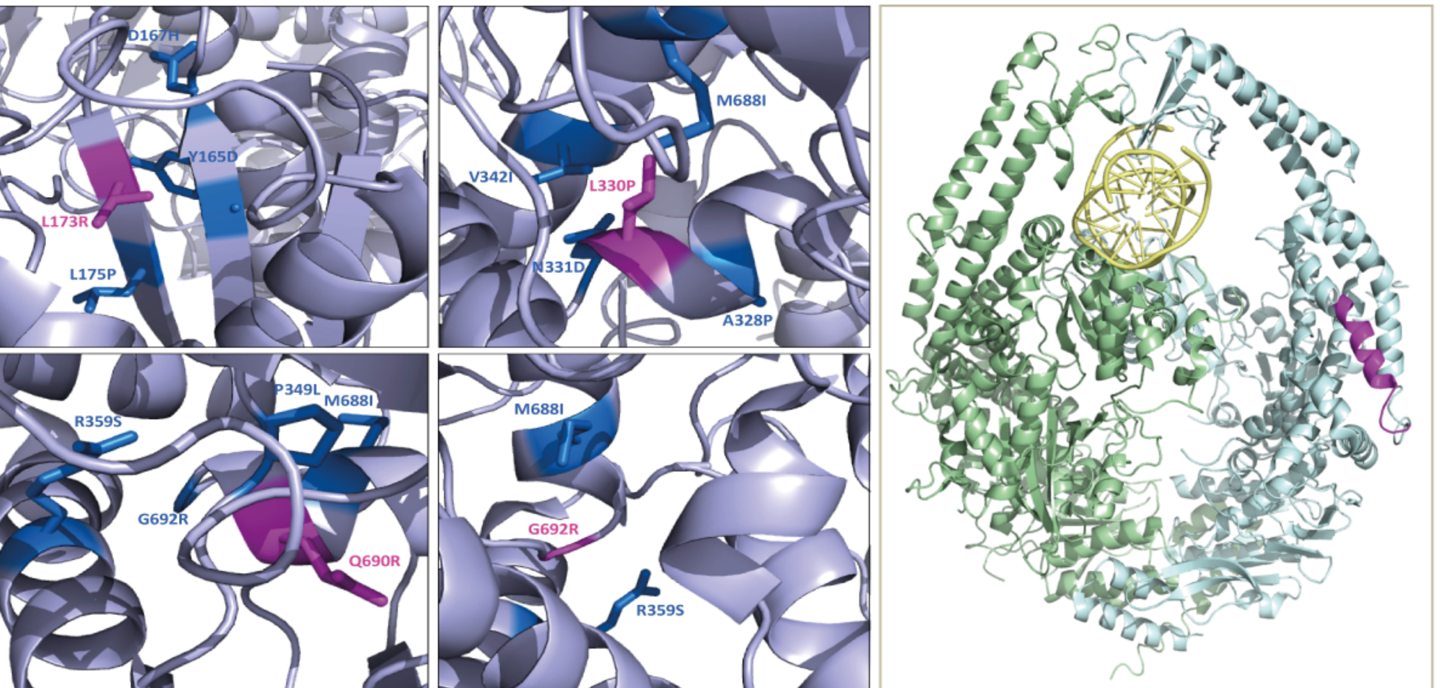
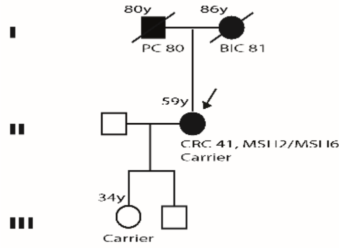


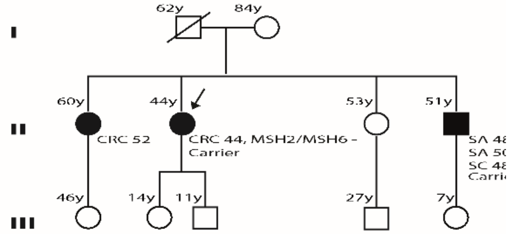
Figure S3. Protein structure location of MSH2 variants. The local environment of the four substitutions described in this work: L173R (top-left), L330P (top-right), Q690R (bottom-left) and G692R (bottom-right) is shown. In each case, the variant residue (in magenta) and their close neighbors which accommodate 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.

A

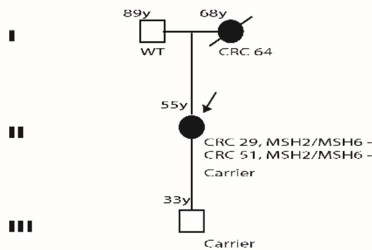
Family 122: c.518T>G



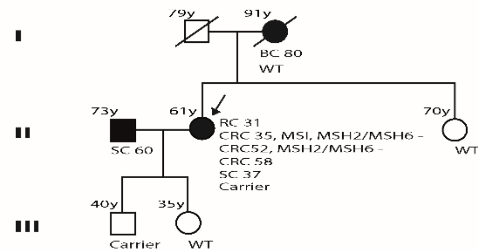
Family 117: c.518T>G



Family 264: Duplication of E11



Family 118: c.2069A>G



Family 120: Duplication of E11-16

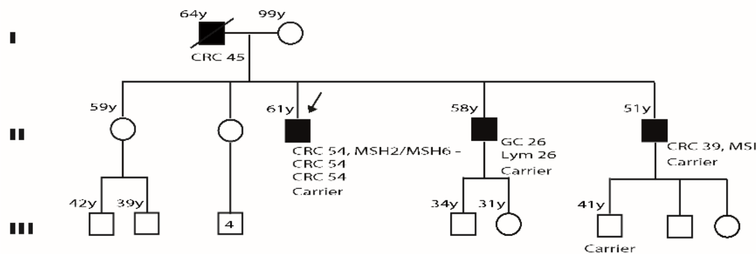
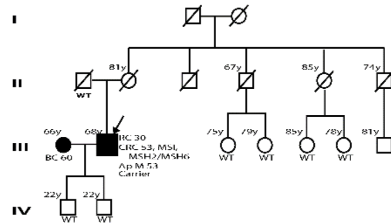


Figure S4. Family pedigrees from carriers of *MSH2* class 3 variants (A), class 4 variants (B) and further identified c.518T>G variant (C). Filled symbol, cancer; arrow, index case. Cosegregation results are indicated below individual's symbols as "carrier" or "WT". Current ages and ages at death, when 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

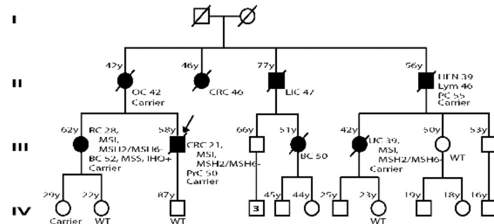
stable; IHC+, conserved MMR protein expression; the pattern of expression of MSH2 and MSH6 proteins is indicated (-, loss; NV, non-valuable).

B

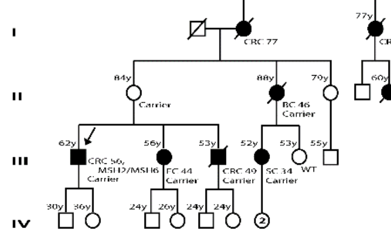
Family 240: c.[2635-3C>T;2635-5T>C]



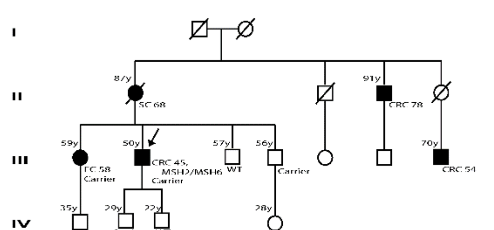
Family 235: c.[2635-3C>T;2635-5T>C]



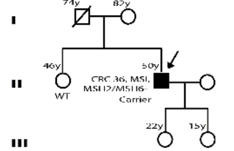
Family 232: c.[2635-3C>T;2635-5T>C]



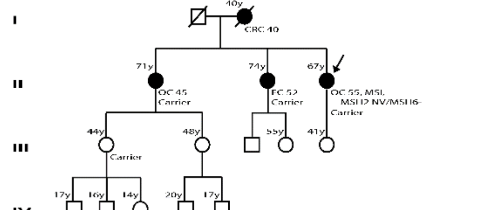
Family 234: c.211G>C



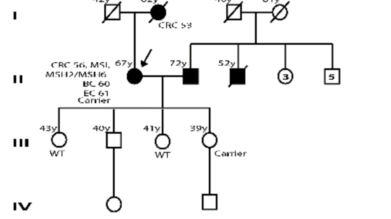
Family 239: c.2074G>C



Family 228: c.989T>C



Family 248: c.1511-1G>A



Family 258: c.1276G>A

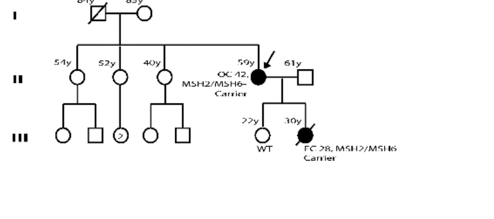
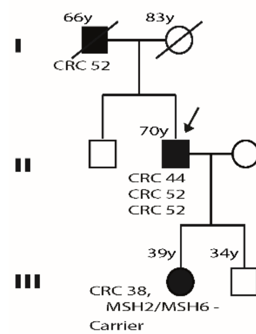


Figure S4. Family pedigrees from carriers of *MSH2* class 3 variants (A), class 4 variants (B) and further identified c.518T>G variant (C). **(Cont.)**

C

Family A1: c.518T>G



Family A2: c.518T>G

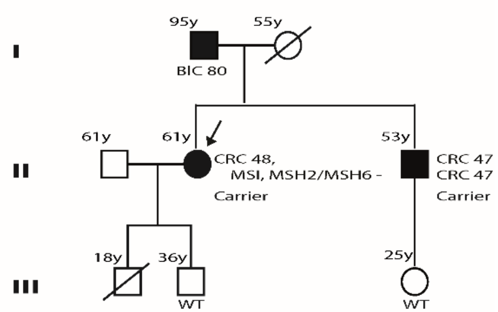


Figure S4. Family pedigrees from carriers of *MSH2* class 3 variants (A), class 4 variants (B) and further identified c.518T>G variant (C). **(Cont.)**

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)	BlC (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	**BlC*** (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)	BlC (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)	BlC (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** n (%)	**LS** n (%)	**LLS** 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	_	_	_	
108_C1				0.12 (20/167x)	NA		NA		0.07 (2/27x)	0,89	0,62	NI-MSI	**Possible**
108_C2				0.18 (23/130x)	NA		NA		0.06 (4/72x)	0,77	0,83	0,99	**Possible**
114	0.504		0.486							_	_	_	
114_C1	0.609		NA							0,93	NI - homo	NA	**Not apparent**
121		0.47				0.448		0.466		_	_	_	
121_C1		0.48				0.485		0.484		0,6	NI - homo	NI- MSI	**Not apparent**
111					0.46		0.53		0.46	_	_	_	
111_C3					0.53		NA		0.49	0,82	NI - homo	NI - homo	**Not apparent**

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 2

Comprehensive constitutional genetic and epigenetic characterization of Lynch-like individuals

Hipótesis: La predisposición a desarrollar cáncer colorrectal (CRC) con deficiencia reparadora observada en los pacientes Lynch-like (LLS) podría ser causada por epimutaciones constitucionales no identificadas.

Objetivo: Dilucidar la causa subyacente a la deficiencia reparadora observada en individuos LLS mediante un análisis exhaustivo de los casos a nivel genético y epigenético.

Resumen de los resultados obtenidos: Se incluyeron 115 pacientes que cumplen criterios de LLS, 23 de los cuales habían sido previamente reportados por nuestro grupo (Vargas-Parra et al 2017 – Artículo 1). El reanálisis mediante un panel personalizado 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 genes MMR, de las cuales 5 pudieron ser reclasificadas como patogénicas. También se encontraron 13 variantes presuntamente patogénicas por su estudio *in-silico* en otros genes de predisposición a CRC. El análisis del metiloma identificó un nuevo caso de epimutación constitucional de *MLH1*. Sin embargo, no se identificaron regiones diferencialmente metiladas en pacientes LLS cuando son comparados con individuos Lynch o controles sanos.

El estudio de subexomas combinado con la evaluación de patogenicidad de variantes de significado desconocido permite la identificación de mutaciones deletéreas en genes MMR, así como nuevos genes candidatos LLS. Las epimutaciones constitucionales fuera de los genes MMR no son responsables del fenotipo de deficiencia de MMR observado en pacientes con LLS.

Contribución del doctorando: Coordinación y recogida de muestras de centros externos. Gestión de los bloques de parafina, corte y tinción de laminillas, macrodisección y extracción del DNA del tejido parafinado. Enriquecimiento y preparación de librerías del panel NGS. Identificación de variantes, filtrado y anotación de resultados NGS. Predicción in-silico de las variantes encontradas por el panel NGS. Diseño y validación de los resultados obtenidos mediante el panel de NGS. Estudio de MLPA e interpretación de resultados .Elaboración de las matrices de comparación del estudio del metiloma. Análisis, interpretación y discusión de resultados. Preparación de tablas y figuras. Escritura del manuscrito.

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], Fátima Marín[1,2], Anna Fernández[1], Carolina Gómez[1], Àngela Velasco[6], Ares Solanes[1], Sílvia Iglesias[1], Gisela Urgell[5], Consol López[4], Jesús del Valle[1], Olga Campos[1], Maria Santacana[6], Xavier Matias-Guiu[6], Conxi Lázaro[1], Laura Valle[1,2], Joan Brunet[1,2,6], Marta Pineda[1]*, Gabriel Capellá[1]*

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 Creus I Sant Pau, Barcelona, Spain.
(5) Genetic Counseling Unit, Hospital Arnau de Vilanova, Lleida, Sapin.
(6) Hereditary Cancer Program, Catalan Institute of Oncology, Institut d'Investigació Biomèdica de Girona (IDIBGI), Girona, Spain.

Short title: Constitutional (epi)genetic characterization of Lynch-like syndrome

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 MMR-deficiency in LLS patients throughout a comprehensive (epi)genetic analysis.

One hundred and fifteen LLS patients harboring MMR deficient tumors and no pathogenic germline mutations identified in MMR genes were included in this study. Pathogenicity of MMR VUS was assessed by mRNA analysis and multifactorial likelihood calculations. Mutational analysis of 26 CRC-associated genes was performed by a customized NGS panel. Methylome analysis was perfomed by Infinium 450K array.

NGS analysis revealed the presence of two MMR truncating mutations not previously found. Five out of 15 MMR VUS were reclassified as pathogenic in 6 individuals. Methylome analysis identified one case harboring a constitutional *MLH1* epimutation. In addition, 13 predicted deleterious variants in other CRC-predisposing genes were found in 12 patients. Differentially methylated regions were not identified in samples from LLS patients compared to LS or healthy individuals.

In conclusion, the use of subexome gene panels combined with pathogenicity assessment of VUS allows the identification of deleterious MMR mutations as well as new LLS-candidate genes. Moreover, constitutional epimutations outside MMR 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 and SAF2015-68016-R), CIBERONC, RTICC Network (RD12/0036/0031 and RD12/0036/0008), the Spanish Association Against Cancer (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 Spanish Association Against Cancer (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.

Disclosures

The authors declare no conflict of interest.

Correspondence:

Gabriel Capellá, M.D., Ph.D. and Marta Pineda, Ph.D.
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 73193
E-mail: gcapella@iconcologia.net; mpineda@iconcologia.net

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 (Bellido *et al*, 2015; Elsayed *et al*, 2014; Palles *et al*, 2012; Rohlin, 2014; Spier *et al*, 2015; Valle *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 -either MMR loss of expression and/or microsatellite instability (MSI)- were included (Table S1), being 23 of them previously reported (Vargas-Parra et al., 2017). The IHC pattern of MMR protein was as follows: MLH1/PMS2 loss (n=56), MSH2/MSH6 loss (n=27), MSH6 loss (n=12), PMS2 loss (n=5) and MMR conserved expression/MSI (n=15). In patients with tumors showing loss of MLH1/PMS2 protein expression by IHC, somatic *MLH1*-hypermethylation and/or *BRAF* V600E mutation were previously discarded. Of note, 3 patients with non-informative tumor methylation analysis and wildtype *BRAF* were included (cases 7, 9 and 78). MMR candidate genes to be mutated according the IHC pattern 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 criteria (72,2%) and 11 Amsterdam criteria (9,6%) (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 global methylome analysis (Dámaso *et al*, 2018) (Table S2).

All patients were assessed at Cancer Genetic Counseling Units at 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

Blood DNA was collected from included individuals. FFPE blocks of normal colorectal mucosa and CRC tissue were also collected when available. Isolation of genomic DNA from blood was performed using FlexiGene DNA kit (Qiagen, Hilden, Germany) or Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. 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 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).

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), including the LLS cases previously reported (Vargas-Parra *et al*, 2017).

Array processing and data analysis were performed as previously described (Dámaso *et al*, 2018). In brief, PBL DNA integrity was evaluated by agarose gel electrophoresis. Purity of non-degraded samples was determined by spectrophotometry. Samples 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 delta-Ct values lower than 5 were restored using Infinium HD FFPE Restore kit (Illumina), following the manufacturer's instructions 1000 ng PBL and 500 ng FFPE DNAs 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 the methylated and unmethylated controls and one sample from each batch. Genome wide methylation profiling was performed using the Infinium Human Methylation 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 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 by Integrative Genome Viewer (Broad Institute). GRCh37/hg19 was used as the reference genome (date of release: February 2009). Only positions that reach FDR p-value<0.05 when comparisons are done between groups bigger than 10 samplesare considered.

Mismatch repair genes mutational analysis

Mutational analysis of coding regions of MMR genes

According to the IHC pattern in tumors, point 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 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 HGVS recommendations.

Direct sequencing of MMR promoter regions and 3'UTR of 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

The tumor of case 53, harboring a *POLE* variant within the exonuclease domain, was further characterized. Whole exome sequencing (WES) was carried out in tumor DNA using Agilent Sure Select Human All Exon v5 kit in a Hi-Seq2000 (coverage >100). Variants identified in blood DNA by WES were eliminated to obtain the number of accumulated somatic mutations in tumors and the relative presence of the described mutational signatures (Alexandrov & Stratton, 2014; Nik-Zainal *et al*, 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). T*he reaction was visualized with the EnVision*^{TM} *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 defined when >30% of the markers showed instability.

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.

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 S5).

Case ID	Results from previous MMR mutational analysis by Sanger sequencing / SSCP	Results from the analysis of CRC-predisposing 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	
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 (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)

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

B. Results obtained in 7 additional cases harboring MMR variants identified by previous Sanger sequencing

Case ID	Results from previous MMR mutational analysis by Sanger sequencing / SSCP	Results from the analysis of CRC-predisposing 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	
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**

RESULTS

Reassessment of germline genetic variants in MMR genes

The presence of missed MMR genetic alterations were reassessed in PBL 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 staining displayed loss of MLH1 protein expression in his first tumor, being non-informative the second one. Previous Single Strand Conformation Polymorphism 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 proteins expression. No mutation was identified by Sanger sequencing in *MLH1* and *MSH2* genes. The panel allowed the identification of a truncating mutation in *MSH6,* c.2219T>A (p.Leu740*) (Figure 1B). Finally additional MMR VUS (*MSH6* c.2092C>G, *MSH6* c.3150_3161dup, *PMS2* c.1320A>G and MSH2 c.2802G>A) were also detected (Table 1).

This reanalysis was complemented with sequencing of the MMR promoter regions that identified a *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 may direct 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 P248 kit revealed the presence of a *MSH2* exon 8 duplication in case 57.

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*. Splicing evaluation and stability analyses were possible for *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 c.1618_1620del variant carriers (cases 70 and 75, Figure 1D and 1E) 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) (Figure S1). This is in agreement with the allelic imbalance detected at the c.1618 position (Table 2A). The remaining 2 variants analyzed had no apparent effect on mRNA splicing and stability

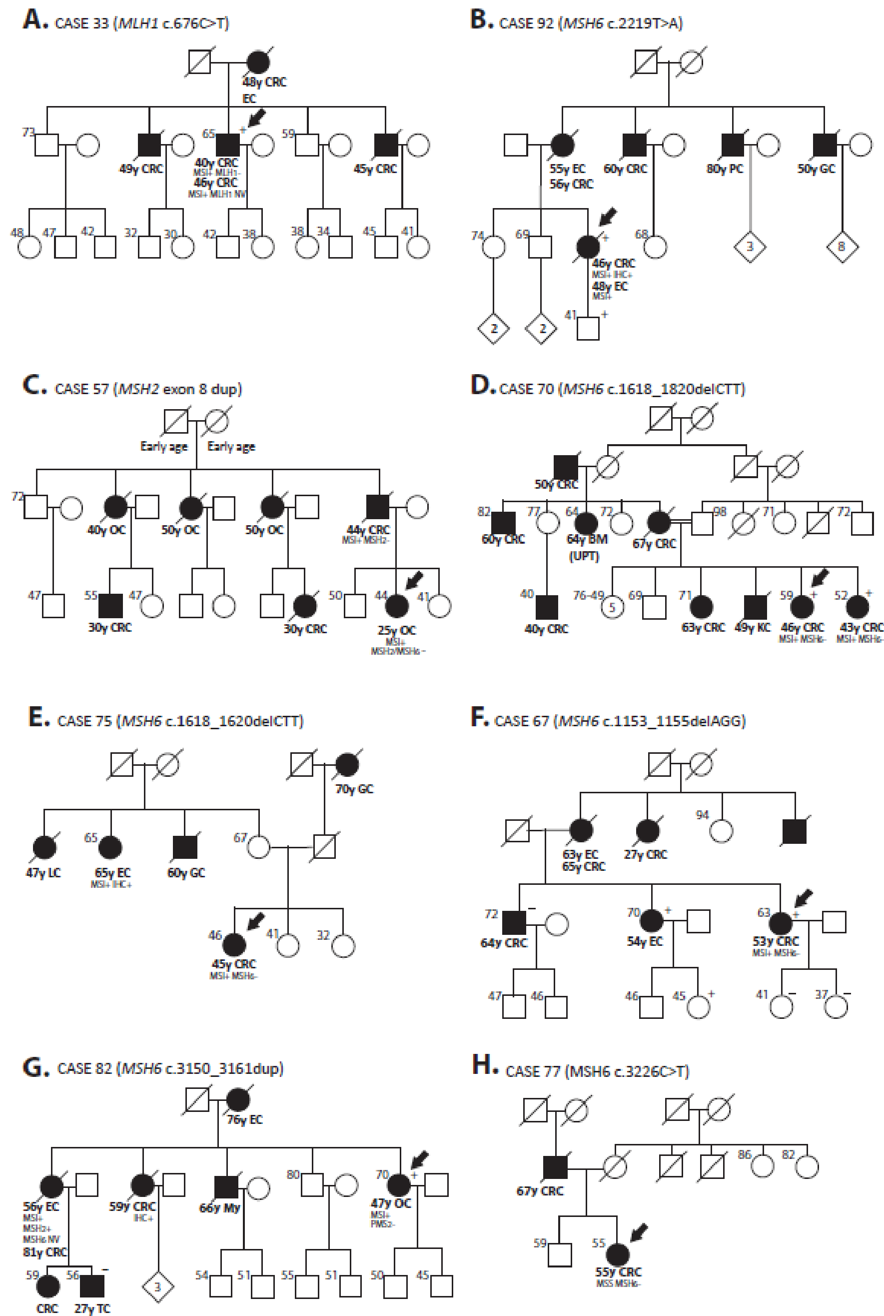


Figure 1: Pedigrees from patients reclassified as Lynch syndrome during the current study. Abbreviations: CRC=colorectal càncer, EC=endometrial càncer, PC=prostate càncer, GC=gàstric càncer, OC=ovarian càncer, BM(UTP)=brain metàstasis from unknown primary tumor, KC=kidney càncer, TC=testis càncer, My=mieloma, MSI+=microsatellite instability, NV=No valuable, +=validated mutation carrier, -=no-carrier.

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 S4 and S5)

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.1607_3172del;=; p.Ser536_Asp1058delinsAsn;Leu540del	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 4	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 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	Ascertain-ment	Cancer (age)	MSI/ IHC status	MSI CRC LR	Tumor Characteristics LR	Bayes	Segrega-tion 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	105,9181058	105,9181058	0,990647047	**Class 5 Pathogenic**
					AF1	clinic	EC (59)	MSH6 loss			7,0805					
					AF1	clinic	CRC (59)	MSH6 loss								
c.1618_1620delCTT; p.Ser536_Asp1058delinsAsn; Leu540del	NR /NR	Not reported	0,95943958	0,9	70	clinic	CRC (46)	MSI-H & MSH6 loss		6,96	1,84534682	1,84534682	12,84361387	115,5925248	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	18670,30345	168032,7311	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 .

(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 (Figure 1F and 1G, respectively) had been identified in additional 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 (Figure 1H), 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 2C 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 being also classified as VUS. The latter MMR variant and the 8 additional ones 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).

Identification of variants in other CRC-predisposing genes

Additionally, subexome analysis identified rare germline variants in other CRC-predisposing genes in 32 LLS cases (Table S5). Thirteen of them were variants predicted as pathogenic by *in silico* tools in genes such as *APC, MUTYH, MSH3* and *FAN1* (Table 3 and S6). Interestingly, four variants were located in the *MSH3* (Table 3), and two of them coexisted in *cis* in the same patient (case 74; Figure S4): the variant c.2732T>G (p.Leu911Trp) affects a highly conserved residue along MutS proteins, and the variant c.685T>C (p.Tyr229His) is located next to the DNA recognition domain and affects a highly conserved in MSH3 (Adzhubei *et al*, 2010; Obmolova *et al*, 2000). Immunohistochemical staining showed conserved MSH3 nuclear expression in normal and tumor tissue from case 74, while tetranucleotide repeats analysis displayed instability in 2 out of 6 microsatellites, indicating EMAST (Figure S4). It is noteworthy that the *EXO1* c.2212-1G>A variant identified in case 58 cause an in-frame deletion of 6 amino acids in the MSH2 interaction domain (Table 3).

Table 3. Variants identified in non LS-associated genes and predicted pathogenic by *in silico* predictions (see Table S5 and S6)

Case ID	Variant calling			rs ID	MAF	In silico predictions						ClinVar Classification / Bibliography
						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)/++

The *FAN1* c.149T>G (p.Met50Lys) 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 recently reported in controls in The Genome Aggregation Database (GnomAD).

Furthermore, the *POLE* c.898A>G (p.Ile300Val) variant was found in patient 53 predicted as benign by *in silico* tools. However, since it was located within the exonuclease domain, whole exome sequencing in tumor was performed in order to elucidate its causality. Unfortunately tumor analysis displayed neither POLE signature nor hypermutability (data not shown).

Constitutional epigenetic alterations in MMR genes

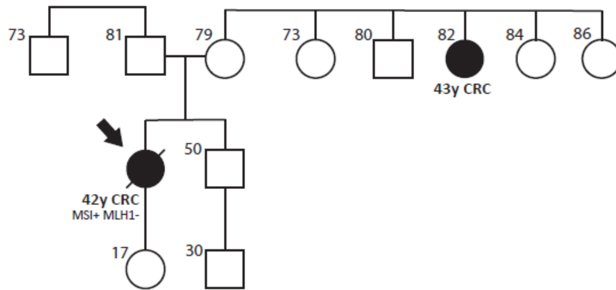
Methylome analysis was firstly used to evaluate the existence of constitutional epigenetic alterations in MMR genes. Interestingly, blood DNA from case 7 displayed *MLH1* promoter hypermethylation that was further validated in blood using MS-MLPA (mean methylation in C/D regions 48%; data not shown). She was a woman who 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 (Figure 2C). No other cases with hypermethylation in MMR genes were found.

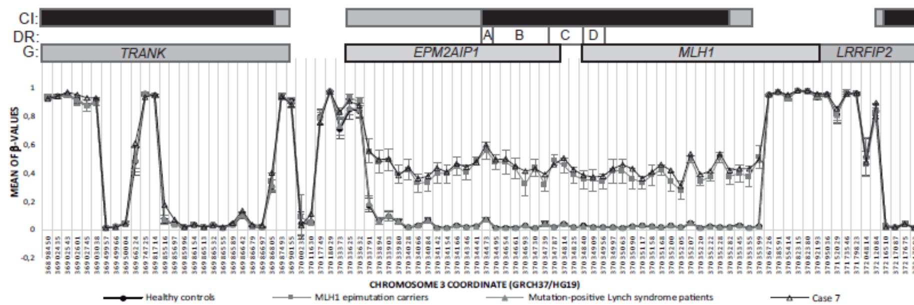
Global epigenetic characterization of Lynch-like cases

We also performed a global epigenetic characterization of LLS cases in order to evaluate the relative contribution of this type of aberration. 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 sole DM region identified in blood when 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 S5). 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.





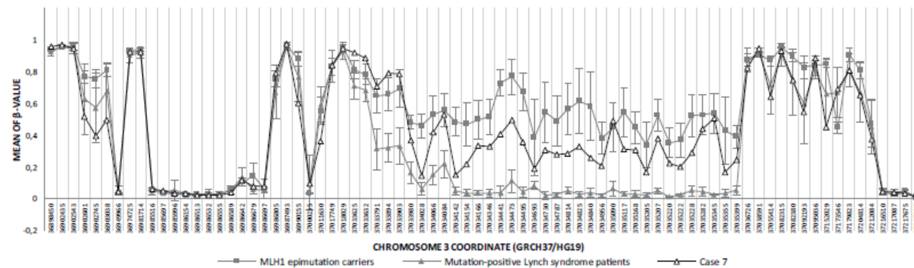
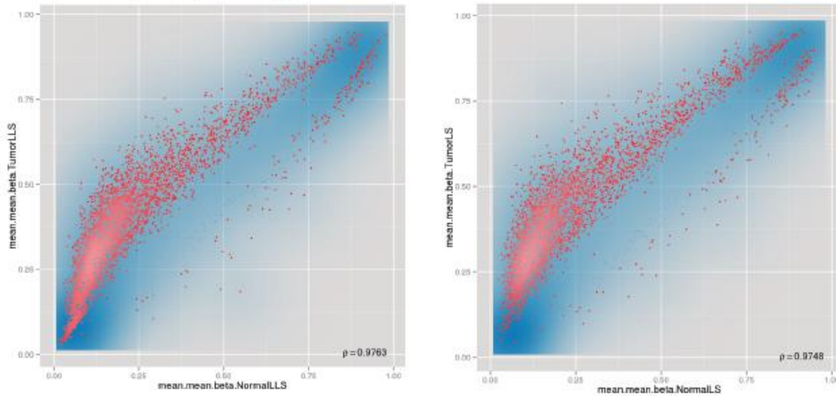
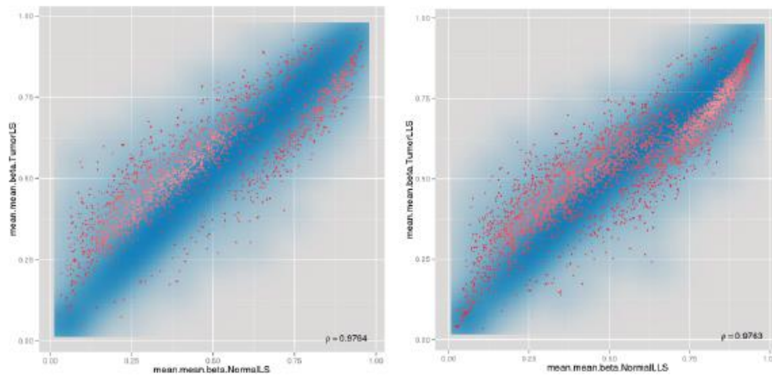


Figure 2: Identification of a new case of constitutional *MLH1* epimutation. **(A)** Pedigree of case 7. Representation of mean β-values in blood DNA **(B)** and FFPE normal colorectal mucosa **(C)** from case 7 against *MLH1* epimutation carriers, mutation-positive Lynch syndrome patients and healthy 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.

A. Differentially methylated CpG islands



B. Differentially methylated genes



C.

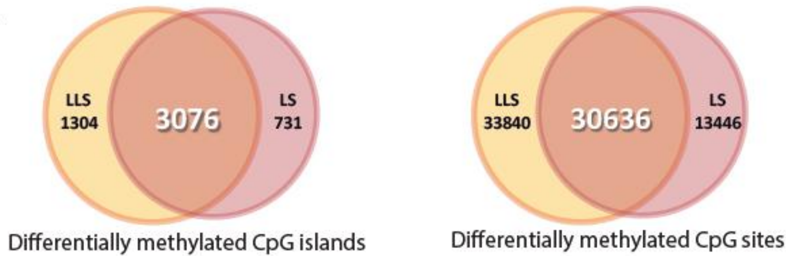


Figure 3: Scatterplot of the normalized mean B-values obtained using the Infinium 450k Human Methylation array to identify differentially methylated CpG islands **(A)** and genes **(B)** during tumorogenesis of LLS cases (left) and LS controls (right). The transparency corresponds to point density. One % 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 subexome analysis 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 13% (15/115) reclassification rate. Also, predicted deleterious variants in other CRC predisposing genes were found. Finally, other clinically relevant DMRs were not identified.

In the present work a customized NGS panel for the analysis of CRC predisposition genes allowed us to identify 2 previously missed *bona fide* MMR pathogenic variants in Amsterdam families. Fifteen additional MMR variants (depicted by NGS or Sanger sequencing) 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 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 a *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 clinical

actionability of germline *MUTYH* monoallelic mutations is limited and a matter of ongoing controversy. A modest increased susceptibility to cancer has been observed in several studies (Croitoru *et al*, 2004; Jones *et al*, 2009; Khalaf *et al*, 2013; Win *et al*, 2014), whereas others have failed to replicate these findings (Balaguer et al 2007; Lubbe et al 2009; Ma, Zhang and Zheng 2014; Theodoratou et al 2010).

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. Met50Lys) 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 (the 73rd primary epimutation carrier reported word-wide; (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 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; Ligtenberg *et al*, 2013). 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 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

Bellido F, Pineda M, Aiza G, Valdés-Mas R, Navarro M, Puente DA, Pons T, González S, Iglesias S, Darder E, Piñol V, Soto JL, Valencia A, Blanco I, Urioste M, Brunet J, Lázaro C, Capellá G, Puente XS, Valle L (2015) POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. *Genetics In Medicine* **18:** 325

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

Croitoru M, Cleary S, Di Nicola N, Manno M, Selander T, Aronson M, Redston M, Cotterchio M, Knight J, Gryfe R, Gallinger S (2004) Association Between Biallelic

and Monoallelic Germline MYH Gene Mutations and Colorectal Cancer Risk. *JNCI: Journal of the National Cancer Institute* **96**(21)**:** 1631–1634

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

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

Jones N, Vogt S, Nielsen M, Christian D, Wark PA, Eccles D, Edwards E, Evans DG, Maher ER, Vasen HF, Hes FJ, Aretz S, Sampson JR (2009) Increased Colorectal Cancer Incidence in Obligate Carriers of Heterozygous Mutations in MUTYH. *Gastroenterology* **137**(2)**:** 489-494.e1

Khalaf R, Jones C, Strutt W, Williamson P (2013) Colorectal Cancer in a Monoallelic MYH Mutation Carrier. *Journal of Gastrointestinal Surgery* **17**(8)**:** 1500-1502

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

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

Ligtenberg MJL, Kuiper RP, Geurts van Kessel A, Hoogerbrugge N (2013) EPCAM deletion carriers constitute a unique subgroup of Lynch syndrome patients. *Familial Cancer* **12**(2)**:** 169-174

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

Palles C, Cazier J-B, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I, Sherborne A, Chubb D, Carvajal-Carmona LG, Ma Y, Kaur K, Dobbins S, Barclay E, Gorman M, Martin L, Kovac MB, Humphray S, The CC, The WGSC, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsey SE, Thomas HJW, McVean G, Houlston RS, Tomlinson I (2012) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics* **45:** 136

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

Rohlin A, Zagoras, T., Nilsson, S., Lundstam, U., Wahlström, J., Hultén, L., Martinsson, T., Karlsson, G. B., Nordling, M (2014) A mutation in POLE predisposing to a multi-tumour phenotype. In *International Journal of Oncology* Vol. 45. pp 77-81

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

Spier I, Holzapfel S, Altmüller J, Zhao B, Horpaopan S, Vogt S, Chen S, Morak M, Raeder S, Kayser K, Stienen D, Adam R, Nürnberg P, Plotz G, Holinski-Feder E, Lifton RP, Thiele H, Hoffmann P, Steinke V, Aretz S (2015) Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. *International Journal of Cancer* **137**(2)**:** 320-331

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

Valle L, Hernández-Illán E, Bellido F, Aiza G, Castillejo A, Castillejo M-I, Navarro M, Seguí N, Vargas G, Guarinos C, Juarez M, Sanjuán X, Iglesias S, Alenda C, Egoavil C, Segura Á, Juan M-J, Rodriguez-Soler M, Brunet J, González S, Jover R, Lázaro C, Capellá G, Pineda M, Soto JL, Blanco I (2014) New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. *Human Molecular Genetics* **23**(13)**:** 3506-3512

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

Win AK, Dowty JG, Cleary SP, Kim H, Buchanan DD, Young JP, Clendenning M, Rosty C, MacInnis RJ, Giles GG, Boussioutas A, Macrae FA, Parry S, Goldblatt J, Baron JA, Burnett T, Le Marchand L, Newcomb PA, Haile RW, Hopper JL, Cotterchio M, Gallinger S, Lindor NM, Tucker KM, Winship IM, Jenkins MA (2014) Risk of Colorectal Cancer for Carriers of Mutations in MUTYH, With and Without a Family History of Cancer. *Gastroenterology* **146**(5)**:** 1208-1211.e5

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

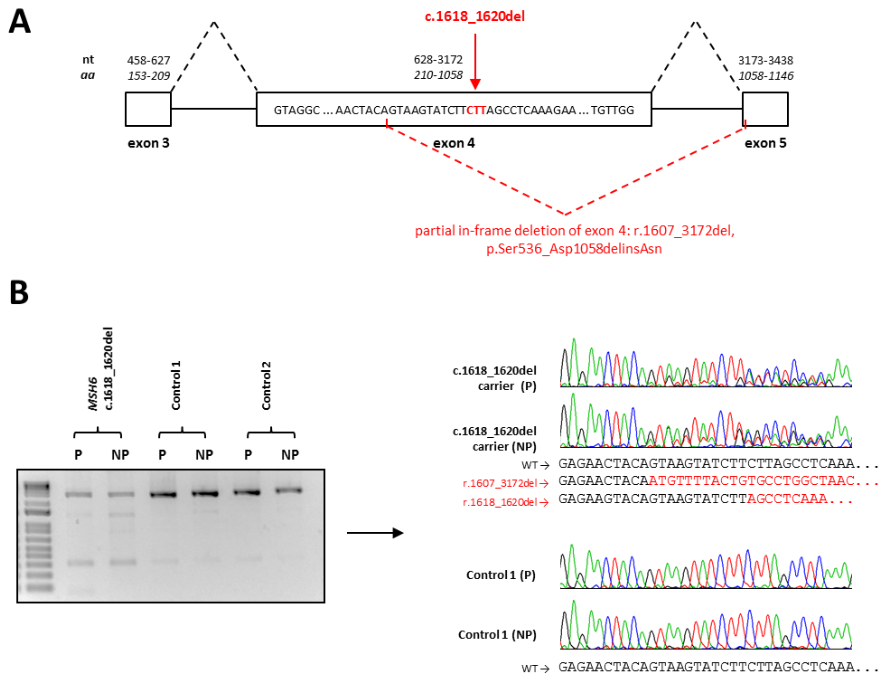
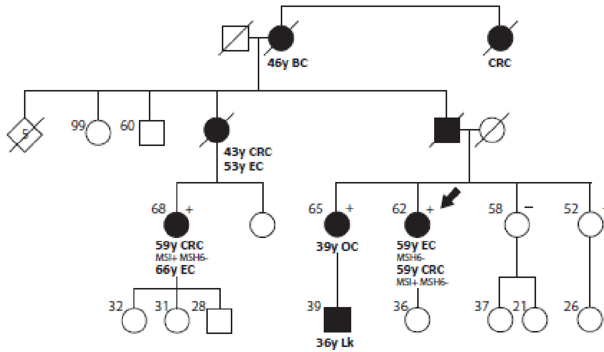
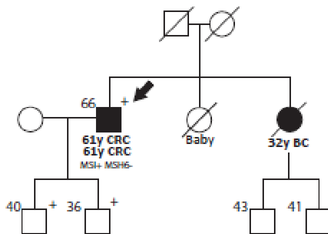


Figure S1. Splicing analysis of *MSH6* c.1618_1620del variant. A) Schematic overview of *MSH6* exons. **B)** Left: agarose gel showing RT-PCR products. Right: direct sequencing of RT-PCR products showing the wt transcript and the aberrant transcript corresponding to a partial in-frame deletion of exon 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)

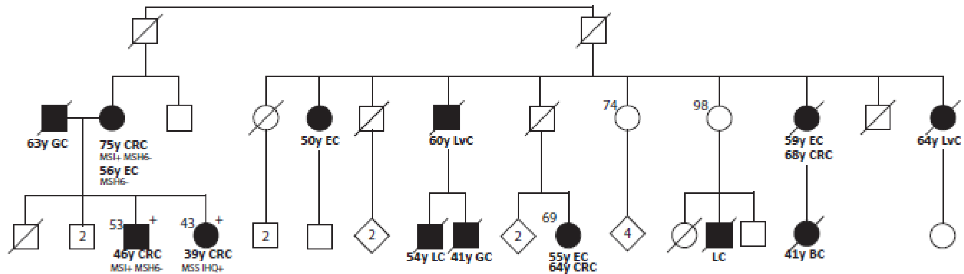


Figure S2: Pedigrees from families included for pathogenicity assessment of MMR VUS by multifactorial analysis. Abbreviations: CRC=colorectal cancer, BC=breast cancer, LvC=liver cancer, LC=lung cancer, EC=endometrial cancer, GC=gastric cancer.

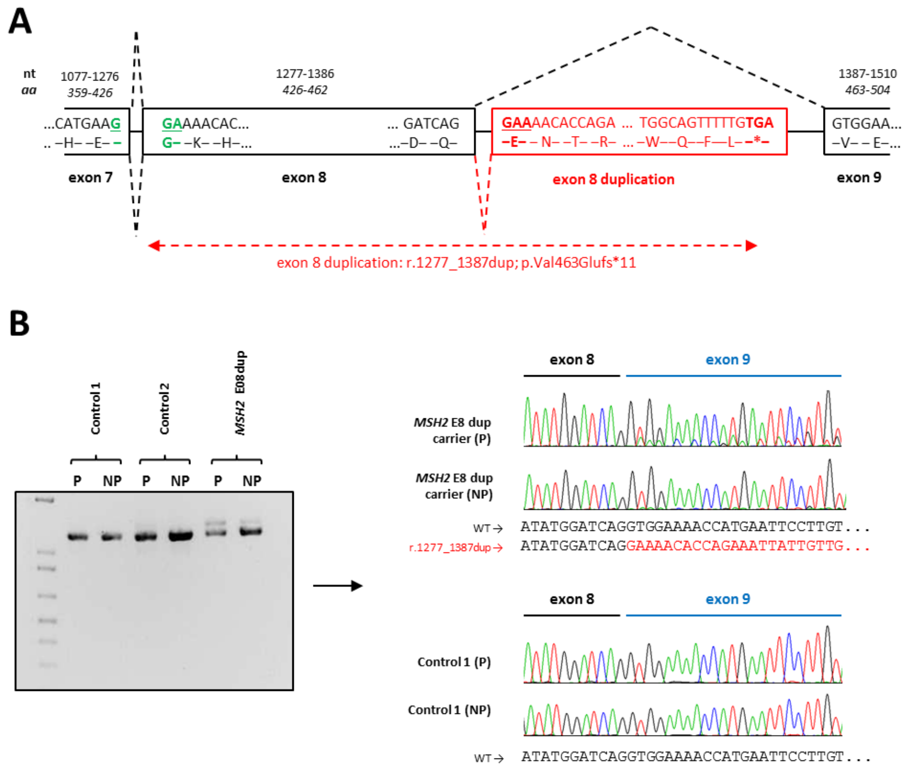
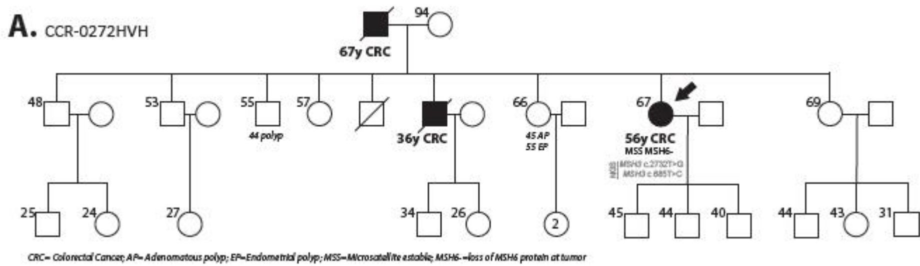
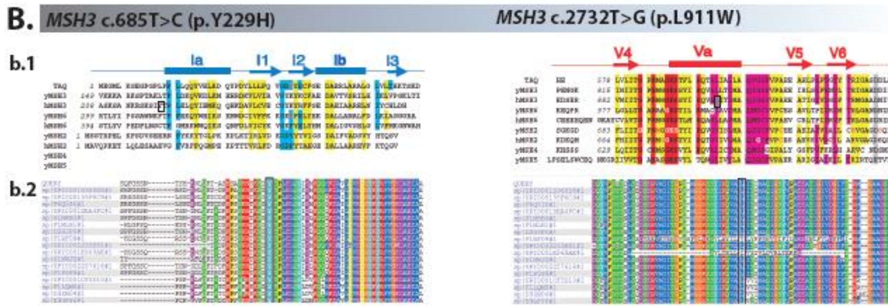
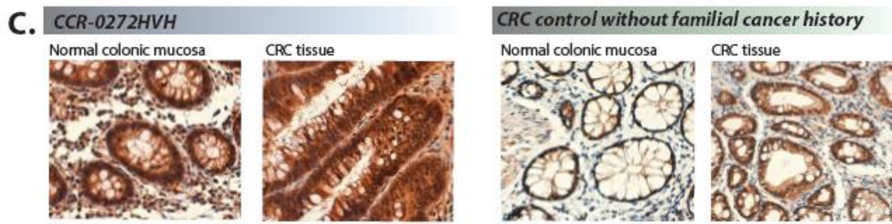


Figure S3: Splicing analysis of *MSH2* exon 8 duplication. A) Schematic overview of *MSH2* exons. **B)** Left: agarose gel showing RT-PCR products. Right: direct sequencing of RT-PCR products showing the wt transcript and the aberrant transcript corresponding to a duplication of exon 8. Abbreviations: nt, nucleotide sequence; aa, amino acid sequence; P, puromycin; NP, no puromycin.







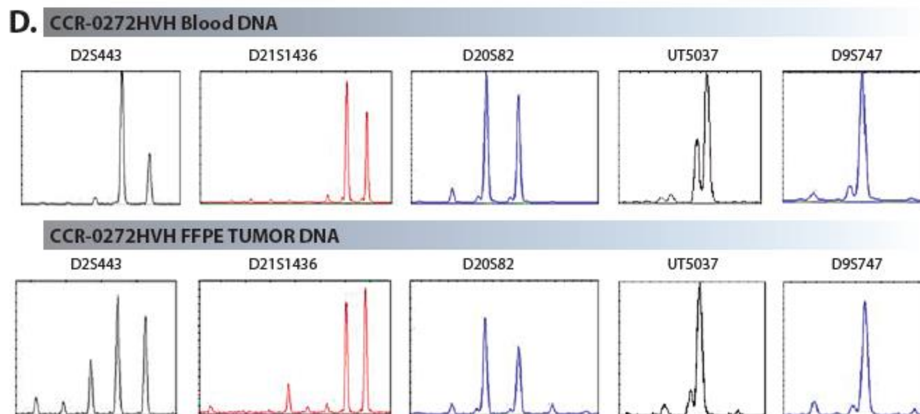


Figure S4: Pathogenicity assessment of *MSH3* variants found in case 57. **(A)** Pedigree of case 7. **(B)** Structure-based sequence alignment of Mut S homologues (upper panel) and multiple sequence alignment (lower panel). **(B.1)** MutS homologues from human (hMSH), S. cerevisiae (yMSH) and T. aquaticus 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 57 and a CRC control without family history of cancer. **(D)** EMAST analysis in blood and FFPE tumor DNA from case 57.

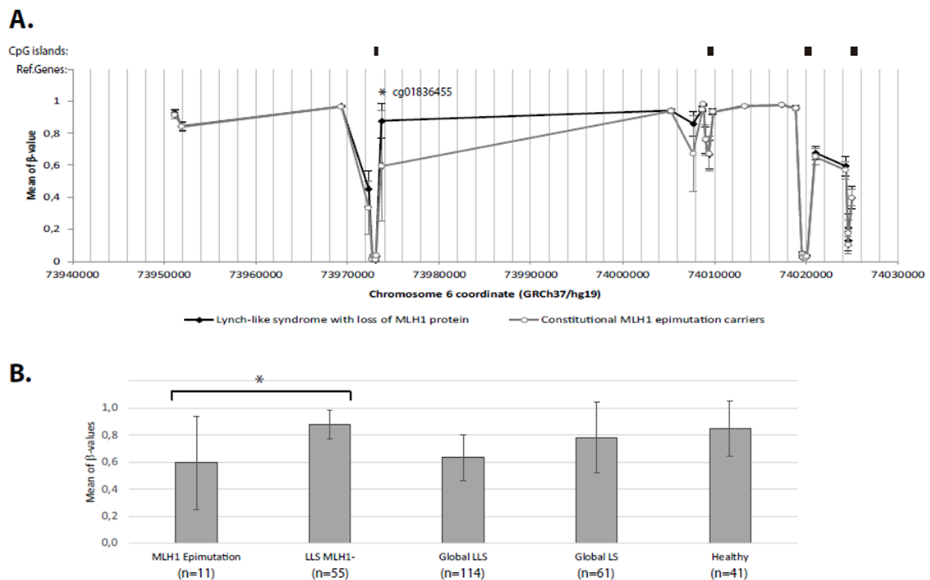


Figure S5: Differentially methylated CpG site was found inside *KHDC1* gene. **(A)** Methylation pattern of the locus in MLH1-deficient Lynch-like syndrome patients (black) and constitutional MLH1 epimutation carriers (grey). The location of reference genes and CpG islands is represented above according 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. Significative 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	61	NP	WT	NP	NP	NP	NP	NP	P+CR	NP	NP	NP	-	B
			CRC	68	MSI+		UM	N	P	P	N						
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

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	
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	69	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B
			EC	69	NP	NP	NP	NP	NP	NP	NP						
50	M	B1	CRC	41	MSI+	WT	NP	N	P	P	N	P+CR	NP	NP	NP	-	B, N, T
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	NP	NM	NP	NP	-	B
			CRC	81	NP	NP	NP	NP	NP	NP	NP						
			CRC	81	NP	NP	NP	NP	NP	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	CR	P+CR	P+CR	NP	*MSH2* c.2045C>G; p.Thr682Ser (Class 3*)	B
			CRC	61	NP	NP	NP	P	N	N	NP						
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	NP	P+CR	P+CR	NP	-	B, N, T,
			CRC	63	NS	NP	NP	NP	NP	NP	NP						
			BC	78	NS	NP	NP	NP	NP	NP	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	NP	PR	P+CR	NP	MSH6 c.3296T>A; p.Ile1099Asn (Class 3 *)	B
			CRC	53	NP	NP	NP	NP	NP	NP	NP						
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	NP	PR	P+CR	NP	-	B
			BC	69	NP	NP	NP	NP	NP	NP	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

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	
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
#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
#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	P	N	N	NP	NP	P+CR	P+CR	NP	-	B
#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
#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**
with somatic MLH1/PMS2 loss	28	29	56	56±16	5	46±7	11	42±11	3
with somatic MSH2/MSH6 loss	10	17	27	56±15	7	57±13	6	50±17	3
with somatic MSH6 loss	4	8	12	60±11	1	45±0	4	51±6	1
with somatic PMS2 loss	2	3	5	57±6	1	47±0	1	57±0	0
no loss	9	5	14	63±13	1	58±0	3	60±5	1
Mutation-positive Lynch syndrome patients (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					
						SSF [0-100]		MaxEnt [0-16]		NNSPLICE [0-1]		GeneSplicer [0-21]		HSF [0-100]								
	Gene	cDNA change	Predicted protein change	rs ID/ ExAC/ESP	Splice Site	wildtype	variant	wildtype	variant	wildtype	variant	wildtype	variant	wildtype	variant	Interpretation	SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	Interpretation
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**	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 75	*MSH6* *MSH6*	c.1618_1620delCTT c.1618_1620delCTT	p.Leu540del	NR/NR/NR	DSS	–		–		–		–		–		No changes	NP	NP	NP	NP	**D (-10.18)**	Benign
					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**	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**	**Probably damaging**	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

Table S5 (Cont.)

Case ID	Variant calling				Coverage		rs ID	MAF	In silico predictions						ClinVar Classification
										Protein function					
	Gene	Transcript	cDNA change	Predicted protein change	Allele Frequency	Read Depth		ExAC/ESP	Splicing	SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	
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.**	**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

Table S5 (Cont.)

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	
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
76	*POLE*	NM_006231.3	c.3938C>T	p.Thr1313Met	0.473	8828	NA	0.001/NR	No changes	Tolerated (0.06)	Polymorphism (1)	Benign (0.133)	Benign (0.017)	NP	Not reported
	SETD2	NM_014159.6	c.2730A>G	p.Ala910=	0.508	1865	rs377041426	0.016/0.000	No changes	NA	NA	NA	NA	NA	Likely bening (1) / +
81	***BUB1***	**NM_004336.4**	**c.2473C>T**	**p.Pro825Ser**	0.474	3567	rs748392521	NR/NR	No changes	**Deleterious (0)**	**Disease causing (1)**	**Probably damaging (1.000)**	**Probably damaging (0.997)**	NP	Not reported
82	***MSH6***	**NM_000179.2**	**c.3150_3161dup**	p.Val1051_Ile1054dup	0.428	1607	NA	NR/NR	NA	NA	NA	NA	NA	**Deleterious**	Not reported

Table S5 (Cont.)

Case ID	Variant calling				Coverage		rs ID	MAF	In silico predictions						ClinVar Classification
										Protein function					
	Gene	Transcript	cDNA change	Predicted protein change	Allele Frequency	Read Depth		ExAC/ESP	Splicing	SIFT (score)	Mutation Taster (p-value)	Polyphen2 /HumDiv (score)	Polyphen2 /HumVar (score)	Provean	
85	*FAN1*	NM_014967.4	c.418G>T	p.Asp140Tyr	0.451	4939	rs761776412	NR/NR	No changes	**Deleterious (0.02)**	Polymorphism (1)	Benign (0.030)	Benign (0.019)	NP	Not reported
	MSH3	**NM_002439.4**	**c.3072G>C**	**p.Gln1024His**	0.496	1007	rs147640909	0.000/0.000	**Loss of a DSS / Inconclusive at ASS**	Tolerated (0.39)	Polymorphism (0.996)	Benign (0.007)	Benign (0.013)	NP	Not reported
	PMS2	NM_000535.5	c.1320A>G	p.Pro440=	0.493	3059	rs138697590	NR/0.000	No changes	NA	NA	NA	NA	NA	Benign/likely benign (5) vs VUS (1) / +
92	*SETD2*	NM_014159.6	c.393C>T	p.Ser131=	0.42	4486	rs556549304	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
	MSH6	**NM_000179.2**	**c.2219T>A**	**p.Leu740Ter**	0.373	3385	rs745483465	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
93	*PMS1*	NM_000534.4	c.628C>G	p.Leu210Val	0.51	3824	rs757185669	NR/NR	No changes	Tolerated (0.44)	**Disease causing (0.912)**	Benign (0.125)	Benign (0.114)	NP	Not reported
	BUB3	NM_004725.3	c.*160A>T	p.?	0.29	850	rs566010640	NR/NR	NA	NA	NA	NA	NA	NA	Not reported
95	*FAN1*	NM_014967.4	c.1577+476C>T	p.?	0.49	2284	rs138151213	NR/NR	No changes	NA	NA	NA	NA	NA	Not reported
96	***APC***	**NM_000038.4**	**c.7514G>A**	**p.Arg2505Gln**	0.501	1765	rs147549623	0.001/0.001	No changes	**Deleterious (0.04)**	**Disease causing (1)**	**Probably damaging (1.000)**	**Probably damaging (0.961)**	NP	Benign/Likely benign (8) / ++
97	*APC*	NM_000038.4	c.3949G>C	p.Glu1317Gln	0.496	2043	rs1801166	0.004/0.009	No changes	Tolerated (0.49)	**Disease causing (0.995)**	Benign (0.001)	Benign (0.003)	NP	Benign/Likely benign(13) vs Pathogenic (2) vs Uncertain
	APC	NM_000038.5	c.5265G>A	p.Ala1755=	0.419	269	rs34506289	NR/0.010	No changes	NA	NA	NA	NA	NA	Benign/Likely benign (11) / ++
98	*MSH2*	NM_000251.2	c.2802G>A	p.Thr934=	0.476	1064	rs150259097	0.000/0.001	No changes	NA	NA	NA	NA	NA	Likely benign (5) vs VUS (2) / +++

Table S6: Detailed in silico predictions of predicted pathogenic variants identified in non LS-associated genes. 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	=	0,83	=	14,26	14.00 (-1.8%)	88,88	=							
30	*APC*	c.7936C>G	p.Gln2646Glu	NR/NR/NR	DSS											No changes	D (0.02)	D (1)	PsD (0.688)	B (0.182)	NP	Damaging
					ASS																	
39	*FAN1*	c.149T>G	p.Met50Arg	rs148404807 /0.002/0.003	DSS											No changes	T (0.08)	D (1)	PrD (0.991)	PsD (0.690)	NP	Damaging
					ASS																	
55	*PMS1*	c.497A>C	p.Lys166Thr	NR/NR/NR	DSS	81,76	=	8,14	=	0,98	=			84,71	=	No changes	D (0)	D (1)	PsD (0.757)	PsD (0.599)	NP	Damaging
					ASS	85,7	=	11,95	=	0,91	=	5,81	6.19 (+6.4%)	89,48	=							
					ASS*			—	0,09					66,74	76.13 (+14.1%)							
58	*EXO1*	c.2212-1G>C	p.Val738_Lys743del	rs4150000 /0.0019/0.00	DSS											Loss of an ASS	NA	NA	NA	NA	NA	NA
					ASS	91,77	—	8,44	—	0,62	—			93,10	—							
59	*APC*	c.1966C>G	p.Leu656Val	rs577466163 /NR/NR	DSS											Gain of a DSS	D (0)	D (1)	PrD(0.999)	PrD (0.998)	NP	Damaging
					DSS*	—	72,54	—	6,29	—	0,71			—	80.95							
					ASS	93,56	=	7,53	=	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											No changes	NA	NA	NA	NA	D (-7.78)	Damaging
					DSS*			3,6	=	0,56	=	4,61	3.06 (-33.6%)	75,54	=							
					ASS*	72,47	=	3,69	4.44 (+20.3%)					83,14	=							
					ASS*									74.58 ⇒ —	—							
					ASS*	77,67	=	1,06	1.58 (+48.9%)					87,82	=							
62 & 74	*MSH3*	c.2732T>G	p.Leu911Trp	rs41545019 /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
					ASS	95,37	=	8,44	=	0,94	=	6,12	6.33 (+3.4%)	89,83	=							
					ASS*	70,86	=	4,17	2.56 (-38.5%)					81,32	=							
74	*MSH3*	c.685T>C	p.Tyr229His	NR/NR/NR	DSS											No changes	D (0.01)	D (0.999)	PrD (1.000)	PrD (0.973)	NP	Damaging
					ASS																	
81	*BUB1*	c.2473C>T	p.Pro825Ser	rs748392521 /NR/NR	DSS											No changes	D (0)	D(1)	PrD (1.000)	PrD (0.997)	NP	Damaging
					ASS	86,44	=	8,26	=	0,77	0.80 (+4.0%)	5,58	5.54 (-0.8%)	89,5	=							
85	*MSH3*	c.3072G>C	p.Gln1024His	rs147640909 /0.000/0.000	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
					DSS*	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	=							
					ASS*	79,71	—	3,82	—					89,25	—							
96	*APC*	c.7514G>A	p.Arg2505Gln	rs147549623 /0.001/0.001	DSS											No changes	D (0.04)	D (1)	PrD (1.000)	PrD (0.961)	NP	Damaging
					ASS																	

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	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	*3/CTB-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*
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*

Table S8. B) Differentially methylated CpG sites (Cont.)							
CG ID	**Chromosome**	**Start**	**Comparison Group**	**Mean difference**	**p.value**	**FDR p.value**	**Gene**
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) Differentially methylated CpG sites (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*

ARTÍCULO 3

Primary constitutional *MLH1* epimutations: a focal epigenetic event

Hipótesis: La completa caracterización de las epimutaciones primarias en *MLH1* podría facilitar el descubrimiento de su mecanismo causal.

Objetivo: Caracterizar las epimutaciones en *MLH1* mediante el estudio genético y epigenético de la región aberrantemente hipermetilada y otros loci genómicos.

Resumen de resultados obtenidos: En este trabajo se incluyeron 12 portadores de epimutación en *MLH1* ya demostradas Además, también se incluyeron 61 pacientes con síndrome de Lynch y 41 controles sanos como controles en el estudio del metiloma. No se identificaron variantes puntuales ni estructurales en *cis* en el alelo epimutado en diez portadores. En dos de estosse demostró el borrado de la metilación intergeneracional, lo que sugiere que la epimutación es primaria. En otros 2 casos, se encontraron variaciones en el número de copia en regiones situadas fuera del locus *MLH1*, aunque no se pudo determinar si estaban relacionadas con la epimutación. La isla CpG que engloba los promotores de *EPM2AIP1-MLH1* fue la única región diferencialmente metilada del genoma en portadores de epimutación en *MLH1* cuando fueron comparados frente a controles. Esta región diferencialmente metilada es distinta a la identificada en tumores esporádicos con hipermetilación somática de *MLH1*, lo que sugiere que el mecanismo causal es diferente en ambos casos.

Las epimutaciones constitucionales primarias de *MLH1* son una alteración epigenética focal que tiene lugar alrededor de la isla CpG de *EPM2AIP1-MLH1* y no se asocian a variantes genéticas en cis. Es necesaria una caracterización molecular más profunda de estos casos para determinar el mecanismo causal de las epimutaciones primarias de *MLH1* y su heredabilidad/reversibilidad.

Contribución del doctorando: Coordinación y recogida de muestras de centros externos. Gestión de los bloques de parafina, corte y tinción de laminillas, macrodisección y extracción del DNA del tejido parafinado. Diseño y puesta a punto de ensayos. Secuenciación y estudios de metilación (secuenciación con bisulfito, MS-MLPA). Secuenciación clonal con bisulfito. Inmortalización de linfocitos. Realización de ensayos de impacto funcional. Estudio de cosegregación en una familia. Análisis e interpretación de resultados de array CGH. Elaboración de las matrices de comparación del estudio del metiloma. Análisis, interpretación y discusión de resultados. Preparación de tablas y figuras. Escritura del manuscrito.

FULL PAPER

BJC

British Journal of Cancer (2017), 1–11 | doi: 10.1038/bjc.2017.466

BJC OPEN

Keywords: constitutional *MLH1* epimutation; Lynch syndrome; methylation; focal epigenetic event

Primary constitutional *MLH1* epimutations: a focal epigenetic event

Estela Damaso[1], Adela Castillejo[2], María del Mar Arias[3], Julia Canet-Hermida[1], Matilde Navarro[1], Jesús del Valle[1], Olga Campos[1], Anna Fernandez[1], Fátima Marín[1], Daniela Turchetti[4], Juan de Dios García-Díaz[5], Conxi Lazaro[1], Maurizio Genuardi[6], Daniel Rueda[7], Ángel Alonso[3], Jose Luis Soto[2,8], Megan Hitchins[9], Marta Pineda[1,10] and Gabriel Capella*[,1,10]

[1]*Hereditary Cancer Program, Catalan Institute of Oncology - Bellvitge Biomedical Research Institute (ICO-IDIBELL), CIBERONC, Avinguda Gran Via de l'Hospitalet 199-203, 08908 L'Hospitalet de Llobregat, Barcelona, Spain;* [2]*Hereditary Cancer Program Valencian Region, Molecular Genetics Laboratory, Elche University Hospital, Camino de la Almazara 11, 03203 Elche, Alicante, Spain;* [3]*Genetics Service, Complejo Hospitalario de Navarra, Pamplona, Calle de Irunlarrea 3, 31008 Navarra, Spain;* [4]*Dipartimento di Scienze Mediche e Chirurgiche, Alma Mater Studiorum - Università di Bologna, Via Massarenti 11, 40138 Bologna, Italy;* [5]*Unidad de Genética Clínica, Servicio de Medicina Interna, Hospital Universitario Príncipe de Asturias, Carretera Alcalá-Meco, 28805 Alcalá de Henares, Madrid, Spain;* [6]*Istituto di Medicina Genomica, Università Cattolica del Sacro Cuore–Fondazione Policlinico Universitario Agostino Gemelli, Largo Agostino Gemelli 8, 00168 Rome, Italy;* [7]*Hereditary Cancer Genetic Diagnostic Laboratory, Doce de Octubre University Hospital, Avenida de Córdoba, 28041 Madrid, Spain;* [8]*Alicante Institute for Health and Biomedical Research (ISABIAL-FISABIO Foundation), Alicante, Spain and* [9]*Department of Medicine, Division of Oncology, Stanford University, 1291 Welch Road, Stanford, CA 94305, USA*

Background: Constitutional *MLH1* epimutations are characterised by monoallelic methylation of the *MLH1* promoter throughout normal tissues, accompanied by allele-specific silencing. The mechanism underlying primary *MLH1* epimutations is currently unknown. The aim of this study was to perform an in-depth characterisation of constitutional MLH1 epimutations targeting the aberrantly methylated region around MLH1 and other genomic loci.

Methods: A total of 12 *MLH1* epimutation carriers, 61 Lynch syndrome patients and 41 healthy controls were analysed by Infinium 450K array, and targeted molecular techniques were used to characterise the *MLH1* epimutation in carriers and their inheritance pattern.

Results: No nucleotide or structural variants were identified in *cis* on the epimutated allele in 10 carriers, in which intergenerational methylation erasure was demonstrated in 2, suggesting primary type of epimutation. The CNVs outside the *MLH1* locus were found in two cases. The EPM2AIP1-MLH1 CpG island was identified as the sole differentially methylated region in *MLH1* epimutation carriers compared with controls.

Conclusions: Primary constitutional *MLH1* epimutations arise as a focal epigenetic event at the EPM2AIP1-MLH1 CpG island in the absence of *cis*-acting genetic variants. Further molecular characterisation is needed to elucidate the mechanistic basis of *MLH1* epimutations and their heritability/reversibility.

*Correspondence: Dr M Pineda; E-mail: mpineda@iconcologia.net or Dr G Capellá; E-mail: gcapella@iconcologia.net
[10]These authors contributed equally to this work and share senior authorship.

Received 28 July 2017; revised 20 November 2017; accepted 27 November 2017

© The Author(s) named above

Lynch syndrome (LS) is characterised by an increased risk for colorectal cancer (CRC) and cancers of the endometrium, ovary, stomach, small intestine, hepatobiliary tract, urinary tract, brain and skin (Lynch *et al*, 2015). It is caused by a germline genetic variant within a mismatch repair (MMR) gene, *MLH1* (*MutL homologue 1*), *MSH2* (*MutS homologue 2*), *MSH6* (*MutS homologue 6*) or *PSM2* (*PMS1 homologue 2*) or a terminal deletion of *EPCAM* (*epithelial cell adhesion molecule* gene) with consequent epigenetic inactivation of *MSH2*. In a small proportion of LS patients, the cancer predisposition is caused by a constitutional epimutation of *MLH1*, in which one allele of the CpG island promoter is aberrantly hypermethylated throughout normal tissues with associated loss of expression from this allele (Hitchins, 2015).

Two types of constitutional *MLH1* epimutation have been defined: secondary that are linked in *cis* to a genetic alteration and follow an autosomal dominant pattern of inheritance (Renkonen *et al*, 2003; Morak *et al*, 2008, 2011; Gylling *et al*, 2009; Hitchins *et al*, 2011; Ward *et al*, 2013; Kwok *et al*, 2014; Cini *et al*, 2015), and primary that occur in the absence of any apparent linked sequence change, typically arise de *novo* (Hitchins *et al*, 2005, 2007; Valle *et al*, 2007; Morak *et al*, 2008; Niessen *et al*, 2009; Goel *et al*, 2011; Pineda *et al*, 2012) and demonstrate null (Suter *et al*, 2004; Hitchins *et al*, 2007; Morak *et al*, 2008; Pineda *et al*, 2012) or non-Mendelian inheritance (Hitchins *et al*, 2007; Ward *et al*, 2013; Sloane *et al*, 2015). To date, 75 index cases with a constitutional *MLH1* epimutation have been reported (Hitchins, 2016; Takeda *et al*, 2016), accounting for 2–3% of mutation negative cases with suspected LS whose tumours are *MLH1* deficient (Peltomäki, 2016). Most of these (66 out of 75) have been considered primary (Hitchins, 2016; Takeda *et al*, 2016). The available evidence from these cases suggests that constitutional *MLH1* epimutations cause a severe LS phenotype, including a young age of cancer onset and multiple primary tumours.

Previous studies on constitutional *MLH1* epimutation have focussed on the role of this molecular defect in cancer causation by confirming the presence of *MLH1* promoter methylation and corresponding transcriptional silencing within normal tissues, refining the selection criteria for patients warranting screening for it, and the inheritance patterns within families. It has been proposed that primary *MLH1* epimutations arise in the germline or early stages of embryonic development, as they are monoallelic and soma-wide, but frequently exhibit mosaic methylation and expression loss (Hitchins, 2015). However, no comprehensive studies have been undertaken to elucidate the mechanism(s) underlying primary *MLH1* epimutations, such as whether they are localised within the *MLH1* locus because of a focal defect, or if additional genes are concomitantly affected because of widespread epigenetic perturbation. The main aim of our study was to perform an in-depth characterisation of the methylation profile in cases with a confirmed constitutional *MLH1* epimutation to define the extent of the aberrantly methylated region around *MLH1*, as well as other loci on a genome-wide scale. The results from this study contribute to the understanding of primary constitutional *MLH1* epimutations by showing that the methylation error occurs in a localised manner.

MATERIALS AND METHODS

Patients and samples. Twelve Caucasian *MLH1* epimutation carriers were recruited from the west Mediterranean area, among whom 8 were previously reported (Pineda *et al*, 2012; Crucianelli *et al*, 2014; Castillejo *et al*, 2015). The four new cases were referred from Complejo Hospitalario de Navarra and Doce de Octubre University Hospital because of the identification of *MLH1* methylation in blood by MS-MLPA. The original numerical code for each proband was maintained from previous publications, however, a prefix corresponding to the referral center (below) was added (Supplementary Table S1). Clinicopathological data collected included age at cancer diagnosis, tumour location, immunohistochemistry of MMR proteins in tumours and *MLH1* methylation status in tumour and blood DNA (Supplementary Table S1). In all, 41 healthy controls (selected to appropriately match the patients by age, race and geographic location), 61 LS patients with confirmed genetic mutations (21 *MLH1*, 28 *MSH2*, 9 *MSH6*, 3 *PMS2*) and 4 CRC patients whose tumours demonstrated somatic *MLH1*-methylation were included as controls (Supplementary Table S2). Written informed consent was obtained from all individuals, and the ethics committee of the respective hospitals approved the study.

Samples of peripheral blood leukocyte (PBL) DNA were collected from biobanks at the Catalan Institute of Oncology (ICO), Elche University Hospital (HGUE), Valencian Biobank Network (IBSP-CV), Complejo Hospitalario de Navarra (CHN), Doce de Octubre University Hospital (H12O) and Policlinico S.Orsola-Malpighi (M). Lymphoblastoid cell lines were established from the B lymphocytes of *MLH1* epimutation carriers by standard EBV transformation using supernatant from the B95.5 cell lines, as described in Supplementary Methods. Colorectal tumours and distal normal mucosa were collected as formalin-fixed, paraffin-embedded (FFPE) tissues when available (Supplementary Table S2).

For each FFPE specimen, 10–20 $\times$ 10 μm sections were cut from a single representative block per case, using macrodissection with a scalpel if needed to enrich for tumour cells. After deparaffinisation using Deparaffinisation Solution (Qiagen, Hilden, Germany), DNA was isolated using the QIAmp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer's instructions.

Methylation analyses. The initial detection or confirmation of the presence of *MLH1* methylation in each proband and consenting family members was performed by methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) on genomic DNA from PBL. For all other methylation analyses, DNA was subject to sodium bisulphite modification using the EZ DNA Methylation Kit (Zymo Research), according to the manufacturer's instructions.

Methylation-specific multiplex ligation-dependent probe amplification. The presence of methylation at the *MLH1* promoter was detected and confirmed using 100 ng of PBL DNA using the SALSA MLPA ME011 Mismatch Repair genes probemix (MRC-Holland), according to the manufacturer's instructions. The DNA from the RKO CRC cell line was used as the *MLH1* methylation-positive control. The amplification products were run on an ABI Prism 3130 DNA sequencer and analysed using GeneMapper v4.0 (Applied Biosystems).

Clonal bisulphite sequencing. Clonal bisulphite sequencing of fragments of the *MLH1* promoter was used to determine the allelic methylation profile in bisulphite modified PBL DNA from *MLH1* epimutation carriers who were found to be heterozygous for a promoter variant. Primers were used to amplify specific promoter regions encompassing each promoter variant (Supplementary Table S3), and PCR products were cloned in *E. coli* cells using the TOPO TA cloning kit (Invitrogen) to separate the amplicons into individual alleles. The inserts from at least 24 individual colonies were sequenced using plasmid vector primers. Additional methodological details are provided in Supplementary Methods. The methylation status at individual CpG site and the allele of each heterozygous variant within each promoter fragment was determined using SeqMan software (DNASTAR) or Mutation Surveyor (SoftGenetics).

Methylation array analysis. The PBL DNA samples (1000 ng) from patients and controls, and 500–1000 ng of available FFPE colorectal cancer DNA and distal normal colon tissue DNA, were randomised within 96-well plates. For internal quality control, *in vitro* methylated and unmethylated DNAs were included in each batch.

For quality control purposes, PBL DNA integrity was evaluated by agarose gel electrophoresis and spectrophotometry. The DNAs from FFPE samples were analysed by qPCR using the Infinium FFPE QC kit (Illumina), and samples showing delta-Ct value below 5 were restored using the Infinium HD FFPE Restore kit (Illumina), following the manufacturer's instructions.

Next, 1000 ng PBL and 500 ng FFPE DNAs were bisulphite converted using the EZ DNA Methylation Kit (Zymo Research), according to the manufacturer's instructions. To determine the efficiency of the bisulphite conversion, a predetermined genomic region was evaluated by Sanger sequencing in the methylated and unmethylated controls and one sample from each batch.

Genome-wide methylation profiling was performed using the Infinium Human Methylation 450K Beadchip (Illumina) that interrogates the methylation status of 485 764 CpG sites across the genome. After hybridisation, sample scanning was performed using the HiScan platform (Illumina) that has a laser scanner with two colours (532 nm/660 nm). The relative intensity of each dye was analysed using GenomeStudio software (Methylation Module). For each analysed CpG site, a β-value was obtained depending on the florescence intensity. The β-measures took values between 0 (unmethylated) and 1 (fully methylated). Analysis for batch effects was performed using RnBeads software (Max-Planck-Institut Informatik). Group comparisons and statistical analyses (based on differentially methylated CpG sites, CpG islands, promoters, genes and tiling) were also performed using RnBeads software. The CpG methylation was visualised by Integrative Genome Viewer (Broad Institute). GRCh37/hg19 was used as the reference genome.

The data discussed in this publication are accessible through GEO Series accession number GSE107353 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107353).

Analysis of Alu elements. The sequence encompassing the *MLH1-EPM2AIP1* CpG island identified as the differentially methylated region (DMR) in carriers of an *MLH1* epimutation was screened for repetitive elements by the RepeatMasker Web Server (Institute for Systems Biology) using 'Cross_match' as the search engine. No repetitive elements were found in this DMR. Downstream of the DMR, Alu elements were found in *MLH1* intron 1, as previously described (Wang *et al*, 2011). The methylation status of these Alu elements was tested in PBL DNA from three *MLH1* epimutation carriers, one healthy control and the RKO cell line by bisulphite sequencing of the PCR fragments designated p24 (8 CpGs from Alu 2) and p25 (1 CpG from Alu 3), as previously described (Wang *et al*, 2011). Amplification of the region p10 (control region including 6 CpGs, unmethylated in normal tissues and methylated in cancer tissues) was performed using methylation-specific primers. Primers and conditions are summarised in Supplementary Table S3.

Mutational analyses

Screen of the EPM2AIP1-MLH1 CpG island for point mutations. The *EPM2AIP1-MLH1* CpG island is defined by its CpG content as the region Chr3: 37 034 229–37 035 355 by the UCSC genome browser (CpG:93). Screening for promoter variants was performed by amplification of two overlapping PCR fragments followed by direct sequencing. Primers and PCR conditions are provided in Supplementary Table S3. Sequencing was performed using the BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems). Mutation Surveyor (SoftGenetics) was used for sequence visualisation.

Detection of structural variants around the MLH1 locus. A screen for the presence of structural alterations was performed on PBL DNA from *MLH1* epimutation carriers using a custom high-definition CGH array designed with eArray Software (Agilent Technologies) and manufactured by Agilent SurePrint oligo technology. This comprised 15 000 probes encompassing the *MLH1* locus (region Chr3: 36 450 000–37 900 000 within cytoband 3p22.2), with an average probe spacing of 100 bp intervals. Bioinformatics analysis was performed in R using the 2.15.12 Bioconductor statistical packages. Results were visualised on CytoGenomics Software (Agilent Technologies).

***In silico* functional evaluation of the genetic variants identified.** The PROMO 3.0 software (Messeguer *et al*, 2002; Farré *et al*, 2003) was used to predict any changes in transcription factor binding between wild-type alleles and promoter variants. Only human transcription factors were considered and 5% was selected as maximum matrix dissimilarity rate.

Structural variants were compared with those previously reported in the Database of Genomic Variants (MacDonald *et al*, 2014), using chromosomal coordinates.

***MLH1* allelic expression analyses.** For allelic expression analyses at the *MLH1* single-nucleotide polymorphism (SNP) rs1799977 (c.655A > G), the relative levels of the two alleles were determined in genomic DNA and cDNA by single-nucleotide primer extension (SNuPE) as previously described (Pineda *et al*, 2012) (see Supplementary Methods for details).

Haplotype analysis. Haplotype analysis was performed using four intragenic *MLH1* SNPs rs1800734, rs9876116, rs1799977 and rs4234259, and seven microsatellite markers, D3S1609, D3S1612, D3S2369, D3S1611, D3S3623, D3S1298 and D3S3564, spanning 12 Mb around *MLH1*, as previously described (Borràs *et al*, 2010). To deduce the methylation-associated haplotype, intrafamilial segregation analysis was performed under the assumption that the number of crossovers between adjacent markers was minimal.

RESULTS

Validation of newly identified *MLH1* epimutations and associated functional impact. The *MLH1* promoter methylation was confirmed by MS-MLPA in blood from newly identified epimutation carriers (Table 1). Three of them presented methylation values higher than 40%, whereas case CHN_2 displayed evidence of epigenetic mosaicism.

The frequent SNP rs1799977 (c.655G > A) within *MLH1* exon 8 was used to determine the effect of the epimutation on the allelic transcriptional activity of *MLH1* in five carriers who were heterozygous for this SNP. Monoallelic expression of the G allele was identified in lymphocytes from case CHN_1, indicating complete loss of expression from the epimutant A allele. In cases CHN_2 and H12O-A, partial transcriptional silencing was observed in lymphoblastoid cell lines derived from patient lymphocytes, consistent with the observed methylation mosaicism (Table 1). CHN_2 had reduced expression of the A allele with allele-specific expression (ASE) 0.53, and H120-A had reduced expression of the G allele, with ASE 0.79 (Table 1). Loss of expression of the epimutant A allele was previously demonstrated in cases ICO_1 and M60 (Pineda *et al*, 2012; Crucianelli *et al*, 2014). The remaining cases were homozygotes for the polymorphism, and thus allelic expression analyses could not be performed in them.

Genetic characterisation of the *MLH1* locus. As secondary constitutional *MLH1* epimutations are linked to rare variants or rearrangements involving the *EPM2AIP1-MLH1* promoter, mutational analysis of the whole CpG island encompassing the

Table 1. Results of molecular characterisation of constitutional *MLH1* epimutation carriers in our series

Case ID at current study	% *MLH1* promoter methylation in blood (MS-MLPA)		Germline variants		Functional impact of epimutation			Inheritance pattern				Other features
	C Deng's region	D Deng's region	Point rare variants at DM CpG island	Structural alterations at 3p22.2 cytoband	Variant associated with loss of transcription	Sample tested	Expression of the analysed variant ±s.d.	*MLH1* promoter variants	Methylation-associated allele determination	% Of methylated clones	Co-segregation of MAA in	
ICO_1	56	56	None	Not identified	*MLH1* c.655A and EPM2AIP1 c.*2570G	Blood lymphocytes	*MLH1* = 0 and *EPM2AIP1* = 0.02	c.-93G>A	*MLH1* c.-93A	100% of c.-93A clones	1 Daughter (unmethylated)	Intergenerational erasure
ICO_34	25	28	None	Not identified	*EPM2AIP1* c.*2570G	Blood lymphocytes	0.48	**c.-593G>C**	***MLH1*c.-593G**	**38% Of c.-593G clones (3 of 8 clones)**	**2 Sisters (both unmethylated)**	
M60	56	57	None	Not identified	*MLH1* c.655A	Blood lymphocytes	0.3	WT	**No promoter variants**	NA	NA	
HGUE_1	48	52	None	**One deletion (20 Kb) outside DM CpG island**	NA	NA	NA	WT	**No promoter variants**	NA	NA	
HGUE_2	64	68	None	**Two deletions (15 and 19Kb) outside DM CpG island**	NA	NA	NA	**c.-93G>A**	***MLH1* c.-93G**	**94% Of c.-93G clones (16 of 17 clones)**	**No transmission of MAA**	
HGUE_3	70	82	None	Not identified	NA	NA	NA	WT	**No promoter variants**	NA	NA	
HGUE_4	49	54	None	Not identified	NA	NA	NA	WT	**No promoter variants**	NA	NA	
HGUE_5	49	50	**c.-234_−236 delCTT in*trans* with methylation**	Not identified	NA	NA	NA	**c.-234_-236 delCTT and c.-93G>A**	***MLH1*c.-93A (allele without deletion)**	**80% Of c.-93A clones (8 of 10 clones)**	NA	
CHN_1	47	48	None	Not identified	***MLH1*c.655A**	**Blood lymphocytes**	**0±0.00**	**c.-93G>A**	***MLH1* c.-93A**	**100% Of c.-93A clones (9 clones)**	**1 Sister and proband's son and daughter (all of them unmethylated)**	**Intergenerational erasure**
CHN_2	36	28	None	Not identified	***MLH1*c.655A**	**EBV cell line (44% methylation)**	**0.53±0.04**	**WT**	**No promoter variants**	NA	NA	
H12O_A	40	41	None	Not identified	***MLH1*c.655G**	**EBV cell line(9% methylation)**	**0,79±0.03**	**c.-93G>A**	***MLH1* c.-93G**	**100% Of c.-93G clones (13 clones)**	NA	
H12O_B	47	43	None	Not identified	NA	NA	NA	**WT**	**No promoter variants**	NA	NA	

Abbreviations: DM = differentially methylated; MAA = methylation-associated allele; *MLH1* = *MutL homologue 1*; MS-MLPA = methylation-specific multiplex ligation-dependent probe amplification; NA = not available; WT = wild type. Novel previously unpublished results are highlighted in bold.

EPM2AIP1-MLH1 promoter was performed. The common SNP rs1800734 (c.-93G>A) was identified in 5 of the 12 patients (42%) and rs34566456 (c.-593G>C) in case ICO_34. Additionally, case HGUE_5 was heterozygous for a novel variant c.-234_-236delCTT (Supplementary Figure S1). *In silico*, this is located within the binding region for the TFII-I and GATA-1 transcription factors, but these were predicted to be unaffected by the variant. No other variants – including c.[-27C>A; c.85G>T] (Hitchins *et al*, 2011) – were identified within the *MLH1* promoter region.

Screening for structural variants surrounding the *MLH1* locus (Chr 3: 36 450 000–37 900 000) identified 3 small deletions in 2 of the 11 patients analysed, for whom sufficient DNA was available (Supplementary Figure S2 and Table 1). Two deletions outside the differentially methylated CpG island were identified in case HGUE_2: one 15 Kb in size located upstream of the *MLH1* gene (Chr3: 36 798 479–36 813 411) and the other of 19 Kb in size located downstream (Chr3: 37 486 324–37 505 162) (Supplementary Figure S2A). Neither of these two deletions has previously been described as copy number variations in the DGV database (Supplementary Table S4). In case HGUE_1, a 20 Kb deletion was identified upstream of the *MLH1* gene (Chr3: 36 396 587–36 416 879) (Supplementary Figure S4B). This region appears to be prone to copy number variations in the control reference group (DGV database), as shown in Supplementary Table S4. In both cases (HGUE_1 and HGUE_2), validation of these deletions using a second method was not possible because of sample depletion.

Determination of the MAA. Clonal bisulphite sequencing was performed on PBL DNA of probands harbouring a heterozygous variant within the *MLH1* promoter region to determine the methylation-associated allele (MAA). Dense monoallelic *MLH1* methylation was linked to the G allele at SNP rs1800734 (c.-93G>A) in cases H12O-A and HGUE_2, and to the allele A in patients CHN_1 and HGUE_5 (Table 1). In case HGUE_5, who was heterozygous for both the novel heterozygous variant c.-234_-236del and the c.-93G>A SNP, methylation was linked to the non-deleted and c.-93A alleles, and the [c.-234-236del; c.-93G] haplotype was unmethylated (Supplementary Figure S3). This finding indicated the c.-234_-236del variant was not causally linked to the epimutation in HGUE_5. In patient ICO_34, methylation was linked to the G allele at SNP rs34566456 (c.-593G>C) with ∼38% of G alleles methylated, indicating mosaicism (Table 1), consistent our prior findings using other methods (Pineda *et al*, 2012).

Inheritance pattern of the epimutant alleles and designation of *MLH1* epimutation as likely primary or secondary. To investigate the inheritance pattern of the genetic allele harbouring the

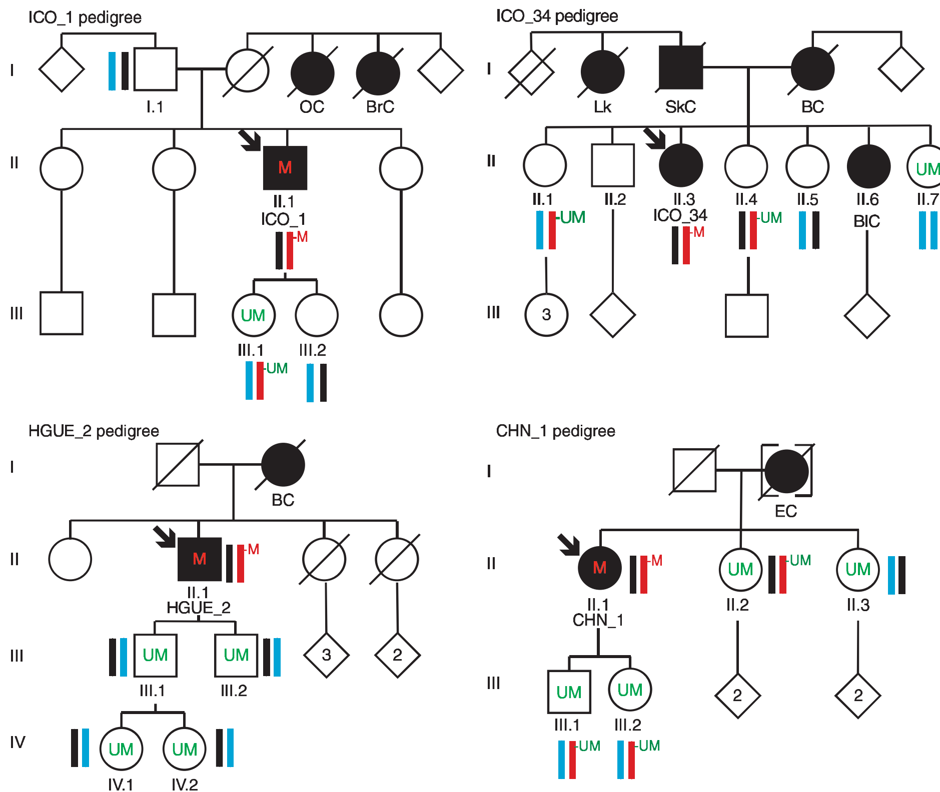


Figure 1. **Pedigrees of the *MLH1* epimutation cases showing that the presence of the epimutation does not co-segregate with any genetic allele.** Circles, females; squares, males; filled, cancer affected. Cancer localisation (BC = breast cancer; BrC = brain cancer; CRC = colorectal cancer; EC = endometrial cancer; Lk = leukaemia; OC = ovarian cancer; PC = prostate cancer) and age at diagnosis are indicated. Epimutation carriers are indicated by an arrow. Generations are indicated on the left margin in Roman numerals and analysed relatives are identified by numbers. Haplotypes are schematised by sticks: Red = methylation-associated allele (MAA), black = haplotype also carried by the proband but not associated with the epimutation, blue = uninformative haplotypes. The presence of methylation (M) or its absence (UM) is indicated on the red MAA haplotypes for those individuals tested.

epimutation, the *MLH1* promoter methylation status and haplotypes within a 12 Mb around *MLH1* were determined in available PBL DNA from the first-degree relatives of epimutation carriers (Supplementary Table S5). The MS-MLPA analysis showed no evidence of *MLH1* methylation in any of the relatives tested from any family, and thus no case of transmission of an epimutation was observed among the family members tested (Figures 1 and 2). Haplotype analyses revealed that the genetic allele bearing the epimutation (in the proband) was shared in unmethylated form by two sisters of case ICO_34, suggesting a *de novo* occurrence of the epimutation in proband ICO_34; one sister and two offspring of case CHN_1, suggesting *de novo* occurrence of the epimutation in proband CHN_1 and indicating intergenerational erasure of the epimutation in her offspring (Figure 1). The intergenerational erasure newly observed in case CHN_1 herein, and previously observed in case ICO_1 (Pineda *et al*, 2012), allowed us to definitively classify the epimutation in these two cases as primary. In six additional cases, the absence of methylation in the relatives tested, and/or the lack of a family history of cancer, also suggested their epimutation was of the primary type.

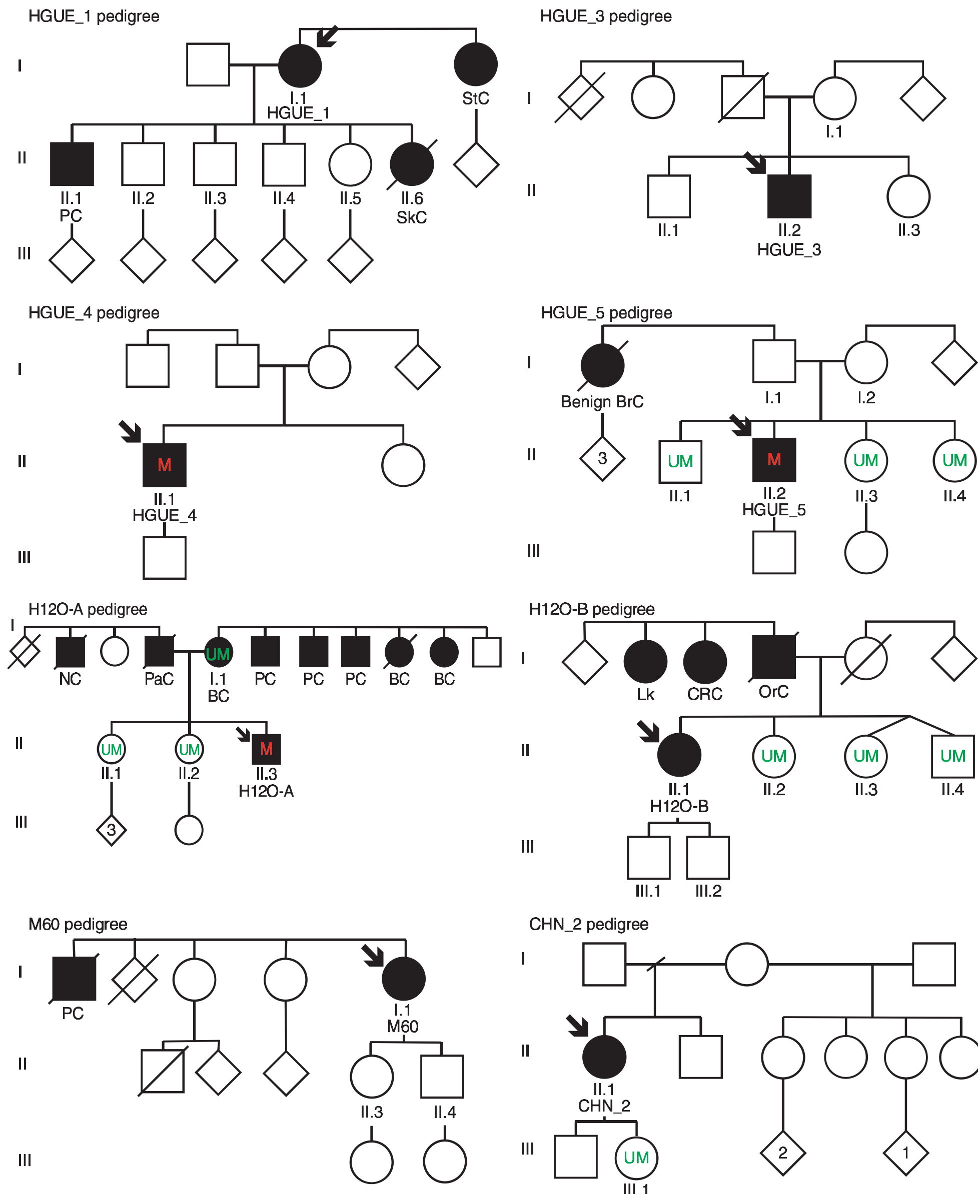
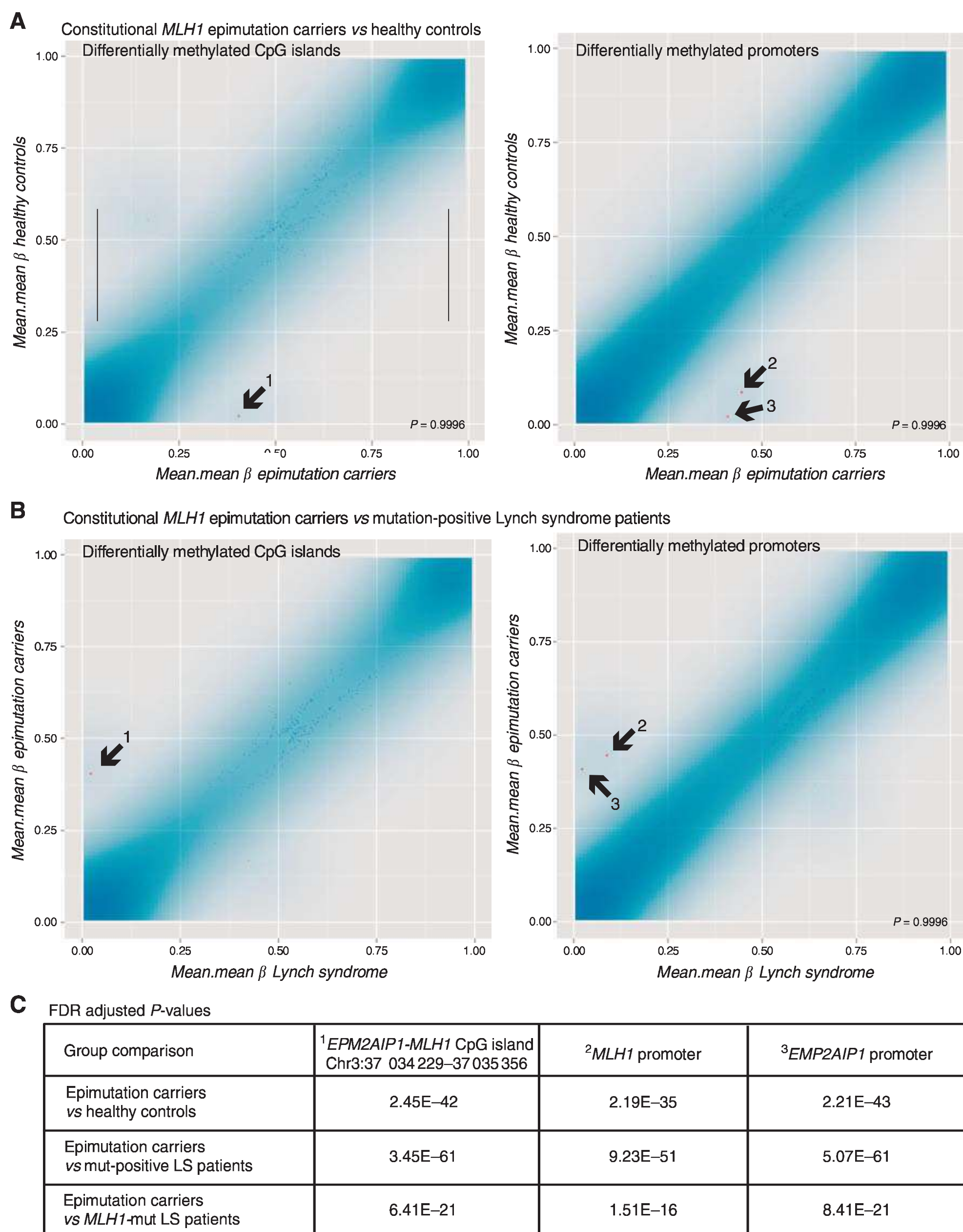


Figure 2. Pedigrees of the *MLH1* epimutation carriers not included in Figure 1. Circles, females; squares, males; filled, cancer affected. Cancer localisation (BC = breast cancer; BrC = brain cancer; CRC = colorectal cancer; GC = gastric cancer; Lk = leukaemia; NC = neck cancer; OrC = oral cancer; PaC = pancreatic cancer; PC = prostate cancer; SkC = skin cancer) and age at diagnosis are indicated. Epimutation carriers are indicated by an arrow. Generations are indicated on the left margin in Roman numerals and analysed relatives are identified by numbers. The presence of methylation (M) or its absence (UM) is indicated if tested.

 www.bjcancer.com | DOI:10.1038/bjc.2017.466

In patient HGUE_2, although clonal bisulphite sequencing showed monoallelic methylation of the c.-93A allele, we were unable to determine whether this was linked in *cis* to either or both of the 15 Kb and 19 Kb deletions flanking *MLH1* that were also identified in this patient. The MAA was not transmitted to either of his offspring, hence this family was non-informative for whether the type of epimutation carried by this proband was primary or secondary. Similarly, in case HGUE_1, we were unable to determine whether the epimutation was linked in *cis* to the 20 Kb deletion identified upstream of *MLH1*, and hence whether the deletion may have been causally related.

Global epigenetic characterisation of *MLH1* constitutional epimutations. In order to evaluate whether aberrant methylation



Group comparison	[1]*EPM2AIP1-MLH1* CpG island Chr3:37 034 229–37 035 356	[2]*MLH1* promoter	[3]*EMP2AIP1* promoter
Epimutation carriers *vs* healthy controls	2.45E–42	2.19E–35	2.21E–43
Epimutation carriers *vs* mut-positive LS patients	3.45E–61	9.23E–51	5.07E–61
Epimutation carriers *vs MLH1*-mut LS patients	6.41E–21	1.51E–16	8.41E–21

Figure 3. Scatterplot of the normalised mean *β*-values obtained using the Infinium 450K Human Methylation array to identify differentially methylated CpG islands and promoters in blood DNA from constitutional *MLH1* epimutation carriers against healthy controls (**A**) and Lynch syndrome patients with a germline sequence mutation. (**B**) The transparency corresponds to point density. The 1% of the points in the sparsest populated plot regions are drawn explicitly. The coloured points represent differentially methylated CpG islands (depicted as arrow 1) and promoters (arrows 2 and 3). The FDR adjusted *P*-values of differentially methylated regions are summarised at the bottom (**C**).

was restricted to *MLH1* or extended to other loci across the genome in *MLH1* epimutation carriers, methylation array analysis was performed using the Infinium 450K array in PBL DNA. This revealed that the *MLH1-EPM2AIP1* promoter CpG island was the sole DMR in *MLH1* epimutation carriers when compared with healthy controls (false discovery rate (FDR) adjusted *P*-value = 2.45E − 42) (Figure 3A). The same result was obtained when *MLH1* epimutation carriers were compared with LS cases (FDR adjusted *P*-value = 3.45E − 61) (Figure 3B), and the subset of Lynch syndrome cases with germline *MLH1* mutation (FDR adjusted *P*-value = 6.41E − 21) (Figure 3C). Accordingly, the same analysis identified *MLH1* and *EPM2AIP1* as the unique differentially methylated promoters in *MLH1* epimutation carriers compared with healthy controls (*EPM2AIP1* FDR adjusted *P*-value = 2.21E − 43, *MLH1* FDR adjusted *P*-value = 2.19E − 35), LS patients (*EPM2AIP1* FDR adjusted *P*-value = 5.07E − 61, *MLH1* FDR adjusted *P*-value = 9.23E − 51) and *MLH1*-mutated LS cases (*EPM2AIP1* FDR adjusted *P*-value = 8.41E − 21, *MLH1* FDR adjusted *P*-value = 1.51E − 16) (Figure 3).

A deeper analysis at the *MLH1* promoter region showed that the methylation levels in PBL from constitutional epimutation carriers spread across the entire CpG island, *MLH1* exon 1 and the first part of intron 1 (Figure 4). The CpG island shores were also affected by hypermethylation, represented by probes at positions chr3: 37 033 791 and chr3: 37 035 400 at the boundaries of the 1.6 Kb DMR. Of note, no other common DMR were found in the genome-wide methylation analysis among the carriers of a *MLH1* constitutional epimutation.

The DMR identified in the PBL of *MLH1* epimutation carriers was also present in subject-matched normal colonic mucosa and CRC (where available for analysis). Although methylation levels across the DMR were marginally higher in CRC tissue than normal tissues, this did not reach statistical significance, suggesting that methylation is not a common second hit in these tumours (Figure 5A). Similarly, no significant differences were found in the levels of methylation among the tissues tested from mutation-positive LS cases (Figure 5B).

In normal colonic mucosa tissue, a lack of *MLH1* methylation was similarly observed in the LS control group and the sporadic cases with somatic *MLH1* methylation in their tumours (Supplementary Figure S3A). In contrast, in tumour tissue, the sporadic *MLH1*-methylated CRC cases showed a more extensive DMR spanning ≈50 Kb (from chr3: 36 985 516 to chr3: 37 035 399), within which four CpG sites located at chr3: 36 986 513; 36 986 532; 36 986 555 and 36 986 642 were differentially methylated only in the sporadic CRC (FDR adjusted *P*-value < 0.01) (Supplementary Figure S3B and C).

Methylation state of Alu elements. In order to evaluate whether demethylation of *MLH1* intron 1 Alu sequences is associated with constitutional epimutation, as has been suggested for sporadic MSI tumours (Wang *et al*, 2011), bisulphite sequencing analysis was performed in PBL DNA from 3 *MLH1* epimutation carriers (cases ICO_1, ICO_34 and CHN_1), one healthy control and the RKO cell line. As previously described, partial demethylation was observed in Alu 3 of RKO cells (Supplementary Figure S4). In contrast, the intron 1 Alu CpGs analysed were found to be fully methylated in carriers of a constitutional *MLH1* epimutation and a healthy control (Supplementary Figure S4), arguing against demethylation of these nearby Alu sequences as a causal mechanism for constitutional epimutation.

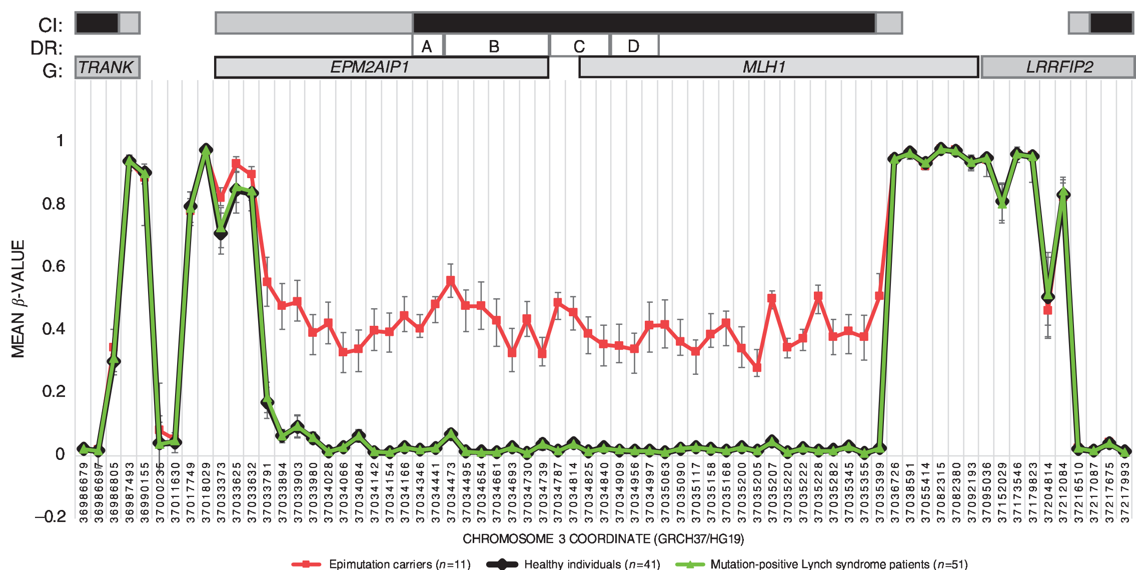


Figure 4. Representation of the differentially methylated region across the *MLH1* locus in blood DNA from constitutional *MLH1* epimutation carriers (red, *n* = 11), Lynch syndrome mutation carriers (green, *n* = 61) and healthy controls (black, *n* = 41). The β-values obtained from Infinium 450K Human Methylation array analysis are displayed against the genomic coordinate for each CpG site interrogated. The relative locations of the CpG sites are not drawn to scale. The CpG sites located between Chr3: 37 033 791 and 37 035 399 coordinates had FDR adjusted *P*-values between 2.5E − 7 and 1.2E − 59 for comparisons of *MLH1* epimutation carriers against Lynch syndrome patients or healthy controls. Above, 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 (A–D). Genes (G) containing the displayed CpG sites are represented as grey rectangles. Ensembl GRCh37 was taken as reference for gene coordinates.

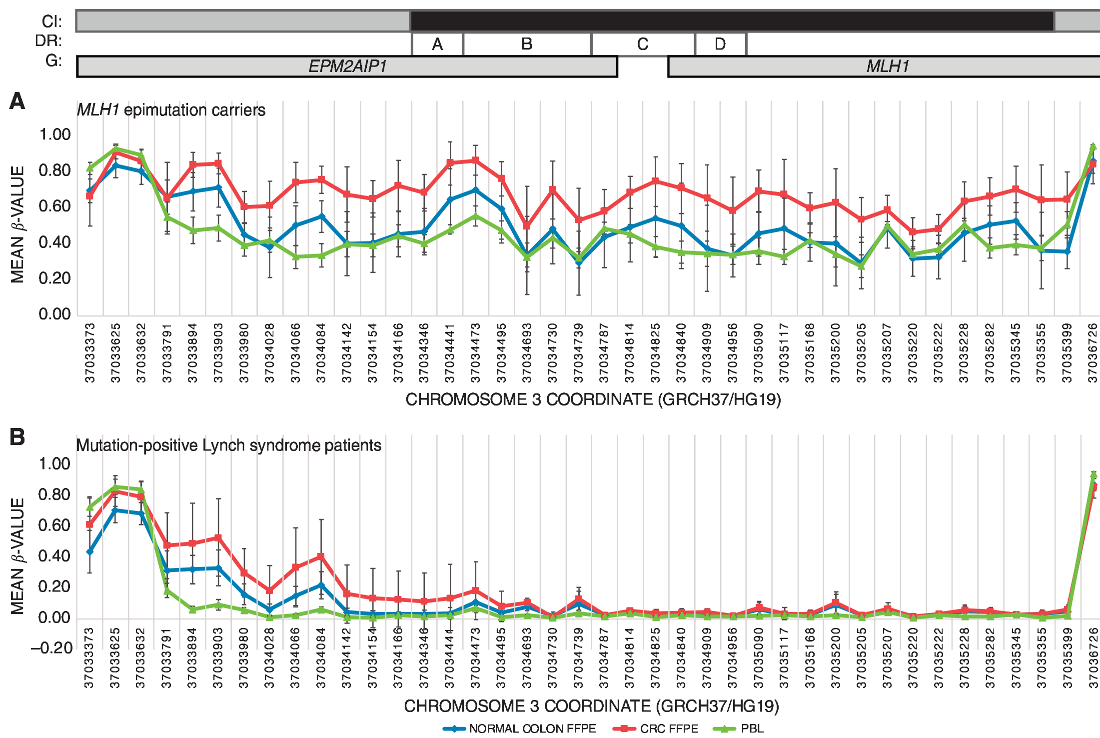


Figure 5. Representation of differentially methylated region in blood, normal colonic mucosa and colorectal cancer tissue in constitutional *MLH1* epimutation patients (**A**) and mutation-positive Lynch syndrome patients. (**B**) The β-values obtained after global methylation analysis are displayed against CpG site coordinate location. The location of the CpG sites is not drawn to scale. Above, *EMP2AIP1-MLH1* CpG island (CI) is represented as black rectangle and their shores are represented in grey. Location of Deng's promoter regions (DR) are indicated as white rectangles (A–D). Genes (G) including displayed CpG sites are represented as grey rectangles. Ensembl GRCh37 was taken as reference for gene coordinates.

DISCUSSION

A comprehensive genetic and epigenetic analysis has provided a deeper and broader molecular characterisation of 12 cases with constitutional *MLH1* epimutation than previously undertaken. The series, which includes 4 novel cases, accounts for 16% of all index cases of constitutional *MLH1* epimutation reported to date (Supplementary Table S6). Our findings suggest that the epimutations in at least 10 of these cases are primary based on the following: (1) the absence of rare in *cis* genetic aberrations within the *MLH1* CpG island, (2) the lack of familial cancer history and the presence of the MAA in an unmethylated state in siblings and/or parents that point to a *de novo* occurrence of the epimutation in the probands (Supplementary Tables S1 and S5); and (3) the demonstration in two cases of the intergenerational erasure of the epimutation in offspring who inherited the MAA. Interestingly, few studies of *MLH1* epimutation that have involved family members from more than one generation have ever been undertaken, but among these, both intergenerational inheritance and erasure of the epimutation between the proband and offspring have been described (Supplementary Table S5). The two cases of intergenerational erasure we describe herein provide further evidence of the reversibility of this defect between generations (Supplementary Table S5 and Figure 1).

In two cases, HGUE_1 and HGUE_2, deletions external to *MLH1* were identified by aCGH. However, we were unable to confirm the presence of these deletions by other methods because of the lack of sample availability. Also, we could not determine if the deletions occurred on the MAA. Noteworthy, secondary *MLH1* epimutations have been reported with concomitant genetic alterations in *cis*, including promoter deletions, a promoter single-nucleotide variant and a large duplication encompassing *MLH1* and neighbouring genes (Renkonen *et al*, 2003; Morak *et al*, 2008; Gylling *et al*, 2009; Morak *et al*, 2011; Hitchins*et al*, 2011; Crépin *et al*, 2012; Ward *et al*, 2013; Kwok *et al*, 2014; Cini *et al*, 2015). These cases have displayed autosomal dominant inheritance of the *MLH1* epimutation with faithful segregation of the methylation with the genetic alteration. Therefore, although we cannot completely rule out that deletions are causative of the epimutation in cases HGUE_1 and HGUE_2, the absence of family history and methylation in blood from first-degree relatives led us to propose to classify them as 'suspected primary' (Supplementary Table S6).

Transcriptional silencing usually correlates with methylation levels (Supplementary Table S6). In case H12O-A, differences in the values of methylation and transcriptional silencing could be found depending upon the sample tested. The lymphoblastoid cell line displayed less methylation than that observed in blood, in line with a previous observation (Kwok *et al*, 2010). These results highlight the limitations of using lymphoblastoid cell lines for epigenetic-based analyses, although this may be a necessity when primary tissue samples from patients are limited.

The array-based genome-wide methylation analysis performed in PBL DNA demonstrated that a region of 1.6 Kb (Chr3: 37 033 791–37 035 400) – encompassing the shared *MLH1-EPM2AIP1* CpG island and its corresponding shores – was the sole DMR unique to the group of *MLH1* epimutation carriers, as compared with mutation-positive LS patients or healthy controls. For cases HGUE_1 and HGUE_2, classified as suspected primary epimutations, the same 1.6 Kb DMR was identified. No differences in the length of the DMR were observed in the normal colon mucosa or CRC tissue among the *MLH1* epimutation carriers.

The absence of any other DMR in the *MLH1* epimutation carrier group suggests that the constitutional epimutation is caused by a shared focal event, as it spans the very same region in all patients tested. Of note, our preliminary analysis does not suggest a genetic cause. Also, we have provided evidence against *MLH1* intron 1 Alu sequence demethylation as a causal mechanism for constitutional *MLH1* epimutation.

Interestingly, the extent of the DMR observed (consistently in all three tissue types analysed) in *MLH1* epimutation carriers was narrower than that observed in sporadic *MLH1*-methylated CRC. Although our cases of sporadic *MLH1*-methylated CRC were few ($n = 4$), the extent of their methylation (50 Kb) is consistent with tumours with available MSI and methylation data from TCGA project ($n = 43$) (Supplementary Figure S5).

Although the causal mechanism of primary constitutional *MLH1* epimutations remains unknown, it may be speculated that it acts in the very early stages of embryonic development (Hitchins, 2015). The *MLH1 de novo* methylation has been induced *in vitro* in human embryonic stem cells, suggesting that pluripotency is a mandatory requirement for establishing the epimutation (Takahashi *et al*, 2017). The detection of non-methylated clones on the MAA in distinct patients, including those with methylation levels close to 50% (Supplementary Table S5), suggests that if an epimutation is established in the germline, some alleles may lose the methylation during the demethylation phase that occurs between zygote formation and the preimplantation embryogenesis (Hitchins, 2015). The methylated allele could behave as a 'leaky' allele when self-propagating the *MLH1* epimutation within individual cells.

Our results show that aberrant methylation of the *MLH1* promoter is targeted, as opposed to disseminated, as we found no other common DMRs among the *MLH1* epimutation carriers. This suggests that a primary epimutation is not derived from a *trans*-event or a cellular perturbation during germline or embryonic development. Further studies are needed to fully understand the mechanistic basis of *MLH1* constitutional epimutations and their heritability/reversibility. Ongoing mechanistic studies should be focussed on specific interactions of epigenetic effectors with the *MLH1* locus, paying special attention to those acting in the germline or during early embryonic development.

ACKNOWLEDGEMENTS

We are indebted to the patients and their families. We thank all members of the involved Genetic Counseling Cancer Units and Valencian Biobank Network. We thank Lauro Sumoy and Gabriel Rech for their support with methylation array data analysis, and Miguel Àngel Peinado and Mireia Jordà for a critical revision of the manuscript. 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 and SAF2015-68016-R), Instituto de Salud Carlos III (Grant PI16/00563), CIBERONC, RTICC Network (RD12/0036/0031 and RD12/0036/0008), the Spanish Association Against Cancer (080253), the Government of Catalonia (Grant 2014SGR338) and Fundación Mutua Madrileña (Grant AP114252013). This article is based upon work from COST Action BM1206, supported by COST (European Cooperation in Science and Technology, www.cost.eu). ED was supported by a grant from the Spanish Ministry of Economy and Competitiveness.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Borràs E, Pineda M, Blanco I, Jewett EM, Wang F, Teulé À, Caldés T, Urioste M, Martínez-Bouzas C, Brunet J, Balmaña J, Torres A, Ramón y Cajal T, Sanz J, Pérez-Cabornero L, Castellví-Bel S, Alonso Á, Lanas Á, González S, Moreno V, Gruber SB, Rosenberg NA, Mukherjee B, Lázaro C, Capellá G (2010) *MLH1* founder mutations with moderate penetrance in Spanish Lynch syndrome families. *Cancer Res* **70**(19): 7379–7391.

Castillejo A, Hernández-Illán E, Rodriguez-Soler M, Pérez-Carbonell L, Egoavil C, Barberá VM, Castillejo M-I, Guarinos C, Martínez-de-Dueñas E, Juan M-J, Sánchez-Heras A-B, García-Casado Z, Ruiz-Ponte C, Brea-Fernández A, Juárez M, Bujanda L, Clofent J, Llor X, Andreu M, Castells A, Carracedo A, Alenda C, Payá A, Jover R, Soto J-L (2015) Prevalence of MLH1 constitutional epimutations as a cause of Lynch syndrome in unselected versus selected consecutive series of patients with colorectal cancer. *J Med Genet* **52**(7): 498–502.

Cini G, Carnevali I, Quaia M, Chiaravalli AM, Sala P, Giacomini E, Maestro R, Tibiletti MG, Viel A (2015) Concomitant mutation and epimutation of the MLH1 gene in a Lynch syndrome family. *Carcinogenesis* **36**(4): 452–458.

Crépin M, Dieu M-C, Lejeune S, Escande F, Boidin D, Porchet N, Morin G, Manouvrier S, Mathieu M, Buisine M-P (2012) Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. *Hum Mutat* **33**(1): 180–188.

Crucianelli F, Tricarico R, Turchetti D, Gorelli G, Gensini F, Sestini R, Giunti L, Pedroni M, Ponz de Leon M, Civitelli S, Genuardi M (2014) MLH1 constitutional and somatic methylation in patients with MLH1 negative tumors fulfilling the revised Bethesda criteria. *Epigenetics* **9**(10): 1431–1438.

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 Res* **31**(13): 3651–3653.

Goel A, Nguyen T-P, Leung H-CE, Nagasaka T, Rhees J, Hotchkiss E, Arnold M, Banerji P, Koi M, Kwok C-T, Packham D, Lipton L, Boland CR, Ward RL, Hitchins MP (2011) De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* **128**(4): 869–878.

Gylling A, Ridanpää M, Vierimaa O, Aittomäki K, Avela K, Kääriäinen H, Laivuori H, Pöyhönen M, Sallinen S-L, Wallgren-Pettersson C, Järvinen HJ, Mecklin J-P, Peltomäki P (2009) Large genomic rearrangements and germline epimutations in Lynch syndrome. *Int J Cancer* **124**(10): 2333–2340.

Hitchins M, Williams R, Cheong K, Halani N, Lin VAP, Packham D, Ku S, Buckle A, Hawkins N, Burn J, Gallinger S, Goldblatt J, Kirk J, Tomlinson I, Scott R, Spigelman A, Suter C, Martin D, Suthers G, Ward R (2005) MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* **129**(5): 1392–1399.

Hitchins MP (2015) Constitutional epimutation as a mechanism for cancer causality and heritability? *Nat Rev Cancer* **15**(10): 625–634.

Hitchins MP (2016) Finding the needle in a haystack: identification of cases of Lynch syndrome with MLH1 epimutation. *Fam Cancer* **15**(3): 413–422.

Hitchins MP, Rapkins Robert W, Kwok C-T, Srivastava S, Wong Justin JL, Khachigian Levon M, Polly P, Goldblatt J, Ward Robyn L (2011) Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5UTR. *Cancer Cell* **20**(2): 200–213.

Hitchins MP, Wong JJL, Suthers G, Suter CM, Martin DIK, Hawkins NJ, Ward RL (2007) Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med* **356**(7): 697–705.

Kwok C-T, Vogelaar IP, van Zelst-Stams WA, Mensenkamp AR, Ligtenberg MJ, Rapkins RW, Ward RL, Chun N, Ford JM, Ladabaum U, McKinnon WC, Greenblatt MS, Hitchins MP (2014) The MLH1 c.-27C>A and c.85G>T variants are linked to dominantly inherited MLH1 epimutation and are borne on a European ancestral haplotype. *Eur J Hum Genet* **22**(5): 617–624.

Kwok C-T, Ward RL, Hawkins NJ, Hitchins MP (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.

Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP (2015) Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* **15**(3): 181–194.

MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42**(D1): D986–D992.

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.

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. *J Med Genet* **48**(8): 513–519.

Morak M, Schackert HK, 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.

Niessen RC, Hofstra RMW, Westers H, Ligtenberg MJL, Kooi K, Jager POJ, de Groote ML, Dijkhuizen T, Olderode-Berends MJW, Hollema H, Kleibeuker JH, Sijmons RH (2009) Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes Chromosomes Cancer* **48**(8): 737–744.

Peltomäki P (2016) Update on Lynch syndrome genomics. *Fam Cancer* **15**(3): 385–393.

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.

Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, Aittomäki K, Järvinen HJ, Mecklin J-P, Lindblom A, Peltomäki P (2003) Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. *J Clin Oncol* **21**(19): 3629–3637.

Sloane MA, Nunez AC, Packham D *et al.* (2015) Mosaic epigenetic inheritance as a cause of early-onset colorectal cancer. *JAMA Oncol* **1**(7): 953–957.

Suter CM, Martin DIK, Ward RL (2004) Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* **36**(5): 497–501.

Takahashi Y, Wu J, Suzuki K, Martinez-Redondo P, Li M, Liao H-K, Wu M-Z, Hernández-Benítez R, Hishida T, Shokhirev MN, Esteban CR, Sancho-Martinez I, Belmonte JCI (2017) Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells. *Science* **356**(6337): 503–508.

Takeda T, Banno K, Yanokura M, Adachi M, Iijima M, Kunitomi H, Nakamura K, Iida M, Nogami Y, Umene K, Masuda K, Kobayashi Y, Yamagami W, Hirasawa A, Tominaga E, Susumu N, Aoki D (2016) Methylation analysis of DNA mismatch repair genes using DNA derived from the peripheral blood of patients with endometrial cancer: epimutation in endometrial carcinogenesis. *Genes* **7**(10): 86.

Valle L, Carbonell P, Fernandez V, Dotor AM, Sanz M, Benitez J, Urioste M (2007) MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. *Clin Genet* **71**(3): 232–237.

Wang X, Fan J, Liu D, Fu S, Ingvarsson S, Chen H (2011) Spreading of Alu methylation to the promoter of the MLH1 gene in gastrointestinal cancer. *PLOS ONE* **6**(10): e25913.

Ward RL, Dobbins T, Lindor NM, Rapkins RW, Hitchins MP (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.

This work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/

© The Author(s) named above 2017

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)

SUPPLEMENTARY MATERIAL

Supplementary figures

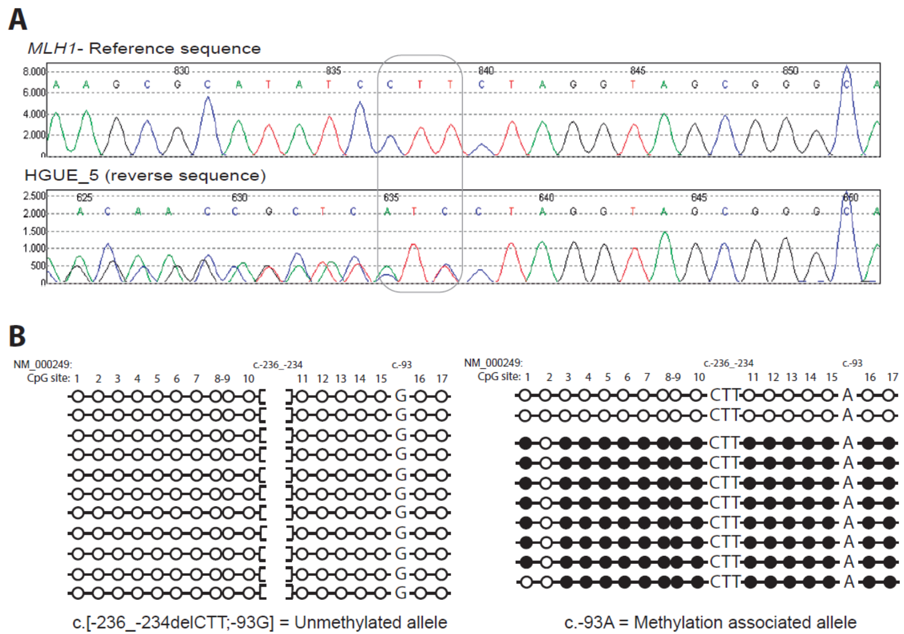


Figure S1 - Characterization of a novel promoter variant found at HGUE_5 case. A) Result of the sequencing analysis of the promoter variant c.-236_-234delCTT in HGUE_5 case. CTT deletion is indicated by a rectangle. B) Clonal bisulfite sequencing of *MLH1* promoter (C Deng's Region) in blood DNA from HGUE_5 patient. Each horizontal line represents a single allele. CpG dinucleotides are depicted by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. The allele at rs1800734 (c.-93G>A) is indicated as A or G. Novel variant c.-234_-236delCTT is indicated as "CTT" for wildtype haplotype and "[ ]" for the deletioned one. Each CpG analyzed is numbered according to its position at amplified region.

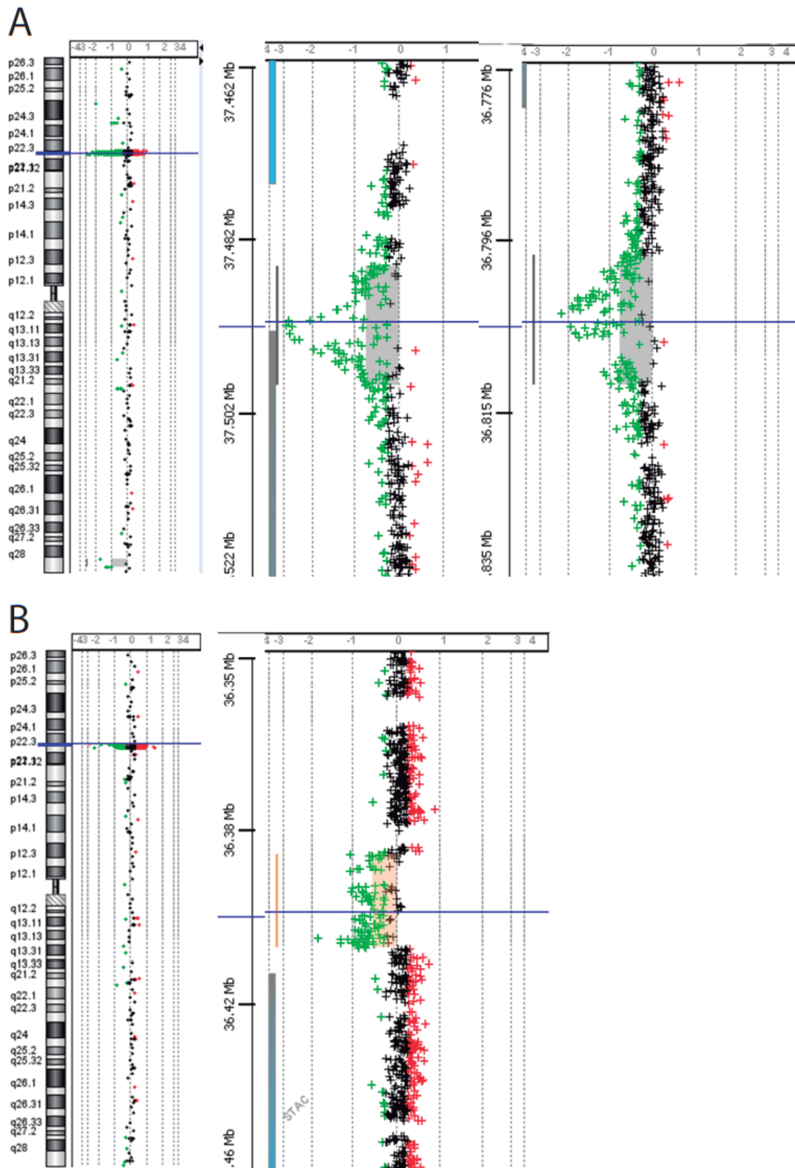


Figure S2 - Structural alterations found in constitutional epimutation carriers by CGH array. Analyzed region is pointed on chromosome 3 representation. Probes are represented as crosses. Black crosses = equal amount of patient sample and reference sample has been hibridizated for a single probe. Red crosses = More patient sample has been hibridizated for a single probe (gain). Green crosses = More reference sample has been hibridizated for a single probe (loss). Shadow squares behind deleted probes indicate deletion sizes for patient HGUE_2 (A) and HGUE_1 (B).

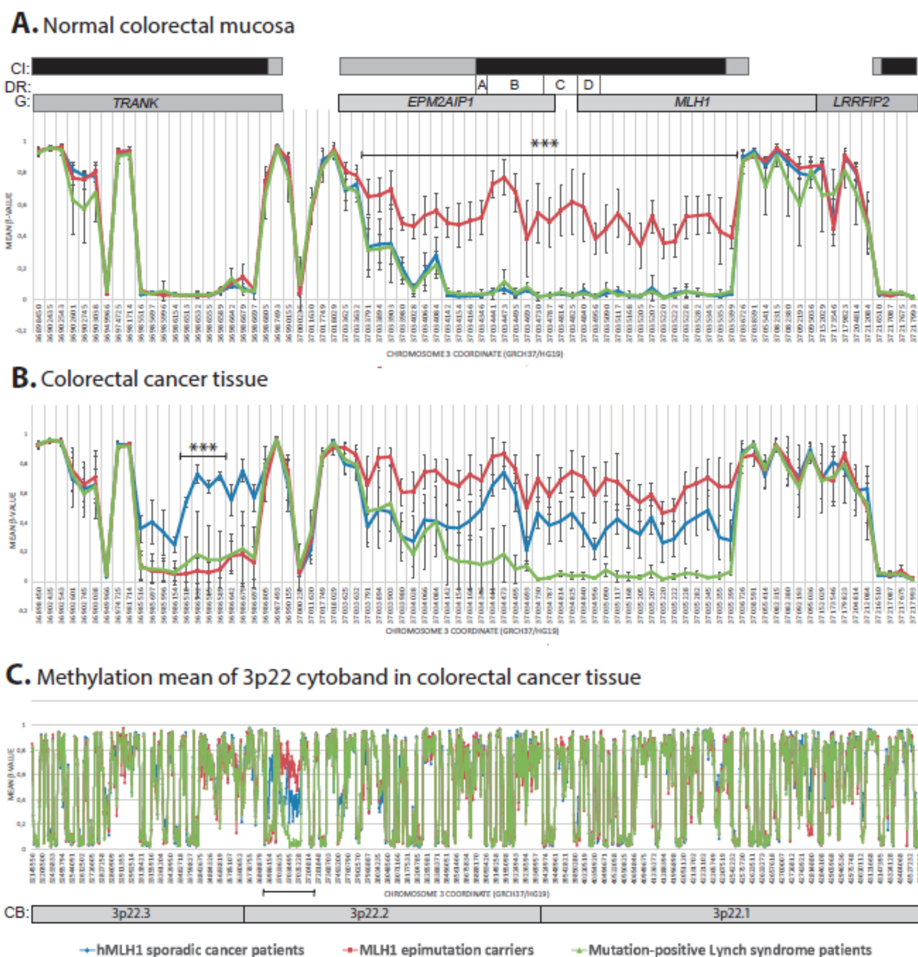


Figure S3 – Representation of mean β-values in FFPE samples from MLH1 epimutation carriers (red), hypermethylated-MLH1 sporadic cancer patients (blue) and mutation-positive Lynch syndrome patients (green). Chromosome coordinates of CpG sites are graphed at axis of abscissa. The locations of the CpG sites are not drawn to scale. Mean β-values of all the analyzed CpG sites in A) normal colorectal mucosa samples and B) colorectal cancer samples. C) Mean β-values of 3p22 cytoband in colorectal cancer samples. 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 (A, B, C and D). 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. Differentially methylated regions with an FDR adjusted p-value lower than 0.005 are signalized by ***.

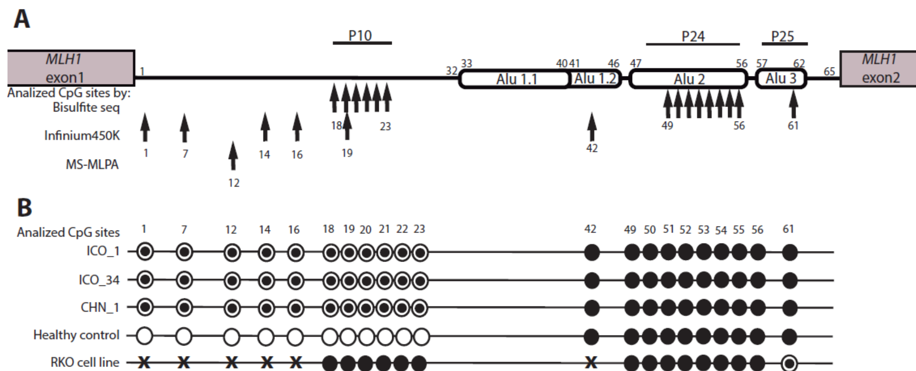


Figure S4– Representation of methylation state of Alu elements present at *MLH1* intron 1. A) Relative position of Alu elements in *MLH1* intron 1. Exon boundaries are represented as grey squares. Alu elements are drawn as white squares. CpG site's number is given according to its position at intron 1, using NM_000249 as reference. Analyzed CpG sites are indicated by arrows and its CpG site number is indicated below them. Bisulfite sequencing amplicons (P10, P24 and P25 (Wang *et al*, 2011)) are graphed at the top part. B) Methylation results of analyzed CpG sites. Each horizontal line represents one individual/cell line. CpG dinucleotides are depicted by circles. Black circles indicate fully methylated CpG sites, white circles unmethylated ones and white circles with an internal black point indicate partially methylation. CpG sites not tested for an individual/cell line are represented by a cross.

Figure S5 –Mean β-values in colorectal cancer samples from TCGA *MLH1*-methylated CRC samples (purple) graphed in combination with data from Figure S3B.

Supplementary tables

Table S1 - Summary of the most relevant clinical features of the constitutional *MLH1* epimutation carriers included in the current								
CaseID at current study	**Publications**	**CaseID given at previous reports**	**Gender**	**Clinical Criteria**	**Age at diagnosis/ location**	**IHC loss**	**BRAF genotype**	***MLH1* promoter methylation status**
ICO_1	Pineda, Mur *et al.,* 2012	1	Male	Bethesda	32 / Left colon 34 / Right colon	MLH1/PMS2 NA	WT NA	M M
ICO_34	Pineda, Mur *et al.*, 2012	34	Female	Bethesda	29 / Left colon 47 / Right colon 49 / Endometrium	NA MLH1/PMS2 MLH1/PMS2	NA WT WT	NA M M
M60	Crucianelli *et al.,* 2014	M60	Female	Bethesda	55 / Endometrium 66 / Uterus 66 / Colon 76 / Breast	MLH1/PMS2 MLH1/PMS2 MLH1/PMS2 MLH1/PMS2	NA NA NA NA	NA NA M M
HGUE_1	Castillejo *et al.,* 2015	1	Female	Bethesda	60 / Bladder 62 / Right colon 63 / Endometrium 66 / Pancreas	MLH1/PMS2 NA NA NA	WT NA NA NA	M NA NA NA
HGUE_2	Castillejo *et al.,* 2015	2	Male	Bethesda	40 / Left colon 48 / Right colon	NA MLH1/PMS2	NA WT	NA M
HGUE_3	Castillejo *et al.,* 2015	3	Male	Bethesda	44 / Right colon	MLH1/PMS2	WT	M
HGUE_4	Castillejo *et al.,* 2015	4	Male	Bethesda	30 / Right colon	MLH1/PMS2	WT	M
HGUE_5	Castillejo *et al.*, 2015	5	Male	Bethesda	39 / Right colon	MLH1/PMS2	WT	M
CHN_1	Unpublished		Female	Bethesda	37 / Colon 43 / Endometrium 55 / Kidney 59 / Colon	MLH1/PMS2 MLH1/PMS2 MLH1/PMS2 MLH1/PMS2	WT WT WT WT	NA NA NA M
CHN_2	Unpublished		Female	Bethesda	38/Colon	MLH1/PMS2	WT	NA
H12O_A	Unpublished		Male	Bethesda	39 / Colon	MLH1/PMS2	WT	M
H12O_B	Unpublished		Female	Bethesda	49 / Colon	MLH1/PMS2	WT	NA

NA=Not available, WT=wild-type, M=methylated

Table S2 - Patients included for genome-wide methylation study.

		Included samples						Analyzed samples (with passed Quality control)						
Group	**Affected gene**	**Included Males**	**Included Females**	**Blood samples**	**FFPE Normal Colon mucosa**	**FFPE Colorectal cancer**	**Paired Normal and Tumor**	**Blood samples**	**Mean age Blood**	**FFPE Normal Colon mucosa**	**Mean age Normal**	**FFPE Colorectal cancer**	**Mean age of onset**	**Paired Normal+ Tumor**
Constitutional *MLH1* Epimutation carriers	***MLH1***	**6**	**6**	**12**	**4**	**7**	**4**	**11**	**59±12years**	**4**	**35±7years**	**6**	**42±8years**	**3**
Healthy controls	**NA**	**12**	**29**	**41**	**0**	**0**	**0**	**41**	**51±11years**	**0**	**NA**	**0**	**NA**	**0**
***MLH1* hypermethylated sporadic CRC patients**	***MLH1***	**1**	**3**	**3**	**3**	**4**	**3**	**3**	**61±12years**	**3**	**59±14years**	**4**	**57±12years**	**3**
Total Mutation-positive Lynch syndrome patients		**30**	**31**	**61**	**19**	**25**	**15**	**61**	**53±8years**	**17**	**50±10years**	**21**	**50±9years**	**14**
Mutation-positive Lynch syndrome patients	***MLH1***	9	12	21	6	9	6	21	49±8years	6	47±8years	8	46±7years	6
Mutation-positive Lynch syndrome patients	***MSH2***	10	18	28	6	8	4	28	47±9years	5	40±9years	7	43±10years	4
Mutation-positive Lynch syndrome patients	***MSH6***	8	1	9	4	5	2	9	55±14years	3	50±12years	3	48±6years	1
Mutation-positive Lynch syndrome patients	***PMS2***	3	0	3	3	3	3	3	64±10years	3	63±13years	3	63±13years	3

Table S3 – Primers and PCR amplification conditions.

Analysis	Sense	Primer sequence (5'-3')	Amplicon size	Annealing temperature	Reference
Mutational analysis of MLH1 promoter					Pineda et al, 2012
PCR and sequencing primer	Forward	AACCCTTTCACCATGCTCTG	1,4 Kb	59	
	Reverse	CCTCGTGCTCACGTTCTTC			
Internal primers for MLH1 promoter sequencing	Forward	TACATGCTCGGGCAGTACCT		55	
	Reverse	TGAAGAGAGAGCTGCTCGTG		55	
Mutational analysis of MLH1 Exon 1					
PCR and sequencing primer	Forward	GGGGCTGGATGGCGTAAG	438 bp	55	
	Reverse	TCGGGGGAGAGCGGTAAA			
Clonal Bisulfite sequencing of MLH1 promoter					Hitchins et al, 2005
A Deng's region amplification	Forward	GTTTTGAYGTAGAYGTTTTATTAGGGT	281bp	52	
	Reverse	TTAACCCTACTCTTATAACCTCCC			
C Deng's region amplification	Forward	TATTTTAGTAGAGGTATATAAGTT	275 bp		
	Reverse	CCTTCAACCAATCACCTCAATAC			
Sequencing primers	T7_Forward	TATTTTAGTAGAGGTATATAAGTT			
	T3_Reverse	CCTTCAACCAATCACCTCAATAC			
SNuPE for *MLH1* c.655A>G variant					Pineda et al, 2012
PCR for cDNA amplification	Forward	CACAATGCAGGCATTAGTTTCTC	326 bp	59	
	Reverse	AGGTACAGGAATGGGTGTGTG			
PCR for genomic amplification	Forward	GTTTCAGTCTCAGCCATGAG	376 bp	55	
	Reverse	ACACATGATTCACGCCACAG			
SNaPShot primer	Reverse	TTCTCGACTAACAGCATTTCCAAAGA			
Bisulfite sequencing of Alu elements of *MLH1* intron 1					Wang et al, 2011
p10 amplification and sequencing - Control region outside Alu element (MSP)	Forward	TATGTTTATAACGGCGGAGGTCGT	226 bp	61	
	Reverse	CATCTTTTAACTTCGCATATTCTACATACA			
P24 amplification and sequencing - Alu element 2	Forward	GGGTAATTTTTGTTTTTTGGTTTTAAG	271 bp	61	
	Reverse	AAAAAAATTATCTATTCCCCACACAAAC			
P25 amplification and sequencing - Alu element 3	Forward	TTGGTTTGTGTTTATAGGTTATTTTT	462 bp	61	
	Reverse	AAATAATTTTCTCTACTATTCCCTTT			

Table S4 – Summary of copy number variations previously reported in the Database of Genomic Variant, which colocalize with deletions found by CGH array in patients HGUE_1 and HGUE_2 (highlighted in bold).

	ID	DGV classification	Start (Chr3)	End (Chr3)	Lenght (bp)	Reported by
Identified CNV	**HGUE_2 (15Kb)**	**Deletion**	**36798479**	**36813411**	**14932**	**Current study**
Reported CNVs colocalizing with HGUE_2 (15Kb)	nsv834660	Deletion	36652421	36852834	200413	Wong *et al*, 2007
	nsv834661	Deletion	36776052	36968161	192109	Wong *et al*, 2007
Identified CNV	**HGUE_2 (19Kb)**	**Deletion**	**37486324**	**37505162**	**18838**	**Current study**
Reported CNVs colocalizing with HGUE_2 (19Kb)	nsv1074583	Deletion	37493499	37493800	301	Thareja *et al,* 2015
	nsv1112822	Deletion	37493500	37494200	700	Alsmadi *et al,* 2014
	esv3595885	Gain	37358053	37497355	139302	1000 Genomes consortium
	esv3595886	Gain	37358299	37570384	212085	1000 Genomes consortium
	esv3595887	Gain	37360630	37919646	559016	1000 Genomes consortium
Identified CNV	**HGUE_1 (20Kb)**	**Deletion**	**36396587**	**36416879**	**20292**	**Current study**
Reported CNVs colocalizing with HGUE_1 (20Kb)	nsv522903	Deletion	36385822	36404564	18742	Shaikh *et al,* 2009
	dgv884e214	Deletion	36396268	36417612	21344	1000 Genomes consortium
	dgv2509e106	Deletion	36396287	36416913	20626	Alsmadi *et al*, 2014
	esv2763782	Deletion	36399248	36422789	23541	Vogler *et al*, 2010
	esv3595859	Deletion	36401878	36402697	819	1000 Genomes consortium
	nsv515984	Deletion	36403964	36405074	1110	Shaikh *et al*, 2009
	gssvL81050	Gold standard variant	36396268	36422789	26521	NR Frequecy = 0,14%
	esv3595854	Gain	36371382	36448953	77571	1000 Genomes consortium
	esv3595858	Gain	36396350	36417612	21262	1000 Genomes consortium
	nsv1006443	Gain	36403321	36517759	114438	Coe *et al*, 2014
	esv3595860	Inversion	36406362	36407627	1265	1000 Genomes consortium

Table S5 – Haplotype analyses in MLH1 epimuation families. Haplotypes associated with the MLH1 methylated allele in probands are highlighted in bold.

Microsatellite/SNP:		D3S1609	D3S1612	D3S2369	rs34566456	rs1800734	rs4234259	rs1799977	D3S1611	rs9876116	D3S3623	D3S1298	D3S3564	Epigenetic state of *MLH1* promoter in blood	Methylation Associated polymorphism at proband by Clonal Bisulfite sequencing
NM_000249.3:		–	–	–	c.-593G>C	c.-93G>A	c.453+79A>G	c.655A>G	–	c.1668-19A>G	–	–	–		
Coordinate (GRCh37/hg19):		3:29940075-29940327	3:34589535-34589642	3:36497205-36497311	3:37034446	3:37034946	3:37048633	3:37053568	3:37068501-37068762	3:37083740	3:37443531-37443747	3:38048763-38048958	3:42418977-42419188		
ICO_1 Family															
II.1-ICO_1	Proband	**250**/250	**92**/104	**102**/102	–	**A**^{m}/G	**A**/G	**A**/G	**250**/256	**A**/G	**218**/222	**200**/210	202/**210**	Methylated	c.-93A
I.1	Father	250/254	104/102	102/104	–	G/G	G/G	G/G	256/256	G/G	222/222	200/210	202/210	Unmethylated	–
III.1	Daughter	**250**/254	**92**/102	100/**102**	–	**A**/G	**A**/A	**A**/A	**250**/260	**A**/A	**218**/212	**200**/208	**210**/210	Unmethylated	–
III.2	Daughter	250/254	104/102	100/102	–	G/G	G/A	G/A	256/260	G/A	222/212	208/210	202/210	Unmethylated	–
ICO_34 Family															
II.3-ICO_34	Proband	**249** / 249	93 / **109**	102 / **104**	^{m}**G**/C	**G**/G	–	**A**/A	**261** / 261	–	**217** / 217	**200** / 206	**203** / 211	Methylated	c.-593G (mosaic)
II.1	Sister	**249** / 253	97 / **109**	**104** / 104	**G**/G	**G**/G	–	**A**/A	**261** / 261	–	**217** / 217	**200** / 208	**203** / 203	Unmethylated	–
II.4	Sister	**249** / 249	93 / **109**	102 / **104**	**G**/C	**G**/G	–	**A**/A	**261** / 261	–	**217** / 217	**200** / 206	**203** / 211	Unmethylated	–
II.5	Sister	249 / 249	93 / 93	102 / 102	G/C	G/G	–	A/G	257 / 261	–	215 / 217	196 / 206	207 / 211	Unmethylated	–
II.7	Sister	249 / 253	93 / 97	104 / 102	G/G	G/G	–	A/G	257 / 261	–	217 / 217	196 / 208	203 / 207	Unmethylated	–
HGUE_2 Family															
II.1-HGUE_2	Proband	**249**/249	97/**105**	102/104	–	**G**^{m}/A	–	**A**/A	261/**271**	–	219/**223**	**210**/212	**203**/211	Methylated	c.-93G (mosaic)
III.2	Son	249/253	97/107	102/104	–	A/A	–	A/A	251/261	–	217/219	212/216	211/211	Unmethylated	–
III.1	Son	249/249	97/107	102/104	–	A/A	–	A/A	251/261	–	217/219	212/216	211/211	Unmethylated	–
IV.1	Granddaughter	249/249	97/105	102/104	–	G/A	–	G/A	257/261	–	219/223	202/212	211/211	Unmethylated	–
IV.2	Granddaughter	249/249	97/105	102/104	–	G/A	–	G/A	257/261	–	219/223	202/212	211/211	Unmethylated	–
CHN_1 Family															
II.1-CHN_1	Proband	**249**/253	**93**/93	**104**/104	–	G/A^{m}	G/**A**	G/**A**	212/**214**	G/**A**	**251**/257	**216**/224	**200**/200	Methylated	c.-93A
II.2	Sister	**249**/253	**93**/93	**104**/104	–	G/**A**	G/**A**	G/**A**	212/**214**	G/**A**	**251**/257	**216**/224	**200**/200	Unmethylated	–
II.3	Sister	249/253	93/97	104/104	–	G/G	G/G	G/A	210/212	G/A	257/257	224/224	198/200	Unmethylated	–
III.1	Son	**249**/251	**93**/97	**104**/106	–	G/**A**	G/**A**	G/**A**	204/**214**	G/**A**	**251**/257	**216**/216	**200**/202	Unmethylated	–
III.2	Daughter	**249**/249	**93**/97	**104**/104	–	**A**/A	**A**/A	**A**/A	204/**214**	**A**/A	**251**/251	**216**/218	**200**/200	Unmethylated	–
H12O-A Family															
II.3-H12O-A	Proband	–	93/93	102/102	–	**G**^{m}/A	–	**G**/A	–	–	–	–	–	Methylated	c.-93G
I.1	Mother	–	93/105	102/104	–	G/A	–	G/A	–	–	–	–	–	Unmethylated	–
II.1	Sister	–	93/97	100/102	–	G/A	–	G/A	–	–	–	–	–	Unmethylated	–
II.2 -A3	Sister	–	97/105	104/104	–	G/G	–	G/A	–	–	–	–	–	Unmethylated	–

Table S6 – Reported constitutional MLH1 epimutation carriers with their principal features listed. Adapted from Hitchins 2016 and updated with patients in the current study.

Type of epimutation	ID	Publication	Gender	Clinical criteria	Genetic alteration	Methylation in blood	Variant in cis with methylation	Loss of transcription	Mosaicism	*MLH1* Methylation in first degree relatives	MAA in relatives	Intergenational erasure	Onset	Transmission
Secondary epimutation	#4 or Case 2	Morak et al. 2008; Morak et al. 2011	Male		From exon 1 to exon 19 including *MLH1* promoter region	10-20%	Homozygous		Yes	M	Yes			Yes
	F36	Gylling et al. 2009 / Renkonen et al. 200	Female	BC	Deletion from c.-63 to intron 2 (6,4Kb)	50%	c.655A	c.655A (Complete silencing)	No		No	No		Unknown
	WA Family 16	Hitchins et al. 2011	Female	AMS	c.[-27C>A;85G>T]	24,90%	c.[-27A;85T]	*EMP2AIP1* rs9311149A	Yes	M	Yes	No		Yes
	CCFR-H	Ward et al. 2013	Female	AMS	c.[-27C>A;85G>T]	32%	c.[-27A;85T]	c.-27A (Partial silencing)	Yes		Unknown			Unknown
	USA family 1 or Family 1744	Kwok et al. 2014; Raevaara et al.2005	Female	AMS	c.[-27C>A;85G>T]	32%	c.[-27A;85T]	c.85T (partial silencing)	Yes	M	Yes	No		Yes
	USA family 2	Kwok et al. 2014	Male	AMS	c.[-27C>A;85G>T]	29%	c.[-27A;85T]	c.85T (partial silencing)	Yes	M	Yes	No		Yes
	Netherlands family 3	Kwok et al. 2014	Male	AMS	c.[-27C>A;85G>T]	23/30%	c.[-27A;85T]	c.85T (partial silencing)	Yes	M	Yes	No		Yes
	CFS279	Cini et al. 2015	Female	AMS	c.-168_c.116+713del (997 bp)	50%	c.-168_c.116+713del	c.-168_c.116+713del (partial silencing)	Unknown	M	Yes	No		Yes
Supected secondary epimutation	B	Crépin et al. 2012	Female	AMS	Not identified but associated to one genetic allele	20%		x	Yes	M	Yes	No		Yes
Suspected primary epimutation with transgenerational inheritance	A	Hitchins et al. 2007	Female	BC	Not identified	Positive	c.-93A	*MLH1* c.655A, *EPM2AIP1* rs9311149G	Unknown	M / UM	Yes	Yes		Yes, non-Mendelian
	CCFR-N	Ward et al. 2013	Male		Not identified	2-4%	c.+27A	c.+27A (Complete silencing)	Yes	M / UM		No	De novo	Yes
	Proband/III-7	Sloane et al. 2015	Male	RBG	Not identified	40%	c.116+680A	c.655A	No	M / UM		Yes		Yes, non-Mendelian
Primary epimutation with intergenerational erasure	TT	Suter et al. 2004	Male	AMS	Not identified	Positive	c.-93G		Yes	UM	Yes	Yes		No
	B	Hitchins et al. 2007	Female	BC	Not identified	Positive	Homozygous	*EPM2AIP1* rs9311149G	Unknown	UM	Yes	Yes	De novo	No
	3	Morak et al. 2008	Male	BC	Not identified	10-20%	Homozygous		Yes	UM	Yes	Yes		No
	ICO_1; 1	**Current study**; Pineda et al. 2012	Male	BC	Not identified	62%	c.-93A	c.655A	No	UM	Yes	Yes	De novo	No
	CHN_1	**Current study**	Female	BC	Not identified	47,5%	c.-93A	c.655A	No	UM	Yes	Yes		No

Table S6 – Reported constitutional MLH1 epimutation carriers with their principal features listed. Adapted from Hitchins 2016 and updated with patients in the current study.

Type of epimutation	ID	Publication	Gender	Clinical criteria	Genetic alteration	Methylation in blood	Variant in cis with methylation	Loss of transcription	Mosaicism	*MLH1* Methylation in first degree relatives	MAA in relatives	Intergenational erasure	Onset	Transmission
Suspected primary epimutation	no name	Gazzoli et al. 2002	Female	BC	Not identified	Positive	c.-93G		Unknown	NA				
Suspected primary epimutation	H166	Miyakura et al. 2004	Female	BC	Not identified	Positive	c.-93A		No	NA				
Suspected primary epimutation	H403	Miyakura et al. 2004	Male	BC	Not identified	Positive	Homozygous		Unknown	UM				
Suspected primary epimutation	H450	Miyakura et al. 2004	Female	BC	Not identified	Positive	c.-93G		No	NA				
Suspected primary epimutation	H628	Miyakura et al. 2004	Male	BC	Not identified	Positive	Homozygous		Unknown	NA				
Suspected primary epimutation	VT	Suter et al. 2004	Female	AMS	Not identified	Positive	c.-93G		Yes	UM				
Suspected primary epimutation	ST	Hitchins et al. 2005	Male	BC	Not identified	Positive	c.2363-2365CTT	c.2363-2365CTT	Unknown	UM	Yes		De novo	
Suspected primary epimutation	2	Valle et al. 2007	Male	BC	Not identified	Positive	Homozygous		Unknown	UM			De novo	
Suspected primary epimutation	1	Morak et al. 2008	Male	BC	Not identified	45-55%	c.-93A	C.655A	Unknown	NA				
Suspected primary epimutation	2	Morak et al. 2008	Female	BC	Not identified	45-55%	Homozygous		Unknown	NA				
Suspected primary epimutation	5	Morak et al. 2008	Female	BC	Not identified	45-55%	c.-93G		Unknown	UM	Yes		De novo	
Suspected primary epimutation	6	Morak et al. 2008	Male	BC	Not identified	45-55%	c.-269G		Unknown	UM	No			No
Suspected primary epimutation	7	Morak et al. 2008	Male	BC	Not identified	45-55%	c.-93G		Unknown	NA				
Suspected primary epimutation	8	Morak et al. 2008	Male	BC	Not identified	19-31%	Homozygous	c.655G	Yes	NA				
Suspected primary epimutation	9	Morak et al. 2008	Female	BC	Not identified	19-31%	c.-93G		Yes	UM	Yes		De novo	
Suspected primary epimutation	10	Morak et al. 2008	Female	BC	Not identified	45-55%	c.-93A		Unknown	NA				
Suspected primary epimutation	11	Morak et al. 2008	Male	BC	Not identified	10-20%	Homozygous		Yes	NA				
Suspected primary epimutation	12	Morak et al. 2008	Female	BC	Not identified	19-31%	c.-93G		Yes	NA				
Suspected primary epimutation	H10	Zhou et al. 2008	Male	BC	Not identified	Exhaustive	NA		Unknown	NA				
Suspected primary epimutation	H29	Zhou et al. 2008	Female	Am I	Not identified	Exhaustive	NA		Unknown	NA				
Suspected primary epimutation	H32	Zhou et al. 2008	Male	Am I	Not identified	Partial	NA		Yes	NA				
Suspected primary epimutation	H42	Zhou et al. 2008	Male	Am I	Not identified	Partial	NA		Yes	NA				
Suspected primary epimutation	H46	Zhou et al. 2008	Male	Am I	Not identified	Partial	NA		Yes	NA				
Suspected primary epimutation	N2	Gylling et al. 2009	Female	BC	Not identified	50%	c.655G	c.655G	Unknown	UM				
Suspected primary epimutation	4	Niessen et al. 2009	Male	BC	Not identified	Positive	NA		Unknown	UM			De novo	
Suspected primary epimutation	5	Niessen et al. 2009	Female	BC	Not identified	Positive	NA		Unknown	UM				
Suspected primary epimutation	ID60	van Roon et al. 2010	Female	BC	Not identified	Positive	NA		Unknown	NA				
Suspected primary epimutation	ID36	van Roon et al. 2010	Female	None	Not identified	Partial	NA		Unknown	NA				
Suspected primary epimutation	YT	Goel *et al.* 2011	Male	BC	Not identified	40-55%	c.655A	c.655A	Yes	UM	Yes		De novo	
Suspected primary epimutation	BF	Goel *et al.* 2011	Female	BC	Not identified	22%	c.474T	c.474T	Yes	UM	Yes		De novo	
Suspected primary epimutation	84-1	Hitchins et al. 2011	Female	BC	Not identified	10,2%	c.-93G		Yes	NA				
Suspected primary epimutation	291-1	Hitchins et al. 2011	Male	BC	Not identified	41%	c.-93A		Yes	NA				
Suspected primary epimutation	A	Crépin et al. 2011	Male	BC	Not identified	50%	G/G		No	NA			Proposed de novo	
Suspected primary epimutation	836	Auclair et al. 2011	Unknown	None	Not identified	38%	NA	not shown	Yes	NA				
Suspected primary epimutation	2286	Auclair et al. 2011	Unknown	None	Not identified	36%	c.655A	c.655A	Yes	NA				
Suspected primary epimutation	2138	Auclair et al. 2011	Unknown	None	Not identified	30% not confirme	NA		Yes	NA				
Suspected primary epimutation	2255	Auclair et al. 2011	Unknown	None	Not identified	30% not confirme	NA		Yes	NA				
Suspected primary epimutation	61	Auclair et al. 2011	Unknown	None	Not identified	30% not confirme	NA		Yes	NA				
Suspected primary epimutation	1492	Auclair et al. 2011	Unknown	None	Not identified	30% not confirme	NA		Yes	NA				
Suspected primary epimutation	2007	Auclair et al. 2011	Unknown	None	Not identified	30% not confirme	NA		Yes	NA				
Suspected primary epimutation	2395	Auclair et al. 2011	Unknown	None	Not identified	17%	NA	not shown	Yes	NA				
Suspected primary epimutation	**ICO_34**; 34	**Current study**; Pineda et al. 2012	Female	BG	Not identified	32%	c.-593G		Yes	NA			Proposed de novo	
Suspected primary epimutation	G46	Wu et al. 2012	Female	None	Not identified	10%	c.-93A		Yes	NA				
Suspected primary epimutation	CCFR-A	Ward et al. 2013	Male	BG	Not identified	36%	c.-93A		Yes	NA				
Suspected primary epimutation	CCFR-B	Ward et al. 2013	Female	BG	Not identified	17%	c.-269C		Yes	UM				
Suspected primary epimutation	CCFR-C	Ward et al. 2013	Male	BG	Not identified	45%	c.-93G			UM				
Suspected primary epimutation	CCFR-D	Ward et al. 2013	Male	BG	Not identified	11-15%	Homozygous		Yes	UM				
Suspected primary epimutation	CCFR-E	Ward et al. 2013	Male	BG	Not identified	46%	c.-93G		Yes	UM				
Suspected primary epimutation	CCFR-P	Ward et al. 2013	Female	None	Not identified	2%	c.-93G		Yes	NA				
Suspected primary epimutation	CCFR-O	Ward et al. 2013	Female	BG	Not identified	4%	NA		Yes	NA				
Suspected primary epimutation	CCFR-Q	Ward et al. 2013	Male	BG	Not identified	3%	Homozygous		Yes	UM				
Suspected primary epimutation	M60	**Current study**; Crucianelly et al. 2014	Female	RBG	Not identified	58,6%	Homozygous	c.655A	Unknown	UM				
Suspected primary epimutation	**HGUE_1**; #1	**Current study;** Castillejo et al. 2015	Female	RBG	Not identified	50%	Homozygous		Unknown	UM				
Suspected primary epimutation	**HGUE_2**; #2	**Current study;** Castillejo et al. 2015	Male	RBG	Not identified	66%	c.-93G	X	Yes	UM	No			
Suspected primary epimutation	**HGUE_3**; #3	**Current study;** Castillejo et al. 2015	Male	RBG	Not identified	76%	Homozygous		Unknown	UM				
Suspected primary epimutation	**HGUE_4**; #4	**Current study;** Castillejo et al. 2015	Male	RBG	Not identified	51,5%	Homozygous		Unknown	NA				
Suspected primary epimutation	**HGUE_5**; #5	**Current study;** Castillejo et al. 2015	Male	RBG	Not identified	49,5%	c.-93A		Yes	UM			Proposed de novo	
Suspected primary epimutation	no name	Kidambi et al. 2016	Male	RBG	Not identified	Positive	NA		Unknown	NA				
Suspected primary epimutation	2	Takeda et al. 2016	Female	AMS	Unknown	Positive	NA		Unknown	NA				
Suspected primary epimutation	**CHN_2**	**Current study**	Female	BC	Not identified	32%	Homozygous	c.655A	Yes	UM				
Suspected primary epimutation	**H120-A**	**Current study**	Male	BC	Not identified	40,3%	c.-93G	x	No	UM				
Suspected primary epimutation	**H120-B**	**Current study**	Female	BC	Not identified	45,2%	Homozygous	x	Yes	UM				

Supplementary methods

Lymphoblastoid cell line establishment

Nine milliliters of *MLH1* epimutation carriers' blood were collected in EDTA tubes and were processed in the following 24h. Peripheral blood mononuclear cells (PBMC) were isolated using 4ml of lymphocytes isolation solution (Rafer S.L.). PBMC were collected with a Pasteur pipet and a wash was done using phosphate-buffered saline buffer. PBMC pellets were incubated for 3 hours at 37º C with 200 µL of 0.45 µm-filtered supernatant from a culture of the EBV producer cell line B95.8. RPMI medium 1640+GlutaMAX™-I (Glibco) supplemented with 10% fetal bovine serum and 10 µg/ml of Phytohemaglutinin-L (Sigma-Aldrich Quimica S.L.) was then added. The cells were cultured with frequent media changes for 2 or 3 months, until immortalization.

Detailed methylation analysis methods

Clonal bisulphite sequencing of *MLH1* promoter fragments was used to determine the methylation profile of individual alleles in patients who were heterozygous for a genetic variant within the *MLH1* CpG island. PCR amplification was performed on 100-200 ng of bisulfite modified DNA using Platinum Taq polymerase (Invitrogen) using primers and conditions listed in Figure S3. PCR products were purified using the QIAquick PCR purification kit (QIAGEN) and eluted in 15 µl of water, and 4 µl of eluate was ligated with 1 µl of pCR4-TOPO vector. One vial of competent E. coli cells vial was transformed using 2 µl of the ligation reaction, as indicated for the TOPO TA Cloning kit for sequencing (Invitrogen). Transformed competent cells (in 25, 50 and 100 µl volumes) were spread onto selective LB-agar plates containing ampicillin and incubated overnight at 37ºC. At least 24 individual colonies were picked at random for sequencing the plasmid inserts. In order to determine the methylation associated allele, the colony-PCR was performed using T3/T7 primers in 24 colonies. Colony-PCR products were purified by Exosap-it (Affymetrix) and sequenced using BigDye Terminator v.3.1 Sequencing Kit (Applied Biosystems, Carlsbad, CA).

***MLH1* allelic expression analyses**

Lymphocytes from index cases were cultured in PB-MAX™ Karyotyping Medium (Gibco) for up to 8 days, and lymphoblastoid cell lines were cultured in RPMI Medium 1640+GlutaMAX™-I supplemented with 10% fetal bovine serum, at 37 degrees Celcius and 5% carbon dioxide. Total RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's instructions. cDNAs were synthesized using the iScript™ cDNA Synthesis Kit (Bio-Rad).

For allelic expression analyses at the *MLH1* single nucleotide polymorphism (SNP) rs1799977 (c.655A>G), the relative levels of the two alleles were determined in genomic DNA and cDNA by single-nucleotide primer extension (SNuPE). PCR flanking the analyzed heterozygous variants of genomic DNA and cDNA was performed as described (Pineda *et al*, 2012). For SNuPE analysis, the amplified band was analyzed using the ABI PRISM SNaPshot kit (Applied Biosystems) and a specific primer (Table S3). The purified products were run on an ABI Prism 3130 DNA sequencer and analyzed by GeneMapper v4.0 (Applied Biosystems). To obtain ASE (allele-specific expression) values, we used the previously described method (Pineda *et al*, 2012): cDNA (peak height major allele/ peak height minor allele) / gDNA (peak height major allele/ peak height minor allele). The final ASE value was calculated as the mean of the ASE values obtained for the triplicates or quadruplicates studied in each sample. 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.

ARTÍCULO 4:

Highly sensitive *MLH1* methylation analysis in blood allows the identification of low-level epigenetic mosaicism

Hipótesis: Algunos pacientes con epimutación en MLH1 podrían estar siendo infradiagnosticados debido a la presencia de mosaicismo epigenético.

Objetivo: Identificar pacientes con bajos niveles de mosaicismo epigenético en MLH1, los cuales no han sido diagnosticados como portadores de epimutación.

Resumen de resultados obtenidos: Se puso a punto un ensayo de alta sensibilidad para la detección de metilación en la región C de Deng del promotor de MLH1 mediante MS-MCA. Mediante esta técnica se analizaron 22 pacientes con hipermetilación de MLH1 en sus tumores, pero con metilación indetectable en sangre en su análisis mediante MS-MLPA.

El estudio de metilación con alta sensibilidad identificó un caso (4.5%) con bajos niveles de metilación (1-2%) en DNA de sangre. Esta paciente contaba con una historia personal de cáncer muy fuerte, habiendo desarrollado 3 tumores gastrointestinales a las edades de 22, 24 y 25 años. Todos los tumores compartían la hipermetilación del promotor de *MLH1* y la pérdida de heterocigosidad asociada con el alelo c.655A. La presencia de niveles bajos de metilación de MLH1 se confirmó por secuenciación clonal con bisulfito, lo que demostró la asociación con el alelo c.-93G a la metilación. El estudio del metiloma de la paciente demostró que la extensión de la región hipermetilada era igual a la anteriormente descrita para portadores de epimutación primaria en MLH1. No se identificaron variantes genéticas raras en línea germinal ni alteraciones en números de copia, clasificando a esta paciente como portadora de epimutación primaria en MLH1.

El uso de técnicas altamente sensibles como MS-MCA ha demostrado ser útil para la detección de niveles bajos de metilación en sangre.

Contribución del doctorando: Propuesta de la hipótesis. Diseño de experimentos. Obtención, gestión y tratamiento de las muestras. Cultivo primario de linfocitos y fibroblastos. Extracción de RNA. Parte de la secuenciación clonal con bisulfito. Análisis e interpretación de resultados de array CGH y SNP array. Estudio de tejidos mediante Infinium 450K. Análisis, interpretación y discusión de resultados. Preparación de figuras. Escritura de la primera versión del manuscrito.

HIGHLY SENSITIVE *MLH1* METHYLATION ANALYSIS IN BLOOD ALLOWS THE IDENTIFICATION OF LOW-LEVEL EPIGENETIC MOSAICISM

Estela Dámaso*[1], Julia Canet-Hermida*[1], Gardenia Vargas-Parra[1], Àngela Velasco[2], Esther Darder[2], Jesús del Valle[1], Anna Fernández[1], Fátima Marin[1], Àngel Izquierdo[2], Gemma Mateu[3], Virgínia Piñol[4], Hugo-Ikuo Uchima[4], José Luis Soto[5,6], Megan Hitchins[7], Conxi Lázaro[1,8], Bernat Queralt[9], Joan Brunet[2,8], Marta Pineda”[1,8], Gabriel Capellá”[1,8]

(1) Hereditary Cancer Program, Catalan Institute of Oncology, Insititut d’Investigació Biomèdica de Bellvitge (IDIBELL), ONCOBELL Program, L’Hospitalet de Llobregat, Barcelona, Spain.
(2) Hereditary Cancer Program, Catalan Institute of Oncology, Institut d’Investigació Biomèdica de Girona (IDIBGI), Girona, Spain.
(3) Pathology Department, Hospital Universitari Josep Trueta, Girona, Spain.
(4) Department of Gastroenterology, Hospital Universitari Josep Trueta, Girona, Spain.
(5) Hereditary Cancer Program Valencian Region, Molecular GeneticsLaboratory, Elche University Hospital, Elche, Alicante, Spain.
(6) Alicante Institute for Health andBiomedicalResearch (ISABIAL-FISABIO Foundation), Alicante, Spain.
(7) Department of Medicine, Division of Oncology, Stanford University, Stanford, California, United States.
(8) Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain
(9) Department of Medical Oncology, Catalan Institute of Oncology, Girona, Catalonia, Spain.
(*)These authors contributed equally to this work and share first authorship.
(”)These authors contributed equally to this work and share senior authorship.

Short title: Low-level *MLH1* epigenetic mosaicism causing Lynch syndrome

Keywords: Constitutional *MLH1*epimutation, Lynch syndrome, methylation, epigenetic mosaicism, highly sensitive methodologies

ABSTRACT

Constitutional *MLH1* epimutations are a rare cause of Lynch syndrome. Low methylation levels (≤10%) have been occasionally described. The aim of this study was the identification of patients with low levels of epigenetic mosaicism in *MLH1* gene.

Eighteen patients with *MLH1* hypermethylated tumors and undetectable methylation in blood as assessed by Methylation-Specific (MS) Multiplex Ligation-Dependent Probe Amplification were included. Highly sensitive MS-Melting Curve Analysis (MS-MCA) at *MLH1* promoter was used to screen for epigenetic mosaicism. Constitutional methylation was confirmed by other methods. Mutational analysis of hereditary cancer genes including *MLH1* was performed.

MS-MCA analysis identified one case (5.6%) with low levels of methylation (1-2%) in blood DNA. The patient had developed 3 gastrointestinal tumors at ages 22, 24 and 25, sharing *MLH1* promoter hypermethylation and loss of heterozygosity associated with c.655A allele. The presence of low *MLH1* methylation levels was confirmed by clonal bisulfite sequencing, evidencing the association with c.-93G allele. The extension of the

hypermethylated region overlaps with the reported in constitutional *MLH1* epimutation carriers. No rare germline variants were identified.

The use of highly sensitive techniques such as MS-MCA has demonstrated to be useful for the detection of low levels of *MLH1* methylation in blood.

Acknowledgements

We thank the participating patients and families and all the members of the Units of Genetic Counseling and Genetic Diagnostic the Hereditary Cancer Program of the Catalan Institute of Oncology (ICO-IDIBELL). We thank Juana Fernández and Mar Varela for technical assistance. 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 (080253) and the Government of Catalonia (grant 2017SGR1282). ED was supported by a grant from the Spanish Ministry of Economy and Competitiveness. The Mexican National Council for Science and Technology (CONACyT) fellowship to GV. JC and FM were supported by CIBERONC.

Disclosures: The authors declare no conflict of interest.

Correspondence:
Gabriel Capellá, M.D., Ph.D. and Marta Pineda, Ph.D.
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 73193
E-mail: gcapella@iconcologia.net; mpineda@iconcologia.net

INTRODUCTION

Lynch syndrome (LS) is characterized by an increased risk for colorectal cancer (CRC) as well as other cancers (stomach, small intestine and endometrium among others) [1]. It is mainly caused by germline genetic mutations in a mismatch repair (MMR) gene (*MLH1, MSH2, MSH6 or PSM2).* In a small proportion of patients LS is caused by a *MLH1* constitutional epimutation, in which monoallelic hypermethylation of the promoter CpG island throughout normal tissues results in allele-specific silencing [2].

Ninety-seven index cases with a constitutional *MLH1* epimutation have been reported so far [3, 4, 5, 6, 7]. Most are considered primary, arising apparently *de novo* and reversible between generations, whereas secondary ones are associated with an in-cis genetic variant. Recently we demonstrated that *EPM2AIP1-MLH1* CpG island is the sole differentially methylated region in primary *MLH1* epimutation carriers [4]. Available evidence suggests that constitutional epimutations cause a severe LS phenotype, including early-onset and multiple primary tumors [8].

The level of constitutional methylation varies among individuals. Although most cases show hemiallelic methylation in blood, variable levels of germline methylation are frequently reported (reviewed in [4]). Low methylation levels (≤10%) have been described in 7 patients [9, 10, 11, 12] but also in a high proportion of healthy controls as well, hampering its interpretation [13]. The use of robust and sensitive techniques is critical to know the real prevalence of constitutional epimutations and to ascertain the role of *MLH1* epigenetic mosaicism in cancer predisposition.

The main aim of this study was the identification of patients with low levels of epigenetic mosaicism in *MLH1* gene combining a variety of highly sensitive methylation analysis techniques.

PATIENTS AND METHODS (550 words)

Patients and samples

Patients were assessed through Cancer Genetic Counseling Units of the *Institut Català d'Oncologia* (ICO) from 1998 to 2016. A total of 18 individuals presenting *MLH1*-methylated colorectal tumors before 50 years of age or multiple tumors before 60 years were included in this study (Figure S1 and Table S1; Supplementary Methods). Their levels of *MLH1* methylation in blood were previously assessed by MS-MLPA resulting below the limit of detection (between 0% and 4% at Deng C and D regions) and were considered as negative (Table S1). Twenty healthy individuals (selected to appropriately match the patients by age, race and geographic location) were included as controls. In addition, 61 LS cases harboring MMR genetic mutations, 12 constitutional *MLH1* epimutation carriers and 41 healthy controls were included as controls for global methylome analysis [4].

Written informed consent was obtained from all individuals, and the ethics committee of the respective hospitals approved the study. Sample processing is detailed in Supplementary Methods.

Methods

In all cases screening of mutations in hereditary cancer genes (including *MLH1*) was performed as described (see Supplementary Methods). The levels of methylation at the *MLH1* promoter in biological samples were assessed by several methods (see Supplementary Methods): (i) Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) using the SALSA MLPA ME011 Mismatch Repair genes probemix (MRC-Holland). (ii) Methylation-specific melting curve analysis (MS-MCA): Bisulfite treated DNA was used to amplify Deng C and D regions in a nested PCR carried out in a LightCycler 480 II; the analytical sensitivity of the technique is 1% for Deng C region (Figure S2A-B). (iii) Pyrosequencing: Bisulfite treated DNA was used to amplify Deng C and intron 1 regions with biotin-labeled primers; the estimated analytical sensitivity of this technique is 4% and 5% at C-region and intron 1, respectively (Figure S3A-B). (iv) Clonal bisulfite sequencing of fragments of the *MLH1*.

The impact of *MLH1* promoter methylation on allelic expression was assessed by single-nucleotide primer extension (SNuPE) at *MLH1* rs1799977 (c.655A>G) in PBL. The relative levels of the two alleles were determined in cDNA/gDNA as previously described [14]. Loss of heterozygosity (LOH) was determined as the ratio of tumor-DNA / normal-DNA at rs1799977 as assessed by SNUPE. Finally, Genome wide methylation profiling was analysed using Infinium Human Methylation 450K Beadchip as previously described [4].

RESULTS

Highly sensitive *MLH1* methylation screening

MS-MCA analysis at region C in PBL DNA from healthy controls showed the same curve pattern than the unmethylated WGA (Whole Genome Amplification), indicating absence of detectable methylation **(Figure S2C).** Seventeen of 18 patients with *MLH1*-hypermethylated tumors and undetectable methylation in blood by MS-MLPA displayed the same curve pattern than controls. In contrast, case 29 showed levels of methylation around 1% (**Figure 1A and S2D**) that were confirmed in an independently extracted sample (**Figure 1A**).

Case presentation

Patient 29 is a woman who developed three gastrointestinal adenocarcinomas: a colorectal cancer at the age of 22, a small bowel cancer at 24 and a gastric cancer at 25. No family history of cancer was reported in her first-degree relatives (**Figure1B**). Although the presence of methylation in all tumours suggested the presence of constitutional epimutation, MS-MLPA reported background methylation levels in C-D regions in blood DNA and normal gastrointestinal tissues (**Table S2**). Of note, loss of heterozygosity analysis at *MLH1* exonic SNP c.655A>G (rs1799977) evidenced that the three tumors showed loss of expression of the A allele, pointing to a common origin (**Figure S4**).

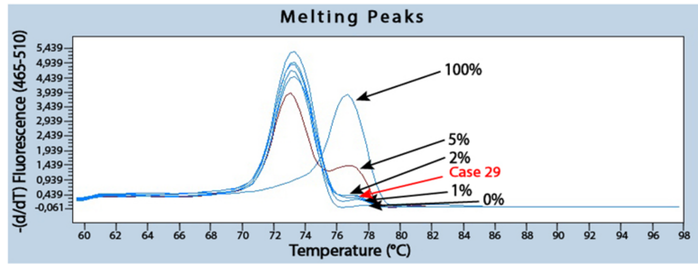
Confirmation of the low level constitutional *MLH1* epimutation

The presence of the low levels of *MLH1* methylation in blood was confirmed by clonal bisulfite sequencing. Nineteen of the 372 clones sequenced (5%; CI 0,01–0,05) were fully methylated. All methylated clones displayed dense monoallelic methylation associated with G allele at rs1800734 (c.-93 G>A) (**Figure 1C**). Of note, no genetic variants associated to methylated G alleles were detected in the analysed region including c.-27A.

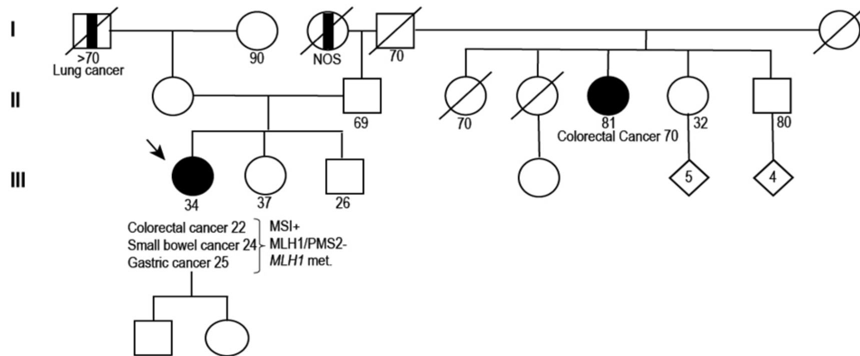
In depth characterization and *MLH1* epimutation classification

Global methylome analysis also revealed slightly higher levels of *MLH1* promoter methylation than controls in blood (**Figure 2A**) and normal colonic mucosa (**Figure 2B**),

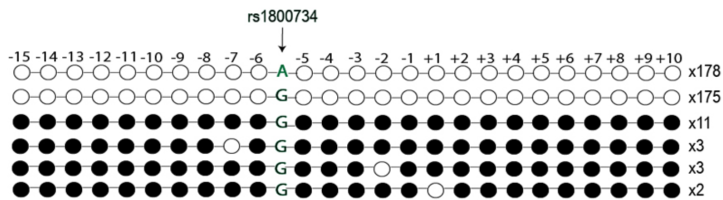
A.



B.

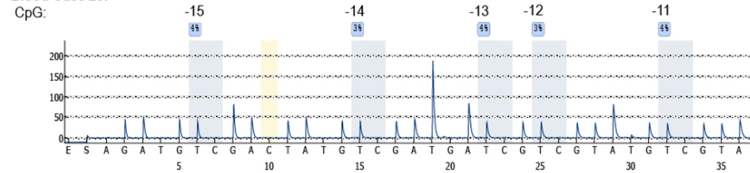


C.



D.

Sequence: GAGYGGATAGYGATTTTTAAYGYGTAAGYGTA



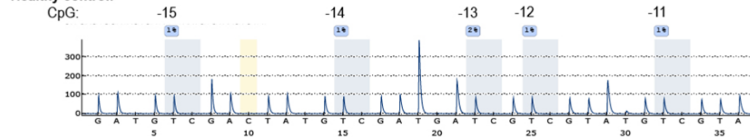
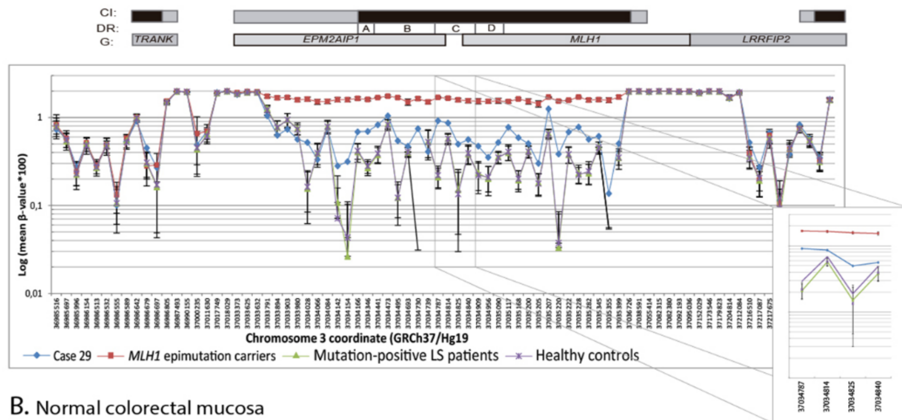


Figure 1. Identification and characterization of epigenetic mosaicism in case 29. A) Methylation analysis by MS-MCA of promoter C-region in blood DNA from case 29. The patient shows levels of methylation around 1%. **B)** Family pedigree from case 29. The epimutation carrier is indicated by an arrow. Circles, females; squares, males; filled, cancer affected; vertical line at center, not otherwise specified. Cancer localization and age at diagnosis are indicated. Generations are indicated on the left margin in Roman numerals. **C)** Clonal bisulfite sequencing of the *MLH1* promoter in PBL DNA from case 29. Each horizontal line represents a specific allele. CpG dinucleotides are depicted by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. The allele at rs1800734 (c.-93G>A) is indicated as A or G. Methylation is confined to the G allele. Each CpG analyzed is numbered according to its position relative to the translation initiation codon. **D)** *MLH1* methylation analysis in the promoter C-region by pyrosequencing from case 29 and one healthy control. Each CpG is numbered according its position relative to the translation initiation codon.

---

affecting the same region involved in constitutional epimutations [4]. Similar levels of *MLH1* methylation was found by MS-MCA in gastric, small bowel and colon mucosa (endoderm), oral mucosa and skin fibroblasts (ectoderm), confirming the presence of methylation in all embryonic layers (**Figure S2)**. Finally, the levels of germline methylation detected by pyrosequencing were slightly higher than in controls but below the threshold of the analytical sensitivity of the technique (**Figure S3**).

No germline variants were found in *MLH1* promoter and coding regions in patient 29 when analyzed by Sanger sequencing or Next Generation Sequencing. In addition, no *MLH1* germline copy number alterations were identified (**Figure S5**), further suggesting a primary type of the epimutation. Further mutational analysis of hereditary cancer genes only identified two variants of unknown significance in *XRCC1* and *NBN* (**Table S4**). In accordance with the low percentage of methylation detected in blood, a slight reduction in the expression of *MLH1* c.655G (rs1799977) allele transcript was observed in lymphocytes from patient 29 compared with two controls (**Figure S4B**).



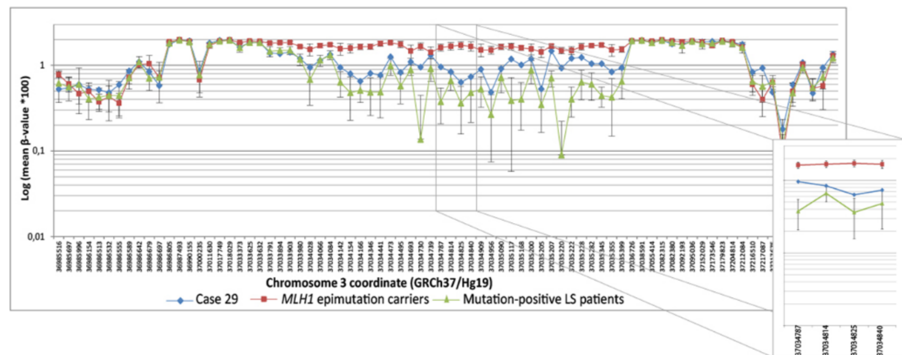


Figure 2. Representation of the differentially methylated region across the *MLH1* locus in blood and colorectal mucosa from case 29, *MLH1* epimutation carriers (n=12), Lynch syndrome mutation carriers (n=61) and healthy controls (n=41). A) Representation of the differentially methylated region across the *MLH1* locus in blood DNA. **B)** Representation of the differentially methylated region across the *MLH1* locus in normal colorectal mucosa. β-values obtained from Infinium 450k Human Methylation array analysis are displayed as a log(mean β-value*100) against the genomic coordinate for each CpG site interrogated. The relative locations of the CpG sites are not drawn to scale. CpG sites are located between Chr3:36,985,516-37,219,077 coordinates. Above, CpG islands (CI:) are represented as black rectangles and their shores are represented in grey. The location of the Deng regions (DR:) A, B, C and D of the *MLH1* promoter are indicated by white rectangles. Genes (G:) containing the displayed CpG sites are represented as grey rectangles, using the Ensembl GRCh37 database as the reference for gene coordinates.

DISCUSSION

Here we report the presence of low-level *MLH1* epigenetic mosaicism in a woman who suffered from 3 LS-spectrum *MLH1*-methylated tumors in her early twenties. A comprehensive approach that combined highly sensitive MS-MCA with clonal bisulfite sequencing and array analysis confirmed the presence of dense allele-specific methylation covering the whole *EPM2AIP1-MLH1* CpG island in a low proportion (around 1%) of the *MLH1* alleles.

The MS-MCA approach for *MLH1* methylation analysis has allowed the robust detection of low-level methylation in blood. Although MS-MLPA and pyrosequencing are widely used for *MLH1* methylation analysis in the clinical diagnostic routine [7, 10, 15, 16], its analytical sensitivity is lower (5-10%) than MS-MCA (1%) ([17] and current study), preventing the detection of low-level epigenetic mosaicism. The low analytical sensitivity of pyrosequencing could account for the high proportion (78%) of low-level methylation (<10%) previously reported in healthy controls [13]. In contrast, no evidence of methylation was detected in 20 controls analyzed by MS-MCA in the current study. The presence of constitutional methylation was only confirmed in one of the 5 patients (case 29) showing methylation levels between 1-4% by MS-MLPA. Since the MS-MLPA is based on the use of methylation-sensitive enzymes, incomplete digestion could account for the low confirmation rate.

Epigenetic mosaicism in *MLH1* is observed among constitutional epimutations (reviewed in [4]), in contrast to MMR genetic mosaicism that is very rare [18, 19]. To date, although several cases with low levels of epigenetic mosaicism in *MLH1* (≤10% methylation) have been reported [5, 9, 10, 11, 12], only 7 of them have been validated by other techniques showing highly variable clinical manifestation profiles (Table S4). The case identified here shows an aggressive phenotype associated with multiple tumors and an early age of onset.

The high expressivity of the disease contrasts with the subtle functional impact on *MLH1* expression, in accordance with the low methylation levels. The possibility of a CMMRD in case 29 was formally discarded (absence of germline*MLH1* mutations, biallelic *MLH1* transcription and conserved MLH1 protein expression in normal tissues). Furthermore,

no pathogenic alterations in other hereditary cancer genes were detected, although other genetic and/or environmental factors could be playing a role in the observed phenotype. In contrast, the loss of heterozygosity of the same *MLH1* allele in tumors from case 29 together with the subtle transcriptional silencing associated with the other *MLH1* allele (Figure S6) underlined the role of the identified epigenetic mosaicism in cancer predisposition.

In spite of an extensive search, we have not been able to identify any genetic alteration underlying the epimutated allele in case 29: copy number variations were not detected, the presence of promoter variants at C-D regions were ruled out by clonal sequencing, and no deleterious variants were found in *MLH1* by Sanger or next generation sequencing (in spite of mosaicism pipelines used). To date 9 families carrying secondary epimutation have been described (reviewed in [3, 4, 5]. In one case the epimutation was present in a low proportion of the alleles (<10%), associated with the silent variant c.27G>A (p.Arg9=) [5]. Although we cannot completely rule out that genetic alterations have been missed in case 29, the lack of family history is compatible with a *de novo* primary epimutation. Furthermore, the detection of homogeneous levels of *MLH1* methylation in tissues derived from the three embryonic layers suggested that the epimutation occurred during the epigenetic reprogramming of pre-implantation embryos, at any time between zygote and blastocyst stages[2].

It is clear that intensive surveillance for the early detection of metachronous gastrointestinal and gynecological tumors should be recommended to patient 29 [20]. In contrast, predictive epigenetic testing should be proposed to family members with extreme precaution. Unless stable inheritance of hypermethylation can be determined in descendants as previously shown in one mosaic epimutant [10], in the absence of a known causal mechanism and an established inheritance pattern for a *MLH1* epimutation, the interpretation for cancer risk in relatives became awkward.

In all, we have identified three *bona fide MLH1* epimutations (2 previously reported in Pineda, Mur et al and one in the present study) in out of 71 patients with *MLH1*-methylated CRC (4.2%) and in 3 out of 20 patients with early onset or multiple tumors

(15%) (Figure S1). In all, *MLH1* epimutations represent so far 1% of all LS cases in our series including the case identified by the use of highly sensitive techniques.

In summary, we have identified a *bona fide* low-level *MLH1* epigenetic mosaicism by using highly sensitive *MLH1* methylation analysis. In light of the obtained results, we strongly recommend the use of highly sensitive techniques for the screening of constitutional methylation. Special attention should be paid to patients diagnosed by early onset and/or multiple *MLH1*-methylated tumors.

REFERENCES

1 Lynch HT, Snyder CL, Shaw TG*, et al.* Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* 2015;15(3):181-94.

2 Hitchins MP. Constitutional epimutation as a mechanism for cancer causality and heritability? *Nat Rev Cancer* 2015;15(10):625-34.

3 Morak M, Ibisler A, Keller G*, et al.* 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. *J Med Genet* 2018;55(4):240-8.

4 Dámaso E, Castillejo A, Arias MM*, et al.* Primary constitutional MLH1 epimutations: a focal epigenetic event. *Br J of Cancer (in press)* 2018.

5 Leclerc J, Flament C, Lovecchio T*, et al.* Diversity of genetic events associated with MLH1 promoter methylation in Lynch syndrome families with heritable constitutional epimutation. *Genet Med* 2018.

6 Barrington DA, Champion ML, Boitano TKL*, et al.* Characteristics of African American women at high-risk for ovarian cancer in the southeast: Results from a Gynecologic Cancer Risk Assessment Clinic. *Gynecol Oncol* 2018.

7 Pinto D, Pinto C, Guerra J*, et al.* Contribution of MLH1 constitutional methylation for Lynch syndrome diagnosis in patients with tumor MLH1 downregulation. *Cancer Med* 2018;7(2):433-44.

8 Hitchins MP. The role of epigenetics in Lynch syndrome. *Fam Cancer* 2013;12(2):189-205.

9 Hitchins MP, Owens SE, Kwok CT*, et al.* Identification of new cases of early-onset colorectal cancer with an MLH1 epimutation in an ethnically diverse South African cohort. *Clin Genet* 2011;80(5):428-34.

10 Sloane MA, Nunez AC, Packham D*, et al.* Mosaic epigenetic inheritance as a cause of early-onset colorectal cancer. *JAMA Oncology* 2015;1(7):953-7.

11 Ward RL, Dobbins T, Lindor NM*, et al.* Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. *Genet Med* 2013;15(1):25-35.

12 Wu PY, Zhang Z, Wang JM*, et al.* Germline promoter hypermethylation of tumor suppressor genes in gastric cancer. *World J Gastroenterol* 2012;18(1):70-8.

13 Auclair J, Vaissiere T, Desseigne F*, et al.* Intensity-dependent constitutional MLH1 promoter methylation leads to early onset of colorectal cancer by affecting both alleles. *Genes Chromosomes Cancer* 2011;50(3):178-85.

14 Pineda M, Mur P, Iniesta MD*, et al.* MLH1 methylation screening is effective in identifying epimutation carriers. *Eur J Hum Genet* 2012;20(12):1256-64.

15 Castillejo A, Hernández-Illán E, Rodriguez-Soler M*, et al.* Prevalence of MLH1 constitutional epimutations as a cause of Lynch syndrome in unselected versus selected consecutive series of patients with colorectal cancer. *Journal of Medical Genetics* 2015;52(7):498-502.

16 Crucianelli F, Tricarico R, Turchetti D*, et al.* MLH1 constitutional and somatic methylation in patients with MLH1 negative tumors fulfilling the revised Bethesda criteria. *Epigenetics* 2014;9(10):1431-8.

17 Gausachs M, Mur P, Corral J*, et al.* MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. *Eur J Hum Genet* 2012;20(7):762-8.

18 Pastrello C, Fornasarig M, Pin E*, et al.* Somatic mosaicism in a patient with Lynch syndrome. *Am J Med Genet A* 2009;149A(2):212-5.

19 Sourrouille I, Coulet F, Lefevre JH*, et al.* Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. *Fam Cancer* 2012;12(1):27-33.

20 Giardiello FM, Allen JI, Axilbund JE*, et al.* 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(8):1159-79.

SUPPLEMENTARY MATERIAL

Supplementary methods

Patients and sample processing:

A total of 18 individuals presenting *MLH1*-methylated colorectal tumors before 50 years of age or multiple tumors before 60 years from a series of 132 CRC harboring loss of expression of MLH1/PMS2, being 71 of them *MLH1*-methylated, were included in the study (Figure S1; Table S1). Fourteen of the 18 included patients were previously reported [1] and the 4 remaining cases were subsequently identified (Figure S1; Table S1). Clinico-pathological data was collected, including age at cancer diagnosis, tumor location, and *MLH1* methylation status in tumor and blood DNA (Table S1). All of them fulfilled Bethesda criteria, being four of them diagnosed before 35 years of age and 6 of them presenting multiple tumors. Cases 7 and 21 displayed synchronic tumors. Cases 7 and 29 were positive for *MLH1* methylation in more than one of their tumors (Table S1).

Samples of peripheral blood leukocytes (PBL) DNA were collected from the biobank at the Catalan Institute of Oncology (ICO). For each FFPE specimen, 10-20 x 10-µm sections were cut from a single block using macrodissection with a scalpel if needed to enrich for tumor cells. After deparaffinization using Deparaffinization Solution (Qiagen, Hilden, Germany), DNA was isolated using the QIAmp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer's instructions. Skin fibroblasts were cultured as reported [1] and DNA extraction was carried out using Wizard Genomic DNA purification Kit. Buccal mucosa was obtained using Isohelix DNA buccal swabs and DNA extraction was done by phenol/chloroform extraction.

Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA)

The presence of methylation at the *MLH1* promoter was assessed using the SALSA MLPA ME011 Mismatch Repair genes probemix (MRC-Holland), according to the manufacturer’s instructions. The kit SALSA MLPA ME011 Mismatch Repair genes

probemix (MRC-Holland) includes five probe pairs in *MLH1* promoter (with the respective HhaI sites located at -659, -383, -246, -13 and +208 relative to the start codon; GenBank accession number U26559) that cover five independent regions: regions A to D of the promoter and intron 1 (Table S5). DNA from the RKO CRC cell line was used as the *MLH1* methylation-positive control. The amplification products were run on an ABI Prism 3130 DNA sequencer and analyzed using GeneMapper v4.0 (Applied Biosystems).

Methylation-specific melting curve analysis (MS-MCA):

One µg of DNA was treated with sodium bisulfite using the EZ DNA Methylation-Gold Kit (Zymo Research, Orange, CA, USA) under manufacturer's conditions. One µl of bisulfite treated DNA was used to amplify Deng's C and D regions in a nested PCR reaction using MegaMix double solution (Microzone Ltd., UK). Each promoter region was preamplified using external primers (Table S5). The nested PCR was carried out in a LightCycler 480 II (Roche) using 1µL of amplified *MLH1* promoter fragment, Light Cycler 480 SYBR Green I Master Kit (Roche) and internal primers (Table S5). Results were analyzed with the Light Cycler ®480 software version 1.5.1 [1].

Since MS-MLPA had showed an analytical sensitivity of 5-10% for *MLH1* methylation detection [2] in C/D promoter regions (previously associated with *MLH1* transcriptional silencing [3], MS-MCA was optimized for the highly sensitive analysis of methylation. Its analytical sensitivity for the detection of *MLH1* promoter methylation was assessed using serial dilutions of the RKO cell line (biallelic*MLH1* methylation) with unmethylated WGA (Whole Genome Amplification) DNA. WGA DNA and RKO DNA showed melting peaks at 73 ºC and 77ºC, respectively. Methylation levels in DNA samples from patients and controls were evaluated in combination with the reconstitutions within the same experiment. The technique demonstrated a sensitivity of 1% at C region and 10% at D region (Figure S1A-B).

Pyrosequencing:

One µl of bisulfite-converted DNA was used in a PCR reaction for the amplification of regions C and intron 1 of *MLH1* using Immolase^{TM} DNA polymerase kit (Bioline, UK)- and MegaMix double solution (Microzone Ltd., UK), respectively, and biotin-labeled primers (designed with MethPrimer; Table S5). Purification and subsequent processing of the biotinylated single-stranded DNA was performed according to the manufacter's recommendations at the PyroMark Q24 Vacuum PrepWorkstation (Qiagen). Pyrosequencing reaction was performed using each specific sequencing primer (Table S5) on a PyroMark Q24 pyrosequencer system using the Pyromark Gold Q26 Reagents kit. The sequences analyzed were GAGYGGATAGYGATTTTTAAYGYGTAAGYGTA for the promoter C region and YGATTTAAYGGGTYGYGTTATTTAATGGYGYGGATAYGT for intron 1. The methylation in targeted CpG sites was analyzed using the software PyroMark Q24 2.0.6. Each sample was run in triplicates. Methylation at each specific CpG was calculated as the mean of all triplicates and statistical analysis was performed using Graphpad Prism.

The limit of detection of *MLH1* methylation was calculated as the sum of the mean concentration of the blank (unmethylated DNA) and three times the standard deviation of the blank (S_b+3*SD_b), and assessed using the serial dilutions of the RKO cell line and unmethylated WGA DNA mentioned above. The technique demonstrated a sensitivity of 4% at C region and 5% at intron 1 (Figure S3A-B).

Clonal bisulfite sequencing:

Clonal bisulfite sequencing of *MLH1* promoter was used to determine the allelic methylation profile in bisulfite modified PBL DNA from the *MLH1* epimutation carrier. Primers were used to amplify the promoter region encompassing variant c.-93G>A (Table S5) and PCR products were cloned in *E. coli*cells using the pGEM-T easy vector system (Promega) to separate the amplicons into individual alleles [4]. The inserts from individual colonies were sequenced using vector primers. The methylation status at each individual CpG site was determined using SeqMan (DNASTAR). The number of

clones needed to acquire statistical power was calculated using the formula n= [z^2 x p x (1-p)]/e^2, where z is the mean value deviation, p is the expected proportion of methylated clones and e is the error range. Considering p as 0.03 (based on previous results), e as 0.02 and z as 1.96, the analysis of at least 280 clones was performed in order to achieve a confidence interval between 0,01–0,05.

Methylation array analysis:

Genome wide methylation profiling was analysed using Infinium Human Methylation 450K Beadchip as previously described [4]. One microgram of bisulfite treated DNA from blood and normal colorectal mucosa was hibridized in addition to in vitro methylated and unmethylated DNAs as internal quality control. 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 GenomeStudio software (Methylation Module). For each analyzed CpG site, a β-value was obtained depending on the florescence intensity. β measures took values between 0 (unmethylated) and 1 (fully methylated). Results were compared with the previously obtained from 41 healthy controls, 51 mutation-positive Lynch syndrome patients and 12 constitutional *MLH1* epimutation carriers [4].

Mutational analysis of hereditary cancer genes including *MLH1*:

Mutational analysis of hereditary cancer genes was performed by using two customized NGS panels: a panel of 26 CRC-associated genes [5] and a panel of 126 hereditary cancer genes (I2HCP v2.1) [6]. Identified variants were filtered against common single-nucleotide polymorphisms (MAF>1 according to ExAC and ESP databases). The presence of *MLH1* genetic variants at low proportion were further explored by applying mosaicism detection pipelines (alternative allele ratio ≥0.01).

Structural alterations in *MLH1* were tested by SALSA MLPA P003-B1 MLH1/MSH2 probemix (MRC-Holland) Mismatch Repair genes probemix (MRC-Holland), by a custom-designed high definition CGH array [4] and by a custom SNP array [7].

Statistical analysis

The obtained results were analyzed using the non-parametric Mann–Whitney U test for quantitative data. All reported P values are 2 sided, and $P<0.05$ was considered significant.

1 Pineda M, Mur P, Iniesta MD*, et al.* MLH1 methylation screening is effective in identifying epimutation carriers. *Eur J Hum Genet* 2012;**20**(12):1256-64.

2 Gausachs M, Mur P, Corral J*, et al.* MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. *Eur J Hum Genet* 2012;**20**(7):762-8.

3 Deng G, Chen A, Hong J*, et al.* Methylation of CpG in a Small Region of the hMLH1 Promoter Invariably Correlates with the Absence of Gene Expression. *Cancer Research* 1999;**59**(9):2029-3.

4 Dámaso E, Castillejo A, Arias MM*, et al.* Primary constitutional MLH1 epimutations: a focal epigenetic event. *Br J of Cancer (in press)* 2018.

5 Vargas-Parra GM, Gonzalez-Acosta M, Thompson BA*, et al.* Elucidating the molecular basis of MSH2-deficient tumors by combined germline and somatic analysis. *Int J Cancer* 2017;**141**(7):1365-80.

6 Castellanos E, Gel B, Rosas I*, et al.* A comprehensive custom panel design for routine hereditary cancer testing: preserving control, improving diagnostics and revealing a complex variation landscape. *Sci Rep* 2017;**7**:39348.

7 Castellsague J, Gel B, Fernandez-Rodriguez J*, et al.* Comprehensive establishment and characterization of orthoxenograft mouse models of malignant peripheral nerve sheath tumors for personalized medicine. *EMBO Mol Med* 2015;**7**(5):608-27.

Supplementary figures

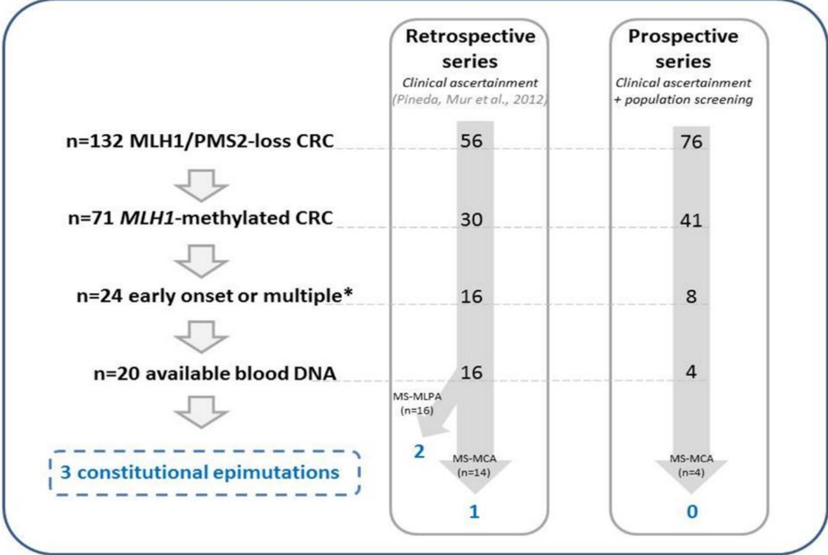


Figure S1. Schematic representation of the origin of the 18 cases included in this study. CRC: colorectal cancer.

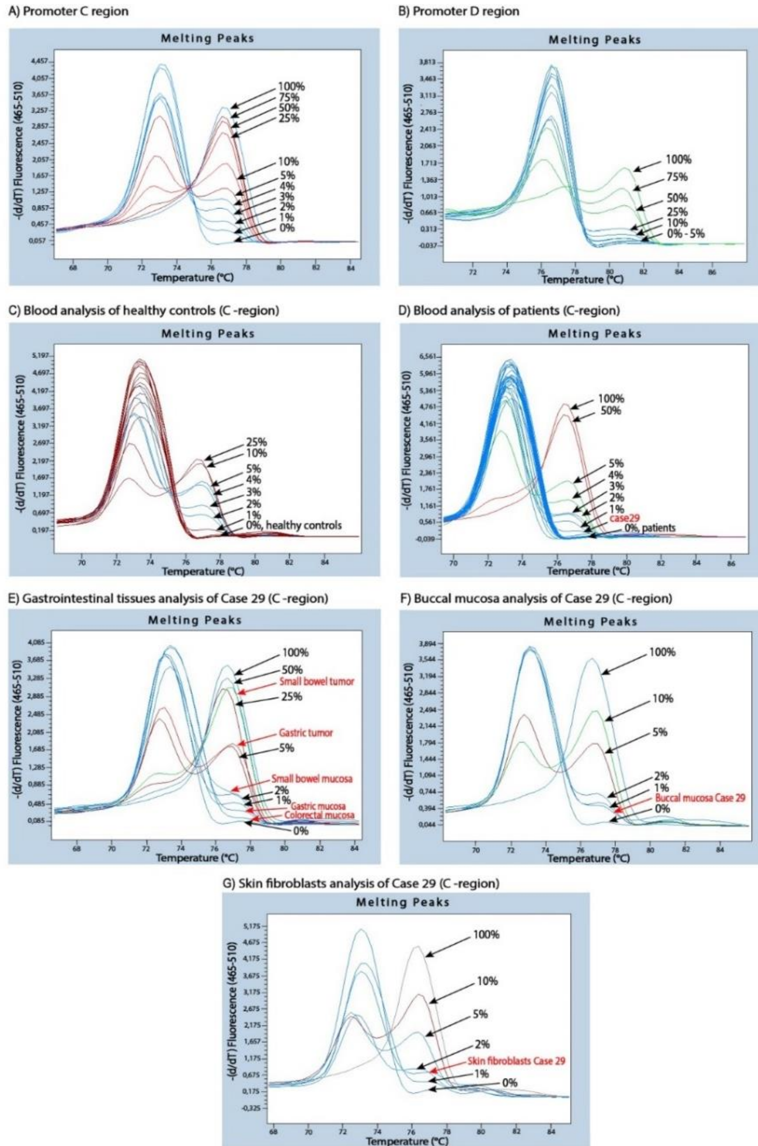


Figure S2. ***MLH1*** **promoter methylation analysis by Methylation-Specific Melting Curve Analysis (MS-MCA). A)** Analytical sensitivity of the promoter C region. The assay displays 1% of sensitivity. **B)** Analytical sensitivity of the promoter D region. The assay shows sensitivity around 10%. **C)** Methylation analysis from 10 healthy controls in the promoter C region. All of them show the same curve pattern than the reconstitution 0%, indicating lack of methylation. **D)** Methylation analysis in the series of patients in the promoter C region. Only the case 29 displays a peak with low levels of methylation (around 1%). **E)** Methylation analysis in tumor and normal gastrointestinal tissues of case 29. **F)** Methylation analysis in buccal mucosa of case 29. **G)** Methylation analysis in skin fibroblasts of case 29.

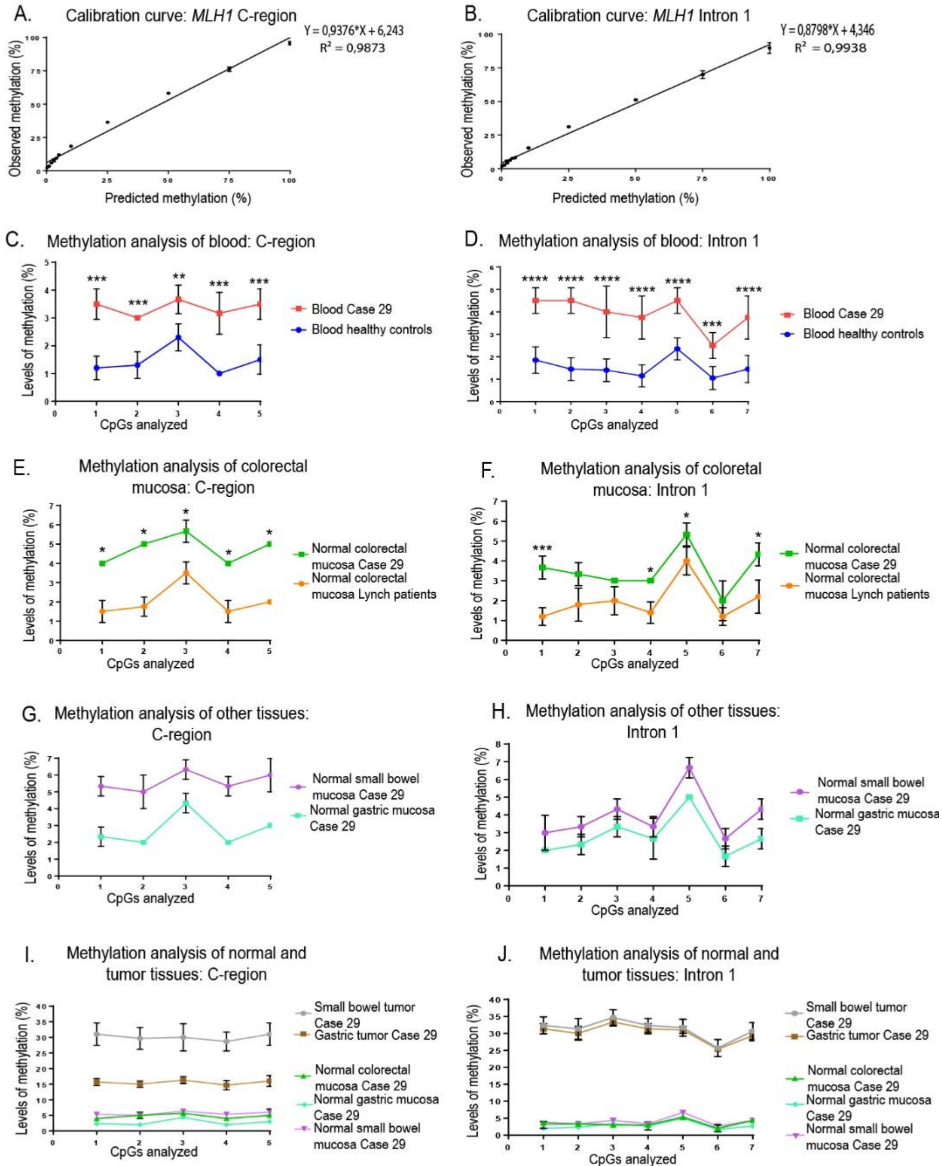


Figure S3. *MLH1* methylation analysis of the promoter C-region and intron 1 by pyrosequencing.A) Analytical sensitivity analysis for the detection of methylation in *MLH1* C-region and intron 1. The detection limits for both regions are 4% and 5% respectively, enabling the detection of positive samples as from these two values. However, the graphs become lineal from 10% of methylation, allowing the quantification of the samples from this value. **B)** Analytical sensitivity analysis for the detection of methylation in intron 1 of *MLH1*. **C)** Methylation analysis in blood from case 29 and healthy controls (n=10) for *MLH*1 C-region. **D)** Methylation analysis in blood from case 29 and healthy controls (n=20) for the intron 1 of *MLH1.* **E)** Methylation analysis in normal colorectal mucosa from case 29 and Lynch patients (n=4) for *MLH1* C-region. **F)** Methylation analysis in normal colorectal mucosa from case 29 and

Lynch patients (n=4) for intron 1 *MLH1*. **G)** Methylation analysis in normal small bowel mucosa and gastric mucosa from case 29 for *MLH1* C-region. **H)** Methylation analysis in normal small bowel mucosa and gastric mucosa from case 29 for the intron 1 of *MLH1*. **I)** Methylation analysis in normal and tumoral gastrointestinal tissues in case 29 for *MLH1* C-region. **J)** Methylation analysis in normal and tumoral gastrointestinal tissues in case 29 for the intron 1 of *MLH1*.

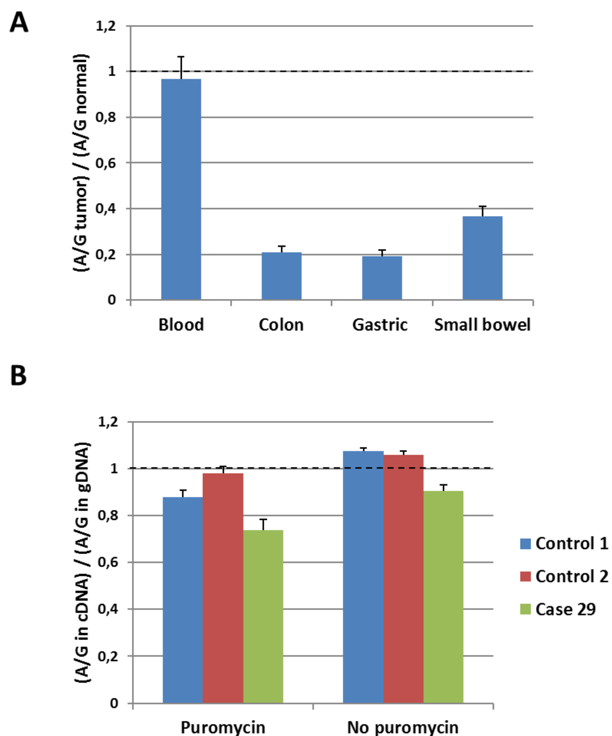


Figure S4. Analysis of the expression of *MLH1* alleles at rs1799977 (c.655A>G). A) Loss of heterozygosis (LOH) analysis in tumors from case 29. Allele specific expression (ASE) values show loss of c.655A allele in the 3 tumors analyzed. **B)** SNuPE analysis at *MLH1* rs1799977 in cDNA derived from case 29 and two heterozygous controls. A slightly diminished expression of the c.655G allele was observed.

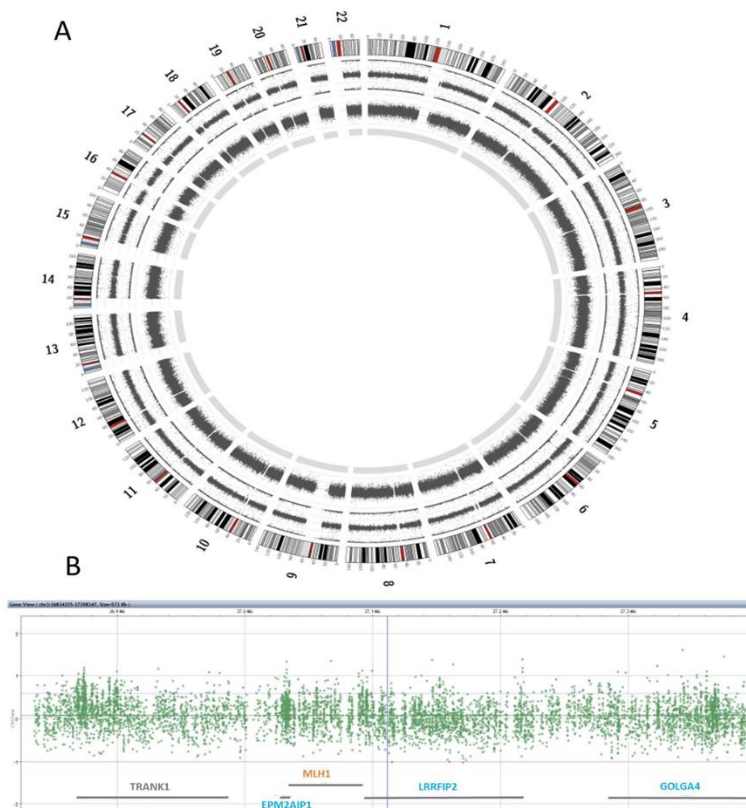


Figure S5. Analysis of structural aberrations in case 29. A) Genome-wide SNP array profiling from case 29 is shown as Circos plots. Circos plot was divided in 3 concentric circles. Chromosomes are represented at the external circle. Their centromeres are painted in red. Middle circle tracks the LOH. External allelic peaks mark homozygous SNPs and internal allelic peaks heterozygous ones. Two middle lines are indicative of LOH, but patient 29 did not show signs of LOH. Internal circle tracks log2 copy number lane. Middle points indicate diploid genomic material; upper points, gains of genomic material and lower points, losses. Patient 29 displayed diploid pattern throughout her genome. **B)** CNV analysis in *MLH1* region of patient 29 by custom CGH array. Genes located at the analyzed region are represented at the bottom of the figure. Probes are displayed as green dots in a log2 graph. Gains and losses of genetic material are considered when more of 5 consecutive probes reach 2 or -2 values respectively. No CNV abnormalities were identified.

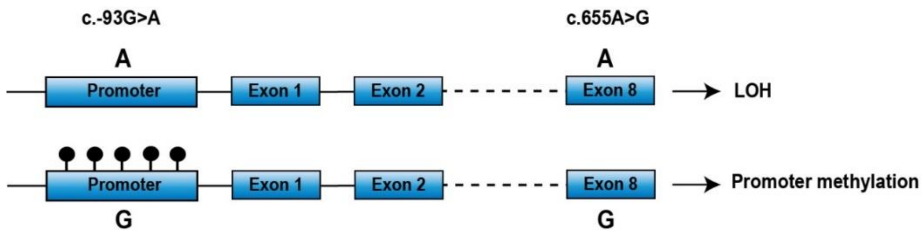


Figure S6. Deduced *MLH1* haplotype in case 29. Promoter methylation is associated with

c.-93G allele, probably located in cis with exonic c.655G allele, which showed subtle transcriptional silencing in blood. In concordance, the *MLH1* c.655A allele showed LOH in the 3 proband tumors, indicating inactivation of the wild type (non-methylated) allele.

Supplementary tables

Table S1 Clinical and molecular features of patients with *MLH1* methylated tumors

ID patient	Gender	Clinical Criteria	Blood: *MLH1* methylation assessed by MS-MLPA (%)			Tumor age of onset	Tumor: *MLH1* methylation assessed by MS-MLPA (%)			Other tumors
			C region	D region	Intron 1		C region	D region	Intron 1	
2^{R}	F	BC	0	0	0	CRC (49)	24,9	36,9	40,3	
3^{R}	M	BC	0	0	0	CRC (37)	29,3	31,7	106	
5^{R}	M	BC	0	0	0	CRC (50)	28,6	33,6	49	TC (49)
7^{R}	M	BC	0	0	0	CRC (42)	24,1	25,2	29,3	CRC^{met}(synch)
8^{R}	M	BC	0	0	0	CRC (29)	25,1	27,6	28,5	
9^{R}	F	BC	0	0	0	CRC (47)	38,5	34,9	39,7	
16^{R}	F	BC	2	0	0	CRC (24)	57,5	75,1	53,6	
18^{R}	M	BC	0	0	0	CRC (48)	32,8	34,8	43,7	
21^{R}	F	BC	0	0	0	CRC (58)	40,6	66,6	74,3	3 CRC (synch)
23^{R}	F	BC	0	0	0	CRC (47)	20,3	39,3	39,9	
24^{R}	F	BC	0	0	0	CRC (59)	11,4	20,6	37,4	CRC (29)
27^{R}	M	BC	0	0	0	CRC (47)	40,1	21,6	64,0	
28^{R}	M	BC	2	0	0	CRC (31)	32	33	35,6	
29^{R}	F	BC	3,6	2,9	3	CRC (22)	76,9	101,6	66,0	SB^{met}(24), GC^{met}(25)
30^{P}	F	BC	0	0	0	CRC (39)	71	67	76,0	
31^{P}	F	BC	0	0	0	CRC (47)	43	39	54,4	
32^{P}	F	BC	3	0	0	CRC (47)	34	23	37,2	
33^{P}	F	BC	2,4	0	0	CRC (63)	49	41	55,1	CRC (41)

Abbreviations: BC, Bethesda Criteria; M, male; F, female; CRC, colorectal cancer; TC, testicular cancer; SB, small bowel cancer; GC, gastric cancer; synch, synchronous. R: retrospective series, patient ID from the original publication Pineda, Mur 2012); P: prospectively collected series.

Table S2. *MLH1* methylation assessed by MS-MLPA in samples from patient 29

Sample	Analyzed region (% of methylation)				
	A	B	C	D	Intron 1
Lymphocytes	6	5	3	3	3
Colorectal mucosa	19	6	6	10	14
Colorectal tumor	60	109	80	50	66
Small bowel mucosa	18	3	2	5	14
Small bowel tumor	45	73	33	37	44
Gastric mucosa	14	2	0	3	10
Gastric tumor	68	93	63	73	69

Table S3. Reported patients with *MLH1* epigenetic mosaicism at low proportion (≤10%)

Patient ID from the original publication	Cancer type (age of onset)	Methylation levels in blood	Technique	Confirmation by other tecniques	Analyzed in other tissues	Reference
29	CRC (22), SBC (24), GC (25)	1-2%	MS-MCA	Clonal BS, Infinium 450k	BM, SF, CRM, SBM, GM	Pineda, Mur et al., 2012 and current study
P	CRC (69)	2%	qMSP	Pyroseq, COBRA, Clonal BS		Ward et al. 2013
Q	CRC (39), skin (44)	3%	qMSP	Pyroseq, COBRA, Clonal BS		Ward et al. 2013
N	CRC (50, 52, 54)	3%	qMSP	Pyroseq, COBRA, Clonal BS		Ward et al. 2013
O	CRC (45)	4%	qMSP	Pyroseq, COBRA, Clonal BS		Ward et al. 2013
II-3	Asymptomatic (?)	4%	Pyroseq.	Clonal BS	saliva, hair	Sloane et al. 2015
84-I	CRC (35)	8%	qMSP	Pyroseq, Clonal BS		Hitchins et al. 2011
G46	GC (60)	10%	Bisulfite seq	Clonal BS		Wu et al. 2012

Abbreviations: CRC, colorectal cancer; SBC, small bowel cancer; GC, gastric cancer; Pyroseq, pyrosequencing; Bisulfite seq, bisulfite sequencing; Clonal BS, clonal bisulfite sequencing; BM, buccal mucosa; SF, skin fibroblasts; CRM, colorectal mucosa; SBM, small bowel; GM, gastric mucosa

Table S4. Variants identified in the mutational analysis of hereditary cancer genes in case 29. NR: not reported; NA: not available.

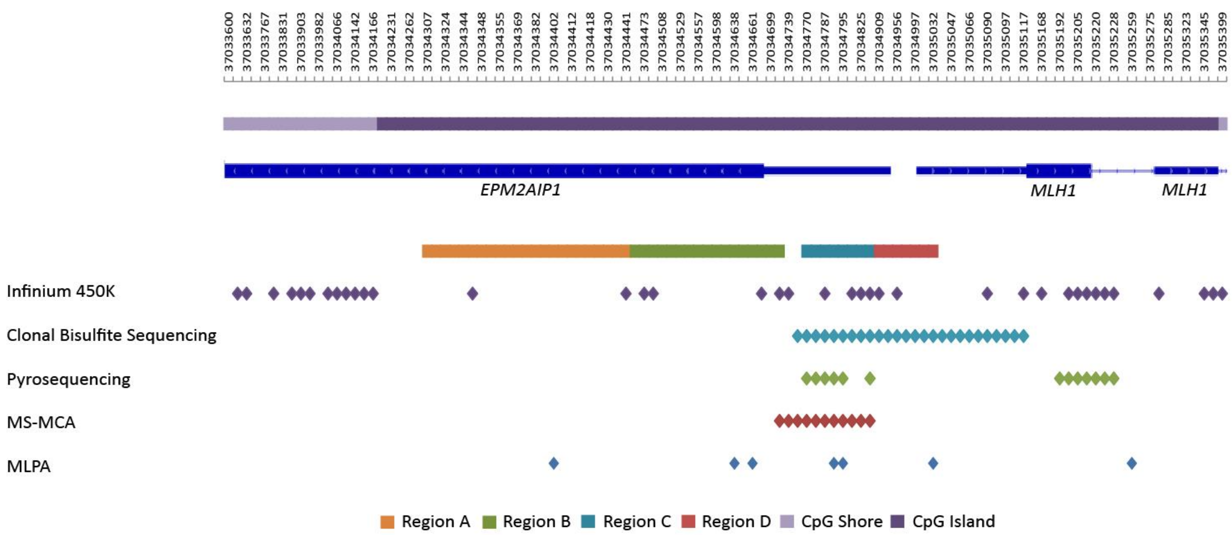
Variant calling				Coverage		rs ID	MAF	In silico predictions					ClinVar Classification
								Splicing	Protein function				
Gene	Transcript	cDNA change	Predicted protein change	Allele Frequency (%)	Read Depth		genomAD (%)		SIFT	Mutation Taster	Polyphen2 /HumDiv	Polyphen2 /HumVar	
XRCC2	NM_005431.1	c.115G>A	p.V39M	45,19	104	rs730882040	0,01	No changes	Deleterious	Disease causing	Possibly damaging	Possibly damaging	Uncertain significance**
NBN	NM_002485.4	c.38-20T>A	p.?	57,74	386	rs774709748	0,0037	No changes	NA	NA	NA	NA	Likely benign*

Table S5. Primers used in the study.

A. Primers and conditions

Gene	Analysis	Primer name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Ta (°C)
MLH1	MS-MCA	MLH1C_PCR_ext	TATTTTTGTTTTTATTGGTTGG	CCAATCAAATTTCTCAACTCTATA	50
		MLH1C_PCR_int	TGTTTTTATTGGTTGGATATTT	CCAATCAAATTTCTCAACTCTATA	50
		MLH1D_PCR_ext	AGGTATTGAGGTGATTGGTTG	CAATTCTCAATCATCTCTTTAATAACA	50
		MLH1D_PCR_int	GGTGATTGGTTGAAGGTATTTT	ATCATCTCTTTAATAACATTAACTAACC	50
	Pyrosequencing	Piro_PCR_MLH1_C	TTTYGGTATTTTTGTTTTTATTGGTTGG	[Btn]AAACAACTTAAATACCAATCAAATTTCTCAAC	58
		Piro_Seq_MLH1_C	TAAAAAYGAATTAATAGGAA	-	-
		Piro_PCR_MLH1_intron1	ATTTAGYGGTTAGTTAATGTTATTAAAGAGATG	[Btn]TAAACATACRCTATACATACCTCTACCC	56
		Piro_Seq_MLH1_intron1	AGTYGGGTTTATTTAAGGGTTA	-	-
	Clonal Bisulfite sequencing	PCR_MLH1_CD	TTTTAAAAAYGAATTAATAGGAAGAG	CAATTCTCAATCATCTCTTTAATAA	50
	ASE (SNuPe)	rs179997_PCR_cDNA	CACAATGCAGGCATTAGTTTCTC	AGGTACAGGAATGGGTGTGTG	59
		rs179997_PCR_gDNA	GTTTCAGTCTCAGCCATGAG	ACACATGATTCACGCCACAG	55
		rs179997_snupe		TTCTCGACTAACAGCATTTCCAAAGA	50

B. Localization probes and regions analyzed in the study of *MLH1* methylation. The *EMP2AIP1-MLH1* CpG island (colored in dark purple) encompass the *MLH1* promoter and intron 1



DISCUSIÓN

El conocimiento de las causas moleculares de la predisposición cáncer colorrectal es importante para la identificación de individuos con predisposición a este tipo de neoplasia, antes de que desarrollen un tumor o, alternativamente, en etapas tempranas, y permite aplicar las estrategias de vigilancia apropiadas con el objetivo de disminuir la morbilidad y mortalidad de estos individuos y sus familiares. Dentro del grupo de pacientes con tumores colorrectales con deficiencia reparadora, el riesgo de cáncer colorrectal en probandos y familiares, así como las estrategias de seguimiento, están bien establecidas para los pacientes con síndrome de Lynch y cáncer colorrectal esporádico por metilación somática de *MLH1* (Figura 28). En cambio, no es así para el grupo de pacientes Lynch-like y epimutaciones constitucionales en *MLH1*, principalmente debido a que se desconoce su base molecular.

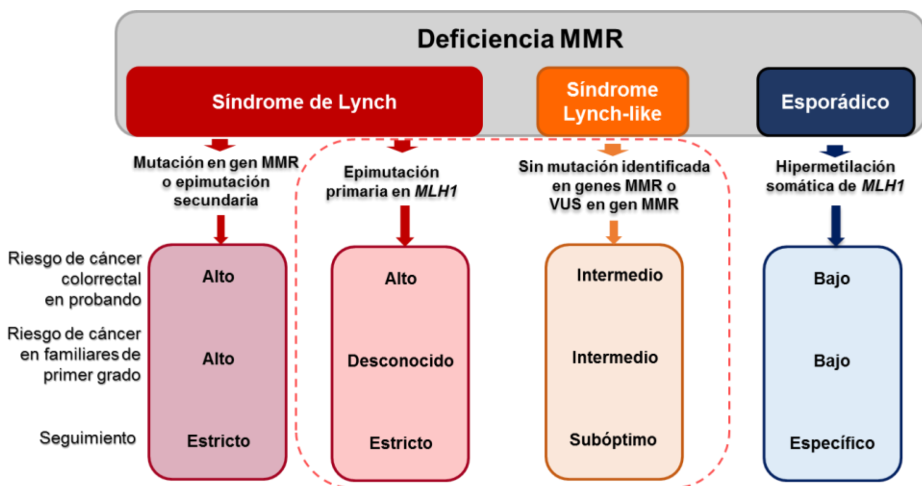


Figura 28: Clasificación de pacientes con deficiencia reparadora y estimación de su riesgo asociado. Con una línea discontinua se señalan los grupos de estudio del presente Trabajo. Modificado de Buchanan et al 2014.

Esta tesis se ha centrado en la identificación y caracterización de epimutaciones constitucionales en individuos con síndrome de Lynch y síndrome de Lynch-like. Dicha caracterización epigenética ha sido realizada en paralelo a la caracterización genética de los mismos casos, útil para el refinamiento de casos Lynch-like y relevante para la identificación de causas genéticas y patrones de herencia asociados a las epimutaciones constitucionales en *MLH1.*

1. El papel de las alteraciones epigenéticas constitucionales en el síndrome Lynch-like

1.1. Refinamiento genético de candidatos

Nuestro trabajo se ha centrado en la caracterización de una serie de 121 individuos Lynch-like, proveniente de una serie previa de 123 casos donde se habían identificado 2 individuos portadores de mutaciones bialélicas en *MUTYH* (Castillejo *et al*, 2014). Previo a la caracterización del metiloma, que tenía el objetivo de buscar causas epigenéticas subyacentes en estos pacientes, se desarrolló una caracterización genética exhaustiva en aquellos que presentaban una fuerte historia individual o familiar de cáncer, que incluía el reanálisis de genes MMR (incluyendo el estudio de promotores), el reanálisis de reordenamientos y la caracterización funcional de variantes en genes reparadores. Esto se completó con el análisis mutacional de un panel de genes de predisposición a cáncer colorrectal.

El papel de los genes reparadores en el síndrome Lynch-like

Identificación de mutaciones genéticas previamente no identificadas en genes MMR

El reanálisis de los genes MMR en una serie de 62 individuos (Vargas-Parra *et al*, 2017-Artículo 1, Dámaso *et al*, manuscrito en preparación-Artículo 2) mediante NGS ha identificado 3 mutaciones en genes MMR previamente no identificadas (3/62; 4.8%): *MSH2* c.211G>C (mutación que afecta al procesamiento del RNA), *MLH1* c.676C>T (p.Arg226*) y *MSH6* c.2219T>A (p.Leu740*). Estos 3 casos representan el 2.5% de la serie (3/121). Asimismo, este análisis ha permitido la identificación de variantes de significado desconocido en genes MMR previamente no identificadas. Finalmente, el reanálisis de número de copias mediante MLPA en un subgrupo de casos deficientes en *MSH2* ha permitido la identificación de una duplicación en el exón 8 de *MSH2* previamente no identificada. En cambio, no se ha detectado ningún portador de la inversión recurrente en *MSH2* (Wagner *et al*, 2002).

En los 121 casos Lynch-like incluidos en nuestra serie, los genes MMR candidatos habían sido analizados previamente según el patrón de IHQ de los tumores. Los resultados obtenidos en el reanálisis de genes reparadores mediante NGS han evidenciado que, en

dos casos, las mutaciones no fueron detectadas por limitaciones metodológicas de las técnicas previamente utilizadas: secuenciación Sanger, debido a la localización de los cebadores o electroforesis en gel con gradiente de desnaturalización, debido a una menor sensibilidad analítica.

Además, los resultados obtenidos evidencian en algunos casos la falta de correlación entre la interpretación de la inmunohistoquímica de las proteínas reparadoras y el gen MMR mutado. A pesar de la elevada sensibilidad (>80%) y especificidad (≈80-92%) que ofrece esta técnica para la identificación de individuos con síndrome de Lynch, su interpretación no está exenta de problemas (Snowsill *et al*, 2017). Así pues, problemas técnicos durante la tinción de las laminillas pueden dar lugar a pérdidas artefactuales de proteínas reparadoras y un exceso de anticuerpo, a una elevada señal de fondo que puede suponer una identificación también artefactual de la proteína reparadora (Markow *et al*, 2017). Además, se han descrito patrones de IHQ poco comunes en tumores de pacientes con síndrome de Lynch, como tinciones conservadas en pacientes con mutaciones patogénicas que provocan una pérdida de función de la proteína sin degradación o patrones de pérdidas de expresión de proteínas diferentes de las habituales debido a la combinación de mutaciones germinales con mutaciones o hipermetilaciones somáticas en el tumor (Graham *et al*, 2015; Hagen *et al*, 2011; Markow *et al*, 2017; Morak *et al*, 2017; Niu *et al*; Pineda *et al*, 2015; Shia *et al*, 2012). Es importante destacar que el estudio de IHQ se recomienda en tumores (Kalady *et al*, 2015; Walsh *et al*, 2012; Yurgelun *et al*, 2012), puesto que el 44% de los adenomas de pacientes con síndrome de Lynch no presentan pérdida de expresión de la proteína afecta, debido a la falta de adquisición de la segunda mutación somática en el gen (Halvarsson *et al*, 2005).

Los resultados obtenidos justifican el reanálisis de genes MMR si se han realizado hace tiempo con técnicas ahora consideradas obsoletas, e independiente del patrón de IHQ en casos Lynch-like altamente sugestivos, especialmente en aquellos con elevada agregación familiar o presencia de varios tumores que muestren deficiencia reparadora.

La evaluación de la patogenicidad de variantes de significado desconocido en genes MMR permite su clasificación

El estudio mutacional de genes reparadores ha identificado un total de 19 variantes de significado desconocido en los genes MMR presentes en 22 de los 121 pacientes (procedentes del estudio previo de genes MMR candidatos o del reanálisis genético posterior). La evaluación de la patogenicidad y revisión de las evidencias ha permitido la reclasificación de 8 de estas variantes a mutaciones patogénicas responsables de síndrome de Lynch en 10 pacientes.

El estudio de RNA ha permitido la clasificación de dos duplicaciones en el gen *MSH2,* debido a la identificación de tránscritos aberrantes que generarían proteínas truncadas. En cambio, la identificación de tránscritos aberrantes no ha sido concluyente en otros casos en los que el impacto en la proteína no es claro. Además, el estudio de RNA también ha permitido la reclasificación a mutación patogénica de la variante probablemente patogénica en *MHS2* c.211G>C, identificada mediante nuestro panel NGS.

El análisis de probabilidad multifactorial, basado en la recopilación de datos clínico-patológicos (inestabilidad de microsatélites en cáncer colorrectal y cosegregación) ha permitido la reclasificación de 5 variantes. Estos modelos podrían mejorarse con la inclusión de la información de los patrones de IHQ de tumores colorrectales o la inclusión de los datos de tumores extracolónicos del espectro del síndrome de Lynch.

Globalmente, la evaluación de la patogenicidad de las variantes, combinando el estudio del RNA y el análisis de probabilidad multifactorial ha permitido la reclasificación a mutación patogénica del 42% (8/19) de las variantes de significado desconocido identificadas en nuestra serie. Cabe destacar que estos resultados han permitido el diagnóstico de síndrome de Lynch en 15 familias, 10 incluidas en nuestra serie LLS y 5 familias adicionales a las cuales se les identificó alguna de las variantes estudiadas. En conjunto el estudio de la evaluación de la patogenicidad de las variantes de significado desconocido ha supuesto el análisis con mayor rendimiento en el refinamiento de pacientes Lynch-like (10/121; 8%) (Figura 29). Este rendimiento podría incluso mejorarse si todas las variantes identificadas fueran evaluadas de manera exhaustiva.

El análisis mutacional de regiones promotoras de genes MMR ha demostrado previamente ser útil en la identificación de pacientes con síndrome de Lynch (Liu *et al*, 2016b). En nuestra serie el análisis de promotores identificó 2 variantes c.-574T>A de *MLH1* y c.[-25C>T(;)-204C>G] de *MSH6*. Desafortunadamente su impacto funcional no ha podido ser evaluado debido a la ausencia de polimorfismos exónicos en heterocigosis o debido a la falta de muestras biológicas necesarias para su análisis.

Los resultados obtenidos resaltan el beneficio de aplicar análisis cuantitativos y cualitativos para la interpretación y clasificación de variantes, así como la utilidad de recolectar muestras de RNA e incluir su análisis en la rutina diagnóstica.

El papel de las mutaciones genéticas en otros genes de predisposición a cáncer colorrectal

En la célula coexisten números mecanismos de reparación del DNA que son en muchas ocasiones redundantes, puesto que esta requiere de la acción coordinada de muchos componentes. Así, los componentes de la vía MMR pueden cooperar con proteínas implicadas en otros mecanismos de reparación del DNA, de modo que alteraciones en estos genes también podrían ser la causa subyacente de la deficiencia reparadora del DNA. Asimismo, la diversidad de manifestaciones clínicas de los diferentes síndromes de predisposición a cáncer colorrectal crea a menudo fenotipos solapantes que dificultan el diagnóstico de algunos pacientes. Por estos motivos, el refinamiento de candidatos LLS incluyó el análisis mutacional de 26 genes de predisposición a cáncer colorrectal, diseñado en nuestro grupo previo al inicio de esta tesis.

La presencia de mutaciones bialélicas en *MUTYH* en la serie de pacientes Lynch-like había sido previamente analizada, permitiendo la identificación de 2 individuos portadores (2/123; 1.6%), además de la identificación de 5 casos adicionales de la serie procedente de la Comunidad Valenciana, suponiendo el 3.1% de los individuos Lynch-like (7/225) (Castillejo *et al*, 2014) están en la misma línea que otros estudios publicados (Colebatch *et al*, 2006; Lefevre *et al*, 2011; Morak *et al*, 2014). El motivo es que la deficiencia en la reparación mediada por *MUTYH* puede causar mutaciones somáticas en los genes de MMR, simulando fenotípicamente tumores procedentes de pacientes con síndrome de Lynch (Castillejo *et al*, 2014; Colebatch *et al*, 2006; Lefevre *et al*, 2011; Morak *et al*, 2014).

Del mismo modo, se han identificado mutaciones germinales en *POLE/POLD1* en algunos casos con deficiencia MMR en tumores debido secundaria a la hipermutabilidad (Elsayed *et al*, 2014; Jansen *et al*, 2016; Yoshida *et al*, 2010). En nuestra serie no hemos identificado variantes patogénicas germinales en *POLE/POLD1*. Además los resultados de rango mutacional y firmas tumorales no indicaron que la única variante localizada dentro del dominio exonucleasa, *POLE* c.898A>G (p.Ile300Val), identificada en uno de los casos fuera patogénica.

Además de *MUTYH* y *POLE*, el gen *MCM9* había sido sugerido como gen candidato a para el síndrome Lynch-like (Goldberg *et al*, 2015). En nuestro trabajo el análisis de un panel de genes ha permitido la identificación de variantes potencialmente patogénicas en otros genes: *BUB1, SETD2* y *FAN1* en el primer estudio (Vargas-Parra *et al*, 2017-Artículo 1) *y EXO1*, *MSH3*, *APC*, *POLD1* en el posterior (Dámaso et al, manuscrito en preparación-Artículo 2).

Algunas variantes de los genes *EXO1* y *MSH3*, que codifican para proteínas directamente implicadas en la reparación de desapareamientos, habían sido identificadas en familias con sospecha clínica, aunque su relevancia clínica no fue determinada (Jagmohan-Changur *et al*, 2003; Peltomäki, 2001). Además, variantes en *MSH3* se han identificado en combinación con variantes de genes asociados a síndrome de Lynch (Duraturo *et al*, 2011; Morak *et al*, 2017). Recientemente se han identificado mutaciones bialélicas en *MSH3* en individuos con poliposis y cáncer colorrectal (Adam *et al*, 2016). En nuestra cohorte, 4 individuos han presentado variantes probablemente patogénicas en *EXO1* y *MSH3*, uno de los cuales presenta EMAST en su tumor. Estos hallazgos sugieren la posibilidad de un efecto oligogénico de las variantes en estos genes y su potencial contribución al desarrollo de los tumores. Sin embargo, son necesarios más estudios para confirmar esta posibilidad.

El gen *BUB1* codifica para un componente del punto control del huso responsable de la segregación cromosómica, y había sido descrito mutado en pacientes con diagnóstico temprano de cáncer colorrectal o con antecedentes familiares de esta patología (de Voer *et al*, 2013; Hanks *et al*, 2004).

SETD2 es una H3K36 trimetiltransferasa necesaria para reclutar el heterodímero MSH2-MSH6 durante la reparación del DNA. Su pérdida de función provoca inestabilidad de microsatélites y una elevada tasa de mutación (Li *et al*, 2013). Este gen está mutado en tumores gástricos inestables (Boussioutas *et al*, 2003) y, frecuentemente, en tumores colorrectales con inestabilidad de microsatélites (Choi *et al*, 2014). El hallazgo de 3 portadores de variantes predichas patogénicas a nivel germinal, sugiere que *SETD2* podría tener un papel importante en la patología Lynch-like.

FAN1 codifica para una nucleasa involucrada en la reparación de las roturas de DNA de doble cadena. Mutaciones monoalélicas en este gen se han asociado a cáncer colorrectal con estabilidad de microsatélites y cáncer de páncreas hereditario (Seguí *et al*, 2015; Smith *et al*, 2016). La identificación en nuestra serie de 2 individuos con variantes en *FAN1* sugería que la deficiencia de FAN1 podría afectar a la actividad MMR. Sin embargo, la frecuencia de variantes en casos y controles, la presencia de individuos homocigotos en población control de algunas variantes reportadas como patogénicas y la falta de cosegregación en algunos casos, ponen en duda el papel de *FAN1* como gen de predisposición hereditaria al cáncer (Broderick *et al*, 2017).

En conclusión, el análisis de paneles de genes permite la identificación nuevos genes candidatos responsables de la deficiencia reparadora observada en los tumores de los pacientes Lynch-like. Sin embargo, se necesitan análisis funcionales y estudios de cosegregación que validen el papel causal de estas variantes en la predisposición a cáncer.

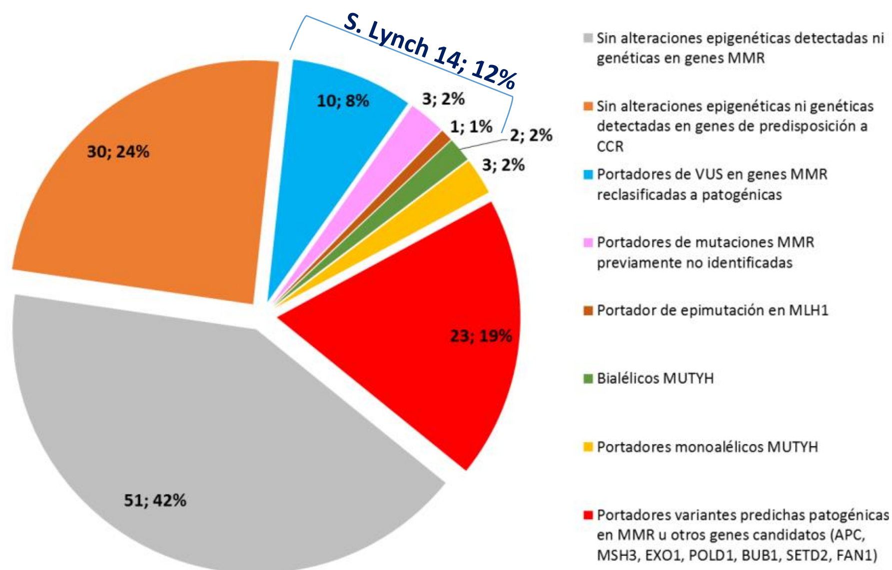


Figura 29: Resultados de la caracterización de la serie de individuos Lynch-like (n=123). Con una horquilla se señalan los grupos de pacientes Lynch-like que fueron reclasificados a síndrome de Lynch.

1.2-Identificación de epimutaciones constitucionales en *MLH1*

Las epimutaciones constitucionales en *MLH1* presentan una prevalencia de alrededor del 1% en series de pacientes sin mutación germinal en genes MMR (Crépin *et al*, 2012; Hitchins *et al*, 2005; Hitchins, 2013; Hitchins *et al*, 2011; Miyakura *et al*, 2004). Sorprendentemente, hemos identificado un nuevo caso de epimutación en *MLH1* en el que el estudio previo de metilación de *MLH1* en tumor había resultado no-informativo. Así, a pesar de que todos los pacientes incluidos en la serie Lynch-like se habían cribado mediante el estudio de *BRAF* y/o metilación somática en *MLH1*, algunos tumores dieron resultados no informativos en su estudio de metilación, posiblemente debido a la mala calidad del DNA extraído de la pieza parafinada. Posteriormente el estudio mutacional germinal no detectó ninguna alteración en estos pacientes que quedaron clasificados como individuos Lynch-like. Este hallazgo apunta a la necesidad de descartar la presencia de metilación constitucional en *MLH1* en casos sugestivos de epimutación germinal (edad temprana al diagnóstico de independientemente de la historia familiar

de cáncer). La caracterización del nuevo caso identificado se discutirá en el apartado 2 de la presente discusión.

El estudio mutacional de *BRAF* y/o de hipermetilación somática en *MLH1* está indicado para el seleccionar aquellos candidatos para el estudio mutacional germinal de *MLH1* en aquellos pacientes que presentan pérdida de expresión de MLH1 en su tumor. Es importante destacar que, aunque estas pruebas permiten diferenciar entre pacientes con síndrome de Lynch y pacientes con cáncer colorrectal esporádico con elevada sensibilidad y especificidad (Gausachs *et al*, 2012; Pérez-Carbonell *et al*, 2010), existe la posibilidad de que los pacientes con epimutaciones constitucionales en *MLH1* sean erróneamente clasificados como pacientes con cáncer colorrectal esporádico. De hecho se ha estimado que hasta el 14% de los individuos con tumores metilados en *MLH1* con una edad de diagnóstico menor de 50 años podrían ser secundarios a epimutaciones constitucionales (Hitchins, 2013).

1.3- Las alteraciones constitucionales epigenéticas fuera de los genes reparadores no parecen jugar un papel en el síndrome Lynch-like

A día de hoy sólo se han descrito epimutaciones constitucionales en los genes *MLH1, MSH2* y *PTPRJ,* relacionadas con la predisposición a cáncer colorrectal. Mientras que las epimutaciones en *MLH1* pueden ser debido tanto a metilación *de novo* como secundarias a diferentes variantes genéticas, las epimutaciones en *MSH2* y *PTPRJ* de deben a la variación en el número de copias de sus regiones reguladoras (deleción en el caso de *MSH2* y duplicación en el caso de *PTPRJ*) (Dymerska *et al*, 2017; Kempers *et al*, 2011; Ligtenberg *et al*, 2013; Venkatachalam *et al*, 2010). Sin embargo, las epimutaciones no son exclusivas del ámbito de la predisposición a tumores gastrointestinales. Así, se han descrito epimutaciones en los genes *BRCA1* y *RAD51C* en pacientes con cáncer de ovario familiar y en casos precoces de cáncer de mama esporádico (Hansmann *et al*, 2012). Se considera que una proporción significativa de las familias con síndrome de Cowden se deben a la epimutación de una región previa a *PTEN* que provoca el silenciamiento transcripcional de *KILLIN* (Bennett *et al*, 2010; Ngeow *et al*, 2011). Además, se han identificado metilaciones constitucionales de manera ocasional en otros contextos: la epimutación de *DAPK1* en una familia con leucemia linfocítica crónica (Raval *et al*, 2007), la epimutación de *RB1* se ha identificado en una familia con alta incidencia de retinoblastoma (Quiñonez-Silva *et al*, 2016) y, recientemente, se han descrito tres casos independientes de deficiencia de cobalamina por la epimutación del gen *MMACHC* (Guéant *et al*, 2018). Todos estos ejemplos hacen referencia a epimutaciones en genes ellas fuera de regiones susceptibles de *imprinting*.

Aunque estas observaciones avalan el papel de la epigenética en la predisposición al cáncer, no hemos identificado anomalías epigenéticas (epimutaciones) comunes en los individuos Lynch-like cuando su patrón de metilación en sangre o tejido colónico es comparado con individuos con síndrome de Lynch (con mutación genética identificada) o controles sanos. Sin embargo, es necesario puntualizar que este estudio se hizo de forma grupal y no podemos descartar completamente que haya alguna alteración puntual en locus no previstos. La carencia de regiones candidatas ha imposibilitado el estudio individual de regiones susceptibles de epimutación más allá de las regiones reguladoras de genes actualmente descritos de predisposición a cáncer colorrectal.

1.4-Alteraciones somáticas en tumores de individuos Lynch-like

El aumento del uso de la secuenciación en tumores puso de manifiesto la existencia de tumores con deficiencia reparadora debido a dobles mutaciones somáticas en los genes MMR. Las dobles mutaciones somáticas se consideran eventos esporádicos que mimetizan a los tumores de síndrome de Lynch debido a la deficiencia reparadora que adquieren. Estos tumores representan entre el 10% y el 52% de los casos clasificados como Lynch-like (Geurts-Giele *et al*, 2014; Haraldsdottir *et al*, 2014; Markow *et al*, 2017; Mensenkamp *et al*, 2014; Sourrouille *et al*, 2013; Vargas-Parra *et al*, 2017-Artículo 1).

El estudio mutacional somático se ha realizado en una subserie de pacientes Lynch-like con pérdida de expresión de MSH2/MSH6, con el fin de estimar la incidencia de las dobles mutaciones somáticas en nuestra serie (Vargas-Parra *et al*, 2017-Artículo 1). De los 5 tumores analizados, 2 tumores presentaron dobles mutaciones somáticas y los otros 3, pérdidas de heterocigosidad, mutaciones somáticas heterocigotas en diferentes genes MMR y/o mutaciones en los genes *POLE/POLD1*. En estos casos, las mutaciones somáticas en *POLE/POLD1* podrían ser el evento iniciador, puesto que una corrección defectuosa por parte de las polimerasas, podría provocar mutaciones en los genes MMR y, por tanto, inestabilidad de microsatélites (Jansen *et al*, 2016).

Sin embargo, y a pesar de la identificación de dobles eventos en genes MMR, no se puede descartar la existencia de una predisposición hereditaria al cáncer en estos pacientes. Así, se ha documentado la existencia de mosaicismo germinal en dos casos Lynch-like cuya mutación en línea germinal no era fácilmente detectable en sangre pero que estaba presente en sus tumores (Pastrello *et al*, 2009; Sourrouille *et al*, 2013) y un portador bialélico de *MUTYH* cuyo tumor presentaba también dobles mutaciones somáticas en *MSH2* (Morak *et al*, 2014). Por ello, las estrategias basadas en el análisis tumoral y germinal de pacientes con sospecha de síndrome de Lynch, con elevada sensibilidad para la detección de variantes, permitirán conocer mejor la naturaleza molecular de estos casos (Gray *et al*, 2018; Vargas-Parra *et al*, 2017-Artículo 1).

En referencia a la hipermetilación de los promotores si bien la inactivación del gen *MLH1* causada ha sido reportada a nivel somático (Herman *et al*, 1998; Yamamoto & Imai, 2015), la contribución relativa de la metilación somática en otros promotores en LLS no parece jugar un papel significativo (Nagasaka *et al*, 2010; Rumilla *et al*, 2011; Vargas-Parra *et al*, 2017-Artículo 1).

Los estudios más recientes muestran que no existe ninguna característica clínico-patológica que permita diferenciar los tumores de pacientes con síndrome de Lynch de los tumores con dobles mutaciones somáticas (Hemminger *et al*, 2018; Mas-Moya *et al*, 2015; Mills *et al*, 2016). De manera análoga, el análisis del metiloma global no permite discriminar entre tumores de pacientes con síndrome de Lynch y Lynch-like. La ausencia de otros grupos control, como individuos con tumores esporádicos con metilación somática de *MLH1* o individuos con tumores estables sin deficiencia reparadora, ha impedido el análisis mediante agrupación no supervisada lo que podría haber sido más informativo. La similitud observada puede ser debida a que los cambios epigenéticos asociados al desarrollo de la neoplasia son compartidos y relacionados con la deficiencia reparadora observada. De hecho, los estudios previos de metiloma de tumores colorrectales han coincidido en la definición de un grupo homogéneo de tumores con deficiencia MMR, principalmente formado por tumores con metilación somática en *MLH1* o mutaciones somáticas en genes MMR (Guinney *et al*, 2015; Hinoue *et al*, 2012; The Cancer Genome Atlas, 2012).

2-La importancia de la identificación y caracterización de las alteraciones epigenéticas constitucionales en síndrome de Lynch

2.1- Las técnicas altamente sensibles permiten la identificación de mosaicismo epigenético

El mosaicismo genético es muy poco frecuente en pacientes con síndrome de Lynch portadores de mutaciones genéticas. En cambio, la presencia de mosaicismo epigenético afecta a la mayor parte de los portadores de epimutación los casos. El rango de metilación es variable entre ellos (Figura 30), aunque un 6% de ellos presentan niveles de metilación por debajo del 10%, que se considera el límite de detección de las técnicas más comúnmente utilizadas en los laboratorios de diagnóstico molecular (Gausachs *et al*, 2012).

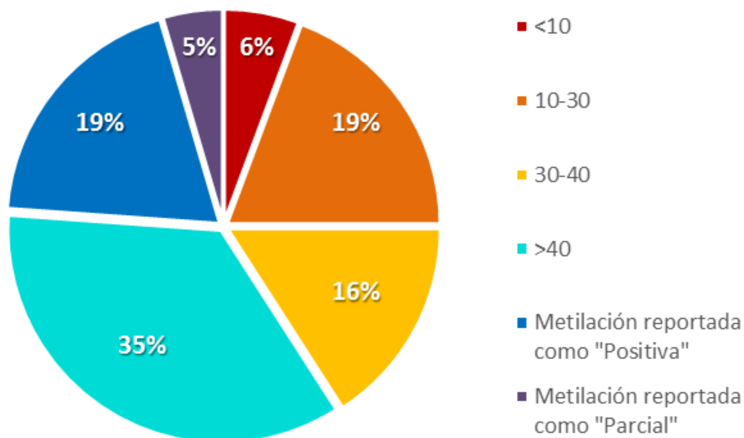


Figura 30: distribución de los casos de epimutación conocidos según el porcentaje de metilación reportado.

La puesta a punto del MS-MCA de la región C de Deng como una metodología de alta sensibilidad analítica para la detección de casos con epimutación en baja proporción ha permitido la identificación de un paciente con epimutación de aproximadamente el 1% (1/18) (Dámaso, Canet-Hermida et al, manuscrito en preparación-Artículo 4). Este resultado fue validado por otras técnicas. La secuenciación clonal por bisulfito ha

permitido validar de manera concluyente el grado de mosaicismo de la epimutación y que la metilación era monoalélica. Asimismo, el estudio del metiloma ha permitido verificar que la región hipermetilada correspondía a la previamente descrita en portadores de epimutación (Chr3: 37,033,791-37,035,400) (Dámaso *et al*, 2018-Artículo 3) y que estaba presente en tejido colónico normal.

A pesar de que la agresividad del fenotipo de la paciente contrasta con los bajos niveles de metilación constitucional, la identificación del mismo patrón (epi)genético en todos sus tumores (metilación, pérdida de heterocigosidad del mismo alelo), es coherente con el papel etiológico del mosaicismo epigenético en *MLH1* en la predisposición al cáncer. Aunque hemos descartado formalmente la presencia de mutaciones en *MLH1* y otros genes de predisposición, no podemos descartar la existencia de otros factores genéticos o ambientales que contribuyan al fenotipo observado.

En la actualidad, la tasa de detección de epimutaciones en series de pacientes con tumores *MLH1* metilados en pacientes diagnosticados antes de los 50 años es del 14% (Hitchins, 2013; Pineda *et al*, 2012; van Roon *et al*, 2010). Sin embargo, esta proporción podría ser mayor con la inclusión de estrategias de alta sensibilidad como técnica de detección de metilación, como es el MS-MCA.

De acuerdo con los resultados obtenidos, proponemos una modificación en el algoritmo diagnóstico de las epimutaciones en síndrome de Lynch, en el cual, tras la verificación de la presencia de metilación en *MLH1* en el tumor, la metilación en línea germinal sea analizada utilizando MS-MCA (Figura 31). De este modo, todos los pacientes con epimutación en *MLH1* puedan ser identificados independientemente de la presencia de mosaicismo epigenético.

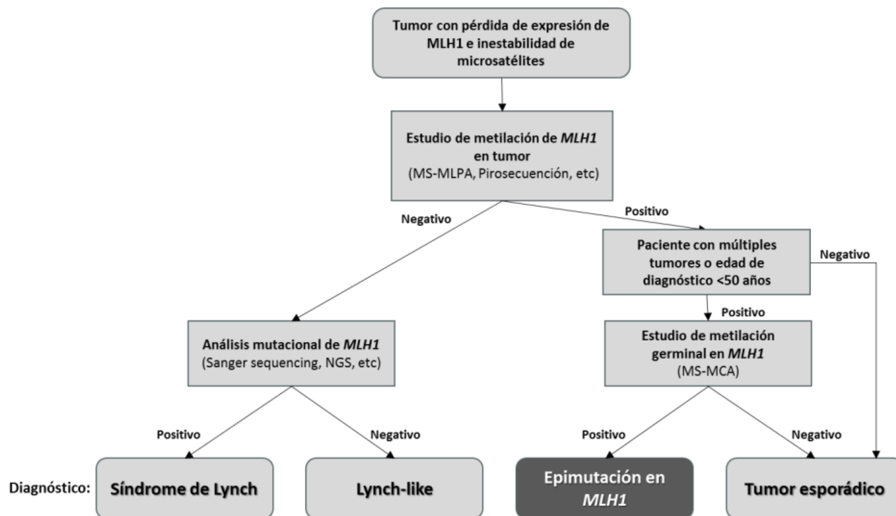


Figura 31: Algoritmo diagnóstico propuesto para la detección de epimutaciones en *MLH1* independientemente de la presencia de mosaicismo epigenético.

2.2- Las epimutaciones constitucionales en *MLH1* son un evento focal

El análisis del metiloma en 14 individuos con epimutación constitucional en *MLH1* ha puesto de manifiesto que la única isla CpG diferencialmente metilada son comparados con individuos con síndrome de Lynch o controles sanos, es la que engloba los promotores de *MLH1* y *EMP2AIP1* (Dámaso *et al*, 2018-Artículo 3). La extensión de la epimutación es constante en todos los casos analizados, 1,6 Kb (Chr3: 37,033,791-37,035,400) y se mantiene en todos los tejidos estudiados: sangre, mucosa colorrectal normal y tumor.

Esta región es específica de portadores de epimutación, puesto que la extensión de la región hipermetilada de tumores colorrectales esporádicos con metilación somática abarca aproximadamente 50 Kb más allá del extremo 5’ del gen de forma continuada (Dámaso *et al*, 2018 – Artículo 3; The Cancer Genome Atlas, 2012). Esta observación concuerda con estudios previos en tumores esporádicos donde la hipermetilación afecta, entre otros, genes situados en la citobanda 3p22 (Frigola *et al*, 2006; Hitchins *et al*, 2007; Smith & Costello, 2006). En este sentido, uno de los mecanismos propuestos

para la hipermetilación somática de *MLH1* ha sido la desmetilación de las secuencias Alu del intrón 1 de *MLH1*, las cuales provocarían una difusión de la metilación desde esta región, habitualmente metilada, hacia el promotor de *MLH1*, metilándolo (Wang *et al*, 2011; Zhang *et al*, 2012). Sin embargo, el estudio del estado de metilación de las secuencias Alu del intrón 1 de *MLH1* descarta este mecanismo causal en pacientes portadores de epimutación primaria. Será interesante conocer que patrón siguen las mutaciones secundarias. Si dicho patrón fuese capaz de unificar las epimutaciones y distinguirlas de los tumores colorrectales esporádicos, podría ayudar a la identificación y diagnóstico de pacientes portadores de epimutación.

Así pues, las epimutaciones primarias en *MLH1* son un evento epigenético focal, específico de este tipo de pacientes y mantenido tanto en los tejidos normales del individuo como sus tumores colorrectales.

2.3- La caracterización exhaustiva de las epimutaciones constitucionales permite su clasificación

Las epimutaciones primarias se definen por aparecer *de novo* en un individuo, no presentar alteraciones genéticas en *cis* y ser reversibles entre generaciones. En nuestra serie de 14 pacientes con epimutaciones constitucionales en *MLH1* hemos demostrado la aparición *de novo* de la epimutación en 3 probandos y borrado intergeneracional en dos, una de ellos caracterizado en profundidad en la presente tesis (Dámaso *et al*, 2018-Artículo 3; Pineda *et al*, 2012). En estas familias los probandos portadores de epimutación transmitían a su descendencia el alelo asociado a metilación en estado no metilado. Además, no se han identificado variantes genéticas en *cis* en la región promotora de *MLH1* en estos casos. Estos resultados apuntan a que la epimutación de estos pacientes es primaria.

Cabe destacar que en un caso de nuestra serie, se identificó una variante previamente no descrita en el promotor de *MLH1* (c.-234_-236delCTT) que, tras un análisis exahustivo resultó no ser causal, debido a que estaba situada en *trans* con el alelo epimutado (Dámaso *et al*, 2018-Artículo 3). El significado de las variantes identificadas en estos casos no siempre es claro. Por un lado, el hallazgo de variantes raras en regiones promotoras de *MLH1* en pacientes portadores de epimutación no debe considerarse como una evidencia sólida de epimutación secundaria. Por otro lado, la presencia de variantes genéticas asociadas al alelo epimutado (situadas principalmente en el promotor del gen) sí que lo puede ser. Sin embargo, su ausencia no debe ser interpretada como un indicio de epimutación primaria, puesto que se han descrito casos donde la epimutación cosegrega con un determinado haplotipo carente de variantes

(Crépin *et al*, 2012) o casos donde se identificaron deleciones del extremo 5’ del gen (Cini *et al*, 2015; Gylling *et al*, 2009; Renkonen *et al*, 2003).

En nuestra serie, hemos identificado 2 pacientes portadores de deleciones fuera del locus de *MLH1* mediante array CGH. Por el momento desconocemos si existe una relación causal entre los grandes reordenamientos identificados (pendientes de confirmar por oros métodos) y la epimutación. La ausencia de historia familiar de cáncer en estos casos y la extensión de la metilación en la región epimutada que fue idéntica a la definida para el resto de pacientes de la serie, sugiere que podrían tratarse de epimutaciones primarias.

En conclusión, una vez identificada una epimutación constitucional en *MLH1* es muy importante su confirmación mediante otras técnicas. Su análisis deberá complementarse con estudios de cosegregación en familiares, análisis de variantes (puntuales y reordenamientos) del promotor del gen y determinación de la fase de las variantes encontradas (si las hubiese). La clasificación de la epimutación como primaria o secundaria será muy útil para el consejo genético de estas familias.

2.4- En búsqueda del mecanismo causal

La definición de las epimutaciones primarias como evento epigenético focal, que abarca la misma región en todos los pacientes independientemente de la presencia de mosaicismo epigenético, sugiere la existencia de un mecanismo subyacente común en este tipo de pacientes. Este mecanismo subyacente además, debe ser distinto al responsable de las hipermetilaciones somáticas en *MLH1*.

La presencia de la epimutación en los tejidos normales de los portadores sugiere que se trata de un evento temprano en la vida del individuo (Figura 32). Concretamente, se cree que la epimutación debe aparecer en un estadio anterior a la aparición de las 3 capas embrionarias, debido a que la metilación está presente en todos los tejidos derivados de ellas en todos los pacientes analizados (Castillejo *et al*, 2015; Dámaso *et al*, 2018-Artículo 3; Goel *et al*, 2011; Gylling *et al*, 2009; Hitchins *et al*, 2005; Hitchins *et al*, 2007; Kidambi *et al*, 2016; Morak *et al*, 2008; Niessen *et al*, 2009; Pineda *et al*, 2012; Pinto *et al*, 2018; Suter *et al*, 2004b; Takeda *et al*, 2016; Valle *et al*, 2007; van Roon *et al*, 2010) Dámaso, Canet-Hermida *et al*, manuscrito en preparación -Artículo 4).

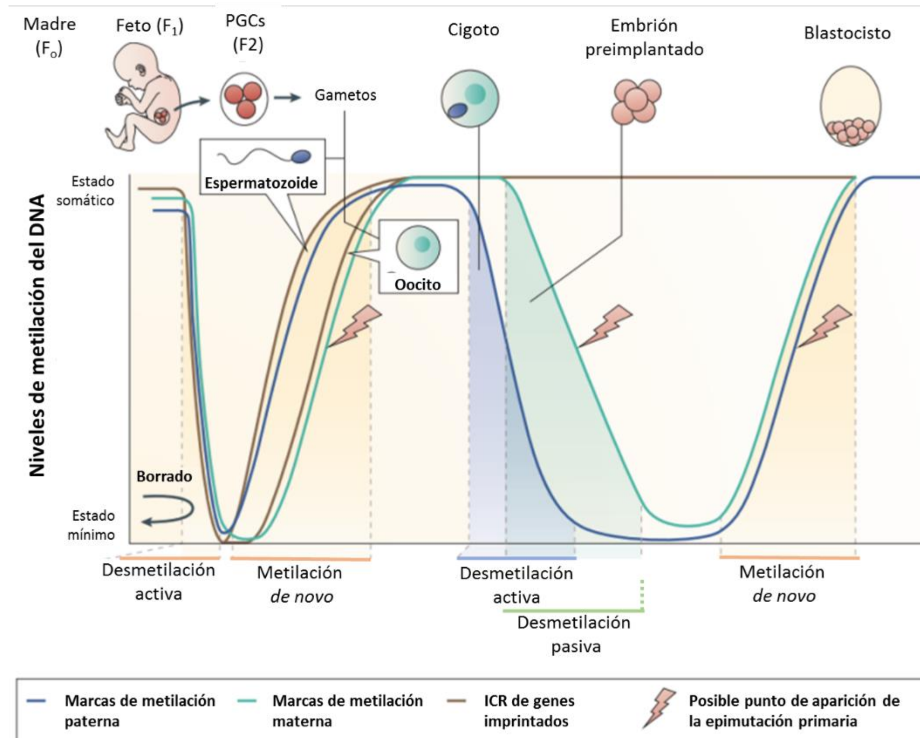


Figura 32: Esquema de los eventos de reprogramación epigenética en el ciclo de vida humano. Durante el desarrollo del feto (generación F1), las marcas epigenéticas parentales (de la generación F0) se borran mediante la desmetilación activa en las células germinales primordiales (PGC), que son los precursores de los gametos (y que se convertirán en la generación F2). Durante la maduración de los gametos, las regiones germinales diferencialmente metiladas se establecen *de novo*, incluso en las regiones de control de imprinting (ICR). Con un rayo anaranjado, se señalan las etapas en las que es más probables que se establezca una epimutatación constitucional primaria. Con la flecha negra curva del margen inferior izquierdo, se muestra la fase en la que es más probable que se produzca el borrado de una epimutatación parental existente, lo que reduce la probabilidad de herencia epigenética transgeneracional.
(Hitchins 2015)

Durante la maduración de los gametos y la creación del blastocisto tienen lugar dos eventos de reprogramación epigenética en los que las células desmetilan completamente su DNA para volver a remetilarlo (Hitchins, 2015). Es en este momento donde se podrían generar las epimutaciones, ya sea por la ausencia de desmetilación o por la retención de una hipermetilación del locus (Figura 32). El momento preciso y el origen celular de este defecto aún no se han determinado, pero se contempla la posibilidad de que puedan diferir entre casos, de modo que justificaría la presencia de casos con mosaicismo epigenético. En cualquier caso, es requisito obligatorio es la

pluripotencia de la célula. Así Takahashi y colaboradores indujeron la metilación *de novo* de *MLH1* en células madre embrionarias humanas (Takahashi *et al*, 2017). Sin embargo, el mismo trabajo no tuvo éxito cuando se realizó en otros tipos celulares (fibroblastos, células madre mesenquimales y la línea celular HeLa).

Nuestros datos no apoyan la existencia de una causa genética dentro de la región diferencialmente metilada como causa de la epimutaciones caracterizadas. Sin embargo, no podemos descartar la presencia de variantes genéticas fuera de sus límites. Los estudios de conformación e interacción del locus de *MLH1* con otras regiones o con efectores epigenéticos podrían ayudar a esclarecer el mecanismo causal de la epimutación. Esta búsqueda debería concentrarse en aquellas interacciones que puedan producirse durante el desarrollo embrionario.

Sumario

1. La evaluación de la patogenicidad de variantes de genes reparadores de desapareamientos del DNA mediante el estudio del RNA y su análisis multifactorial ofrece un buen rendimiento para diagnosticar individuos con síndrome de Lynch.

2. El análisis mediante técnicas de NGS permite identificar la presencia de dobles mutaciones somáticas de genes reparadores, presentes en un porcentaje significativo de pacientes Lynch-like, así como variantes en línea germinal en genes asociados a predisposición a cáncer.

3. Un pequeño porcentaje de pacientes Lynch-like podrían ser debidos a epimutaciones en *MLH1,* especialmente en los casos donde el estudio de metilación de *MLH1* en tumor es no informativo.

4. El uso de técnicas altamente sensibles – como el MS-MCA - permite la identificación de una pequeña proporción de epimutaciones constitucionales con mosaicismo epigenético en *MLH1*.

5. Aparentemente, no existen epimutaciones comunes fuera de los genes reparadores del DNA que jueguen un papel importante en la patología de los pacientes Lynch-like.

6. En las epimutaciones primarias de *MLH1* la hipermetilación se extiende focalmente en 1,6 Kb del locus de *MLH1* (Chr3: 37,033,791-37,035,400). Este evento epigenético no está ligado a variaciones genéticas en *cis* dentro de la región diferencialmente metilada.

CONCLUSIONES

Conclusiones

- El análisis exhaustivo mediante técnicas de NGS de línia germinal y tejido tumoral, combinado con la evaluación de la patogenicidad de variantes en genes reparadores de desapareamientos del DNA, son una estrategia efectiva para reclasificar individuos con síndrome de Lynch entre pacientes Lynch-like.

- La identificación y caracterización exhaustiva de los casos con epimutación en *MLH1* es fundamental para la correcta clasificación de estas familias. La determinación del tipo de epimutación tiene un gran impacto en la identificación de familiares con predisposición y en su asesoramiento genético.

- Las epimutaciones primarias en *MLH1* son un evento epigenético focal. Su mecanismo causal sigue siendo desconocido. Estudios conformacionales y de interacción del DNA con el locus de *MLH1* podrían arrojar luz sobre el establecimiento de la epimutación en el individuo.

REFERENCIAS

A

Achatz MI, Porter CC, Brugières L, Druker H, Frebourg T, Foulkes WD, Kratz CP, Kuiper RP, Hansford JR, Hernandez HS, Nathanson KL, Kohlmann WK, Doros L, Onel K, Schneider KW, Scollon SR, Tabori U, Tomlinson GE, Evans DGR, Plon SE (2017) Cancer Screening Recommendations and Clinical Management of Inherited Gastrointestinal Cancer Syndromes in Childhood. Clinical Cancer Research 23(13): e107-e114

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

Alexandrov LB, Stratton MR (2014) Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Current Opinion in Genetics & Development 24: 52-60

Alonso Sánchez A, Moreno S, Valiente A, Artigas M, Pérez-Juana A, Ramos-Arroyo M (2008) Mecanismos genéticos en la predisposición. 2008 29(1): 18

Aretz S, Stienen D, Uhlhaas S, Loff S, Back W, Pagenstecher C, McLeod DR, Graham GE, Mangold E, Santer R, Propping P, Friedl W (2005) High proportion of large genomic STK11 deletions in Peutz-Jeghers syndrome. Human Mutation 26(6): 513-519

B

Balmaña J, Balaguer F, Cervantes A, Arnold D, on behalf of the EGWG (2013) Familial risk-colorectal cancer: ESMO Clinical Practice Guidelines†. Annals of Oncology 24(suppl_6): vi73-vi80

Baris HN, Barnes-Kedar I, Toledano H, Halpern M, Hershkovitz D, Lossos A, Lerer I, Peretz T, Kariv R, Cohen S, Half EE, Magal N, Drasinover V, Wimmer K, Goldberg Y, Bercovich D, Levi Z (2016) Constitutional Mismatch Repair Deficiency in Israel: High Proportion of Founder Mutations in MMR Genes and Consanguinity. Pediatric Blood & Cancer 63(3): 418-427

Batte BAL, Bruegl AS, Daniels MS, Ring KL, Dempsey KM, Djordjevic B, Luthra R, Fellman BM, Lu KH, Broaddus RR (2014) Consequences of universal MSI/IHC in screening ENDOMETRIAL cancer patients for lynch syndrome. Gynecologic Oncology 134(2): 319-325

Belhadj S, Mur P, Navarro M, González S, Moreno V, Capellá G, Valle L (2017) Delineating the Phenotypic Spectrum of the NTHL1 Associated Polyposis. Clinical Gastroenterology and Hepatology 15(3): 461-462

Bellido F, Pineda M, Aiza G, Valdés-Mas R, Navarro M, Puente DA, Pons T, González S, Iglesias S, Darder E, Piñol V, Soto JL, Valencia A, Blanco I, Urioste M, Brunet J, Lázaro C, Capellá G, Puente XS, Valle L (2015) POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genetics In Medicine 18: 325

Bennett KL, Mester J, Eng C (2010) Germline epigenetic regulation of killin in cowden and cowden-like syndrome. JAMA 304(24): 2724-2731

Boland C, Troncale FJ (1984) FAmilial colonic cancer without antecedent polyposis. Annals of Internal Medicine 100(5): 700-701

Boland CR, Shike M (2010) Report From the Jerusalem Workshop on Lynch Syndrome-Hereditary Nonpolyposis Colorectal Cancer. Gastroenterology 138(7): 2197.e1-2197.e7

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, 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 Research 58(22): 5248-5257

Boland PM, Yurgelun MB, Boland CR (2018) Recent progress in Lynch syndrome and other familial colorectal cancer syndromes. CA: A Cancer Journal for Clinicians 68(3): 217-231

Borràs E, Pineda M, Brieger A, Hinrichsen I, Gómez C, Navarro M, Balmaña J, Ramón y Cajal T, Torres A, Brunet J, Blanco I, Plotz G, Lázaro C, Capellá G (2012)

Comprehensive functional assessment of MLH1 variants of unknown significance. Human Mutation 33(11): 1576-1588

Borràs E, Pineda M, Cadiñanos J, del Valle J, Brieger A, Hinrichsen I, Cabanillas R, Navarro M, Brunet J, Sanjuan X, Musulen E, van der Klift H, Lázaro C, Plotz G, Blanco I, Capellá G (2013) Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. Journal of Medical Genetics 50(8): 552-563

Boussioutas A, Li H, Liu J, Waring P, Lade S, Holloway AJ, Taupin D, Gorringe K, Haviv I, Desmond PV, Bowtell DDL (2003) Distinctive Patterns of Gene Expression in Premalignant Gastric Mucosa and Gastric Cancer. Cancer Research 63(10): 2569-2577

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

Brosens LAA, van Hattem A, Hylind LM, Iacobuzio-Donahue C, Romans KE, Axilbund J, Cruz-Correa M, Tersmette AC, Offerhaus GJA, Giardiello FM (2007) Risk of colorectal cancer in juvenile polyposis. Gut 56(7): 965-967

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

Buhard O, Suraweera N, Lectard A, Duval A, Hamelin R (2004) Quasimonomorphic Mononucleotide Repeats for High-Level Microsatellite Instability Analysis. Disease Markers 20(4-5): 251-257

Burn J, Gerdes A-M, Macrae F, Mecklin J-P, Moeslein G, Olschwang S, Eccles D, Evans DG, Maher ER, Bertario L, Bisgaard M-L, Dunlop MG, Ho JWC, Hodgson SV, Lindblom A, Lubinski J, Morrison PJ, Murday V, Ramesar R, Side L, Scott RJ, Thomas HJW, Vasen HF, Barker G, Crawford G, Elliott F, Movahedi M, Pylvanainen K, Wijnen JT, Fodde R, Lynch HT, Mathers JC, Bishop DT (2011) Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. The Lancet 378(9809): 2081-2087

Byrne RM, Tsikitis VL (2018) Colorectal polyposis and inherited colorectal cancer syndromes. Annals of Gastroenterology 31(1): 24-34

C

Castellsagué E, Liu J, Volenik A, Giroux S, Gagné R, Maranda B, Roussel-Jobin A, Latreille J, Laframboise R, Palma L, Kasprzak L, Marcus VA, Breguet M, Nolet S, El-Haffaf Z, Australie K, Gologan A, Aleynikova O, Oros-Klein K, Greenwood C, Mes-Masson AM, Provencher D, Tischkowitz M, Chong G, Rousseau F, Foulkes WD (2015) Characterization of a novel founder MSH6 mutation causing Lynch syndrome in the French Canadian population. Clinical Genetics 87(6): 536-542

Castillejo A, Hernández-Illán E, Rodriguez-Soler M, Pérez-Carbonell L, Egoavil C, Barberá VM, Castillejo M-I, Guarinos C, Martínez-de-Dueñas E, Juan M-J, Sánchez-Heras A-B, García-Casado Z, Ruiz-Ponte C, Brea-Fernández A, Juárez M, Bujanda L, Clofent J, Llor X, Andreu M, Castells A, Carracedo A, Alenda C, Payá A, Jover R, Soto J-L (2015) Prevalence of MLH1 constitutional epimutations as a cause of Lynch syndrome in unselected versus selected consecutive series of patients with colorectal cancer. Journal of Medical Genetics 52(7): 498-502

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

Cejka P, Marra G, Hemmerle C, Cannavó E, Storchova Z, Jiricny J (2003) Differential Killing of Mismatch Repair-Deficient and -Proficient Cells. Towards the Therapy of Tumors with Microsatellite Instability 63(23): 8113-8117

Cenin DR, Naber SK, Lansdorp-Vogelaar I, Jenkins MA, Buchanan DD, Preen DB, Ee HC, O'Leary P (2018) Costs and outcomes of Lynch syndrome screening in the Australian colorectal cancer population. Journal of Gastroenterology and Hepatology 0(0)

Cini G, Carnevali I, Quaia M, Chiaravalli AM, Sala P, Giacomini E, Maestro R, Tibiletti MG, Viel A (2015) Concomitant mutation and epimutation of the MLH1 gene in a Lynch syndrome family. Carcinogenesis 36(4): 452-458

Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. Nucleic Acids Research 22(15): 2990-2997

Colebatch A, Hitchins M, Williams R, Meagher A, Hawkins NJ, Ward RL (2006) The role of MYH and microsatellite instability in the development of sporadic colorectal cancer. British Journal Of Cancer 95: 1239

Crépin M, Dieu M-C, Lejeune S, Escande F, Boidin D, Porchet N, Morin G, Manouvrier S, Mathieu M, Buisine M-P (2012) Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. Human Mutation 33(1): 180-188

Chika N, Eguchi H, Kumamoto K, Suzuki O, Ishibashi K, Tachikawa T, Akagi K, Tamaru J-i, Okazaki Y, Ishida H (2017) Prevalence of Lynch syndrome and Lynch-like syndrome among patients with colorectal cancer in a Japanese hospital-based population. Japanese Journal of Clinical Oncology 47(2): 108-117

Choi YJ, Oh HR, Choi MR, Gwak M, An CH, Chung YJ, Yoo NJ, Lee SH (2014) Frameshift mutation of a histone methylation-related gene SETD1B and its regional heterogeneity in gastric and colorectal cancers with high microsatellite instability. Human Pathology 45(8): 1674-1681

Chung DC, Rustgi AK (2003) The hereditary nonpolyposis colorectal cancer syndrome: Genetics and clinical implications. Annals of Internal Medicine 138(7): 560-570

D

D'Agostino VG, Minoprio A, Torreri P, Marinoni I, Bossa C, Petrucci TC, Albertini AM, Ranzani GN, Bignami M, Mazzei F (2010) Functional analysis of MUTYH mutated proteins associated with familial adenomatous polyposis. DNA Repair 9(6): 700-707

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

Domingo E, Espín E, Armengol M, Oliveira C, Pinto M, Duval A, Brennetot C, Seruca R, Hamelin R, Yamamoto H, Schwartz S (2004) Activated BRAF targets proximal colon tumors with mismatch repair deficiency and MLH1 inactivation. Genes, Chromosomes and Cancer 39(2): 138-142

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

Durno CA, Sherman PM, 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. European Journal of Cancer 51(8): 977-983

Dymerska D, Gołębiewska K, Kuświk M, Rudnicka H, Scott RJ, Billings R, Pławski A, Boruń P, Siołek M, Kozak-Klonowska B, Szwiec M, Kilar E, Huzarski T, Byrski T, Lubiński J, Kurzawski G (2017) New EPCAM founder deletion in Polish population. Clinical Genetics 92(6): 649-653

E

Eads CA, Laird PW (2002) Combined Bisulfite Restriction Analysis (COBRA). In DNA Methylation Protocols, Mills KI, Ramsahoye BH (eds), pp 71-86. Totowa, NJ: Springer New York

Ebinger M, Sotlar K, Weber A, Bock CT, Bültmann BD, Kandolf R (2006) Simplified detection of microsatellite instability in colorectal cancer without the need for corresponding germline DNA analysis. Journal of Clinical Pathology 59(10): 1114-1115

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

Esteban-Jurado C, Franch-Expósito S, Muñoz J, Ocaña T, Carballal S, López-Cerón M, Cuatrecasas M, Vila-Casadesús M, Lozano JJ, Serra E, Beltran S, Brea-Fernández A, Ruiz-Ponte C, Castells A, Bujanda L, Garre P, Caldés T, Cubiella J, Balaguer F, Castellví-Bel S (2016) The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. European Journal Of Human Genetics 24: 1501

F

Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61(5): 759-767

Foulkes WD, Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C, Shia J, Markowitz A, Figer A, Friedman E, Farber D, Greenwood CMT, Bonner JD, Nafa K, Walsh T, Marcus V, Tomsho L, Gebert J, Macrae FA, Gaff CL, Paillerets BB-d, Gregersen PK, Weitzel JN, Gordon PH, MacNamara E, King MC, Hampel H, de la Chapelle A, Boyd J, Offit K, Rennert G, Chong G, Ellis NA (2002) The Founder Mutation MSH2*1906G>C Is an Important Cause of Hereditary Nonpolyposis Colorectal Cancer in the Ashkenazi Jewish Population. The American Journal of Human Genetics 71(6): 1395-1412

Frigola J, Song J, Stirzaker C, Hinshelwood RA, Peinado MA, Clark SJ (2006) Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nature Genetics 38: 540

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences 89(5): 1827-1831

G

Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M (2010) Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants. In Cancer Susceptibility: Methods and Protocols, Webb M (ed), pp 249-257. Totowa, NJ: Humana Press

Gala MK, Mizukami Y, Le LP, Moriichi K, Austin T, Yamamoto M, Lauwers GY, Bardeesy N, Chung DC (2014) Germline Mutations in Oncogene-Induced Senescence Pathways Are Associated With Multiple Sessile Serrated Adenomas. Gastroenterology 146(2): 520-529.e6

García-Palacios M, Bautista-Casasnovas AL (2014) Síndromes de poliposis intestinales. Anales de Pediatria 12(4): 183-90

Gardiner-Garden M, Frommer M (1987) CpG Islands in vertebrate genomes. Journal of Molecular Biology 196(2): 261-282

Garre P, Martín L, Sanz J, Romero A, Tosar A, Bando I, Llovet P, Diaque P, García-Paredes B, Díaz-Rubio E, de la Hoya M, Caldés T (2015) BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. Clinical Genetics 87(6): 582-587

Gausachs M, Mur P, Corral J, Pineda M, González S, Benito L, Menéndez M, Espinàs JA, Brunet J, Iniesta MD, Gruber SB, Lázaro C, Blanco I, Capellá G (2012) MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. European Journal Of Human Genetics 20: 762

Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD (2002) A Hereditary Nonpolyposis Colorectal Carcinoma Case Associated with Hypermethylation of the MLH1 Gene in Normal Tissue and Loss of Heterozygosity of the Unmethylated Allele in the Resulting Microsatellite Instability-High Tumor. Cancer Research 62(14): 3925-3928

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

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 (2014) Guidelines on Genetic Evaluation and Management of Lynch Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 147(2): 502-526

GLOBOCAN (2012) Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 (http://globocan.iarc.fr/Default.aspx). (web)

Goel A, Nguyen T-P, Leung H-CE, Nagasaka T, Rhees J, Hotchkiss E, Arnold M, Banerji P, Koi M, Kwok C-T, Packham D, Lipton L, Boland CR, Ward RL, Hitchins MP (2011) De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. International Journal of Cancer 128(4): 869-878

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

Graham RP, Kerr SE, Butz ML, Thibodeau SN, Halling KC, Smyrk TC, Dina MA, Waugh VM, Rumilla KM (2015) Heterogenous MSH6 Loss Is a Result of Microsatellite Instability Within MSH6 and Occurs in Sporadic and Hereditary Colorectal and Endometrial Carcinomas. The American Journal of Surgical Pathology 39(10): 1370-1376

Gray PN, Tsai P, Chen D, Wu S, Hoo J, Mu W, Li B, Vuong H, Lu H-M, Batth N, Willett S, Uyeda L, Shah S, Gau C-L, Umali M, Espenschied C, Janicek M, Brown S, Margileth D, Dobrea L, Wagman L, Rana H, Hall MJ, Ross T, Terdiman J, Cullinane C, Ries S, Totten E, Elliott AM (2018) TumorNext-Lynch-MMR: a comprehensive next generation sequencing assay for the detection of germline and somatic mutations in genes associated with mismatch repair deficiency and Lynch syndrome. Oncotarget 9(29): 20304-20322

Graziano A, Gutierrez A, Collia Avila K (2009) Síndromes de poliposis hereditarios. Enciclopedia de Cirugía Digestiva Fernando Galindo Tomo III(324): 1-15

Greenblatt MS, Brody LC, Foulkes WD, Genuardi M, Hofstra RMW, Olivier M, Plon SE, Sijmons RH, Sinilnikova O, Spurdle AB (2008) Locus-specific databases and recommendations to strengthen their contribution to the classification of variants in cancer susceptibility genes. Human Mutation 29(11): 1273-1281

Gruber SB (2006) New Developments in Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer) and Mismatch Repair Gene Testing. Gastroenterology 130(2): 577-587

Guéant J-L, Chéry C, Oussalah A, Nadaf J, Coelho D, Josse T, Flayac J, Robert A, Koscinski I, Gastin I, Filhine-Tresarrieu P, Pupavac M, Brebner A, Watkins D, Pastinen T,

Montpetit A, Hariri F, Tregouët D, Raby BA, Chung WK, Morange P-E, Froese DS, Baumgartner MR, Benoist J-F, Ficicioglu C, Marchand V, Motorin Y, Bonnemains C, Feillet F, Majewski J, Rosenblatt DS (2018) A PRDX1 mutant allele causes a MMACHC secondary epimutation in cblC patients. Nature Communications 9(1): 67

Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S (2015) The consensus molecular subtypes of colorectal cancer. Nature Medicine 21: 1350

Gylling A, Ridanpää M, Vierimaa O, Aittomäki K, Avela K, Kääriäinen H, Laivuori H, Pöyhönen M, Sallinen S-L, Wallgren-Pettersson C, Järvinen HJ, Mecklin J-P, Peltomäki P (2009) Large genomic rearrangements and germline epimutations in Lynch syndrome. International Journal of Cancer 124(10): 2333-2340

H

Hagen CE, Lefferts J, Hornick JL, Srivastava A (2011) “Null Pattern” of Immunoreactivity in a Lynch Syndrome-Associated Colon Cancer Due to Germline MSH2 Mutation and Somatic MLH1 Hypermethylation. The American Journal of Surgical Pathology 35(12): 1902-1905

Halvarsson B, Lindblom A, Johansson L, Lagerstedt K, Nilbert M (2005) Loss of mismatch repair protein immunostaining in colorectal adenomas from patients with hereditary nonpolyposis colorectal cancer. Modern Pathology 18: 1095

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

Hampel H (2016) Genetic counseling and cascade genetic testing in Lynch syndrome. Familial Cancer 15(3): 423-427

Hampel H, de la Chapelle A (2011) The Search for Unaffected Individuals with Lynch Syndrome: Do the Ends Justify the Means? Cancer Prevention Research 4(1): 1-5

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Clendenning M, Sotamaa K, Prior T, Westman JA, 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. Journal of Clinical Oncology 26(35): 5783-5788

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J, Panescu J, Fix D, Lockman J, Comeras I, de la Chapelle A (2005) Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). New England Journal of Medicine 352(18): 1851-1860

Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. Cell 100(1): 57-70

Hanahan D, Weinberg Robert A (2011) Hallmarks of Cancer: The Next Generation. Cell 144(5): 646-674

Hanks S, Coleman K, Reid S, Plaja A, Firth H, FitzPatrick D, Kidd A, Méhes K, Nash R, Robin N, Shannon N, Tolmie J, Swansbury J, Irrthum A, Douglas J, Rahman N (2004) Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nature Genetics 36: 1159

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

Hao H-X, Xie Y, Zhang Y, Charlat O, Oster E, Avello M, Lei H, Mickanin C, Liu D, Ruffner H, Mao X, Ma Q, Zamponi R, Bouwmeester T, Finan PM, Kirschner MW, Porter JA, Serluca FC, Cong F (2012) ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. Nature 485: 195

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

Haraldsdottir S, Rafnar T, Frankel WL, 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 PG, Sigurdsson F, Bekaii-Saab T, Moller PH, Steinarsdottir M, Alexiusdottir K, Hitchins M, Pritchard CC, de la

Chapelle A, Jonasson JG, Goldberg RM, Stefansson K (2017) Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. Nature Communications 8: 14755

Hegde M, Ferber M, Mao R, Samowitz W, Ganguly A, a Working Group of the American College of Medical G, Genomics Laboratory Quality Assurance C (2013) ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genetics In Medicine 16: 101

Heinen CD, Rasmussen L (2012) Determining the functional significance of mismatch repair gene missense variants using biochemical and cellular assays. Hereditary Cancer in Clinical Practice 10(1): 9

Hemminger JA, Pearlman R, Haraldsdottir S, Knight D, Jonasson JG, Pritchard CC, Hampel H, Frankel WL (2018) Histology of colorectal adenocarcinoma with double somatic mismatch-repair mutations is indistinguishable from those caused by Lynch syndrome. Human Pathology 78: 125-130

Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa J-PJ, Markowitz S, Willson JKV, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proceedings of the National Academy of Sciences 95(12): 6870-6875

Hesson LB, Hitchins MP, Ward RL (2010) Epimutations and cancer predisposition: importance and mechanisms. Current Opinion in Genetics & Development 20(3): 290-298

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

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. Clinical Cancer Research 19(9): 2432-2441

Hitchins M, Williams R, Cheong K, Halani N, Lin VAP, Packham D, Ku S, Buckle A, Hawkins N, Burn J, Gallinger S, Goldblatt J, Kirk J, Tomlinson I, Scott R, Spigelman A, Suter C, Martin D, Suthers G, Ward R (2005) MLH1 Germline Epimutations as a Factor

in Hereditary Nonpolyposis Colorectal Cancer. Gastroenterology 129(5): 1392-1399

Hitchins MP (2013) The role of epigenetics in Lynch syndrome. Familial Cancer 12(2): 189-205

Hitchins MP (2015) Constitutional epimutation as a mechanism for cancer causality and heritability? Nat Rev Cancer 15(10): 625-634

Hitchins MP, Lin VA, Buckle A, Cheong K, Halani N, Ku S, Kwok C-T, Packham D, Suter CM, Meagher A, Stirzaker C, Clark S, Hawkins NJ, Ward RL (2007) Epigenetic Inactivation of a Cluster of Genes Flanking MLH1 in Microsatellite-Unstable Colorectal Cancer. Cancer Research 67(19): 9107-9116

Hitchins Megan P, Rapkins Robert W, Kwok C-T, Srivastava S, Wong Justin JL, Khachigian Levon M, Polly P, Goldblatt J, Ward Robyn L (2011) Dominantly Inherited Constitutional Epigenetic Silencing of MLH1 in a Cancer-Affected Family Is Linked to a Single Nucleotide Variant within the 5UTR. Cancer Cell 20(2): 200-213

Hitchins MP, Ward RL (2009) Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. Journal of Medical Genetics 46(12): 793-802

Hitchins MP, Wong JJL, Suthers G, Suter CM, Martin DIK, Hawkins NJ, Ward RL (2007) Inheritance of a Cancer-Associated MLH1 Germ-Line Epimutation. New England Journal of Medicine 356(7): 697-705

Howarth DR, Lum SS, Esquivel P, Garberoglio CA, Senthil M, Solomon NL (2015) Initial Results of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer and Lynch Syndrome. The American Surgeon 81(10): 941-944

Howe JR, Sayed MG, Ahmed AF, Ringold J, Larsen-Haidle J, Merg A, Mitros FA, Vaccaro CA, Petersen GM, Giardiello FM, Tinley ST, Aaltonen LA, Lynch HT (2004) The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B and ACVR1 mutations. Journal of Medical Genetics 41(7): 484-491

Hui VW, Steinhagen E, Levy RA, Sood R, Shia J, Markowitz AJ, Guillem JG (2014) Utilization of Colonoscopy and Pathology Reports for Identifying Patients

Meeting the World Health Organization Criteria for Serrated Polyposis Syndrome. Diseases of the Colon & Rectum 57(7): 846-850

I

Inamura K (2018) Colorectal Cancers: An Update on Their Molecular Pathology. Cancers 10(1): 26

J

Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinimann K, Howarth K, East J, Taylor J, Thomas H, Tomlinson I (2012) Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nature Genetics 44: 699

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

Järvinen HJ, Renkonen-Sinisalo L, Aktán-Collán K, Peltomäki P, Aaltonen LA, Mecklin J-P (2009) Ten Years After Mutation Testing for Lynch Syndrome: Cancer Incidence and Outcome in Mutation-Positive and Mutation-Negative Family Members. Journal of Clinical Oncology 27(28): 4793-4797

Jasperson KW, Tuohy TM, Neklason DW, Burt RW (2010) Hereditary and Familial Colon Cancer. Gastroenterology 138(6): 2044-2058

Jones N, Vogt S, Nielsen M, Christian D, Wark PA, Eccles D, Edwards E, Evans DG, Maher ER, Vasen HF, Hes FJ, Aretz S, Sampson JR (2009) Increased Colorectal Cancer Incidence in Obligate Carriers of Heterozygous Mutations in MUTYH. Gastroenterology 137(2): 489-494.e1

K

Kalady MF, Kravochuck SE, Heald B, Burke CA, Church JM (2015) Defining the Adenoma Burden in Lynch Syndrome. Diseases of the Colon & Rectum 58(4): 388-392

Kapitanović S (2017) Molecular stratification of sporadic and hereditary colorectal cancer: a mini review. Medical Sciences 44: 73-79

Kempers MJE, Kuiper RP, Ockeloen CW, Chappuis PO, Hutter P, Rahner N, Schackert HK, Steinke V, Holinski-Feder E, Morak M, Kloor M, Büttner R, Verwiel ETP, van Krieken JH, Nagtegaal ID, Goossens M, van der Post RS, Niessen RC, Sijmons RH, Kluijt I, Hogervorst FBL, Leter EM, Gille JJP, Aalfs CM, Redeker EJW, Hes FJ, Tops CMJ, van Nesselrooij BPM, van Gijn ME, García EBG, Eccles DM, Bunyan DJ, Syngal S, Stoffel EM, Culver JO, Palomares MR, Graham T, Velsher L, Papp J, Oláh E, Chan TL, Leung SY, van Kessel AG, Kiemeney LALM, Hoogerbrugge N, Ligtenberg MJL (2011) Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. The Lancet Oncology 12(1): 49-55

Kidambi TD, Blanco A, Van Ziffle J, Terdiman JP (2016) Constitutional MLH1 methylation presenting with colonic polyposis syndrome and not Lynch syndrome. Familial Cancer 15(2): 275-280

Kim T-M, Park PJ (2014) A Genome-wide View of Microsatellite Instability: Old Stories of Cancer Mutations Revisited with New Sequencing Technologies. Cancer Research 74(22): 6377-6382

Kim YH, Kakar S, Cun L, Deng G, Kim YS (2008) Distinct CpG island methylation profiles and BRAF mutation status in serrated and adenomatous colorectal polyps. International Journal of Cancer 123(11): 2587-2593

Kinnersley B, Chubb D, Dobbins SE, Frampton M, Buch S, Timofeeva MN, Castellví-Bel S, Farrington SM, Forsti A, Hampe J, Hemminki K, Hofstra RMW, Northwood E, Palles C, Pinheiro M, Ruiz-Ponte C, Schafmayer C, Teixeira MR, Westers H, van Wezel T, Timothy Bishop D, Tomlinson I, Dunlop MG, Houlston RS (2016) Correspondence: SEMA4A variation and risk of colorectal cancer. Nature Communications 7: 10611

Knudson AG (2001) Two genetic hits (more or less) to cancer. Nature Reviews Cancer 1: 157

Kurdyukov S, Bullock M (2016) DNA Methylation Analysis: Choosing the Right Method. Biology 5(1): 3

Kwok C-T, Vogelaar IP, van Zelst-Stams WA, Mensenkamp AR, Ligtenberg MJ, Rapkins RW, Ward RL, Chun N, Ford JM, Ladabaum U, McKinnon WC, Greenblatt MS, Hitchins MP (2014) The MLH1 c.-27C>A and c.85G>T variants are linked to dominantly inherited MLH1 epimutation and are borne on a European ancestral haplotype. Eur J Hum Genet 22(5): 617-624

L

Laghi L, Bianchi P, Roncalli M, Malesci A (2004) Re: Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. JNCI: Journal of the National Cancer Institute 96(18): 1402-1403

Laitman Y, Jaeger E, Katz L, Tomlinson IAN, Friedman E (2015) GREM1 germline mutation screening in Ashkenazi Jewish patients with familial colorectal cancer. Genetics Research 97: e11

Lao VV, Grady WM (2011) Epigenetics and colorectal cancer. Nature Reviews Gastroenterology &Amp; Hepatology 8: 686

Lázaro C, Feliubadaló L, del Valle J (2018) Genetic Testing in Hereditary Colorectal Cancer. In Hereditary Colorectal Cancer: Genetic Basis and Clinical Implications, Valle L, Gruber SB, Capellá G (eds), Chapter 16, pp 209-232. Springer

Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, Wong F, Azad NS, Rucki AA, Laheru D, Donehower R, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Greten TF, Duffy AG, Ciombor KK, Eyring AD, Lam BH, Joe A, Kang SP, Holdhoff M, Danilova L, Cope L, Meyer C, Zhou S, Goldberg RM, Armstrong DK, Bever KM, Fader AN, Taube J, Housseau F, Spetzler D, Xiao N, Pardoll DM, Papadopoulos N, Kinzler KW, Eshleman JR, Vogelstein B, Anders RA, Diaz LA (2017) Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 357(6349): 409-413

Lefevre J, Colas C, Coulet F, Baert-Desurmont S, Mongin C, Tiret E, Frebourg T, Soubrier F, Parc Y (2011) Frequent mutation in North African patients with MUTYH-associated polyposis. Clinical Genetics 80(4): 389-393

Lewis A, Freeman-Mills L, de la Calle-Mustienes E, Giráldez-Pérez Rosa M, Davis H, Jaeger E, Becker M, Hubner Nina C, Nguyen Luan N, Zeron-Medina J, Bond G, Stunnenberg Hendrik G, Carvajal Jaime J, Gomez-Skarmeta Jose L, Leedham S, Tomlinson I (2014) A Polymorphic Enhancer near GREM1 Influences Bowel Cancer Risk through Differential CDX2 and TCF7L2 Binding. Cell Reports 8(4): 983-990

Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li G-M (2013) The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutS. Cell 153(3): 590-600

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

Ligtenberg MJL, Kuiper RP, Geurts van Kessel A, Hoogerbrugge N (2013) EPCAM deletion carriers constitute a unique subgroup of Lynch syndrome patients. Familial Cancer 12(2): 169-174

Lipton LR, Johnson V, Cummings C, Fisher S, Risby P, Eftekhar Sadat AT, Cranston T, Izatt L, Sasieni P, Hodgson SV, Thomas HJW, Tomlinson IPM (2004) Refining the Amsterdam Criteria and Bethesda Guidelines: Testing Algorithms for the Prediction of Mismatch Repair Mutation Status in the Familial Cancer Clinic. Journal of Clinical Oncology 22(24): 4934-4943

Liu Q, Hesson LB, Nunez AC, Packham D, Williams R, Ward RL, Sloane MA (2016a) A cryptic paracentric inversion of MSH2 exons 2–6 causes Lynch syndrome. Carcinogenesis 37(1): 10-17

Liu Q, Thompson BA, Ward RL, Hesson LB, Sloane MA (2016b) Understanding the Pathogenicity of Noncoding Mismatch Repair Gene Promoter Variants in Lynch Syndrome. Human Mutation 37(5): 417-426

Lote H, Cafferkey C, Chau I (2015) PD-1 and PD-L1 blockade in gastrointestinal malignancies. Cancer Treatment Reviews 41(10): 893-903

Lynch D, Murphy A (2016) The emerging role of immunotherapy in colorectal cancer. Annals of Translational Medicine 4(16)

Lynch H, Schuelke G, Kimberling W, Albano W, Lynch J, Biscone K, Lipkin M, Deschner E, Mikol Y, Sandberg A (1985) Hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II. Cancer 56(4): 939-51

Lynch HT, Jascur T, Lanspa S, Boland CR (2010) Making Sense of Missense in Lynch Syndrome: The Clinical Perspective. Cancer Prevention Research 3(11): 1371-1374

Lynch HT, Lynch JF, Attard TA (2009) Diagnosis and management of hereditary colorectal cancer syndromes: Lynch syndrome as a model. Canadian Medical Association Journal 181(5): 273-280

Lynch HT, Lynch JF, Lynch PM (2007) Toward a Consensus in Molecular Diagnosis of Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome). JNCI: Journal of the National Cancer Institute 99(4): 261-263

Lynch HT, Shaw MW, Magnuson CW, Larsen AL, Krush AJ (1966) Hereditary factors in cancer: Study of two large midwestern kindreds. Archives of Internal Medicine 117(2): 206-212

Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP (2015) Milestones of Lynch syndrome: 1895-2015. Nat Rev Cancer 15(3): 181-194

M

Malvezzi M, Carioli G, Bertuccio P, Boffetta P, Levi F, La Vecchia C, Negri E (2017) European cancer mortality predictions for the year 2017, with focus on lung cancer. Annals of Oncology 28(5): 1117-1123

Mangold E, Pagenstecher C, Friedl W, Mathiak M, Buettner R, Engel C, Loeffler M, Holinski-Feder E, Müller-Koch Y, Keller G, Schackert HK, Krüger S, Goecke T, Moeslein G, Kloor M, Gebert J, Kunstmann E, Schulmann K, Rüschoff J, Propping P (2005) Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. International Journal of Cancer 116(5): 692-702

Mardis ER (2013) Next-Generation Sequencing Platforms. Annual Review of Analytical Chemistry 6(1): 287-303

Markow M, Chen W, Frankel WL (2017) Immunohistochemical Pitfalls. Surgical Pathology Clinics 10(4): 977-1007

Mas-Moya J, Dudley B, Brand RE, Thull D, Bahary N, Nikiforova MN, Pai RK (2015) Clinicopathological comparison of colorectal and endometrial carcinomas in patients with Lynch-like syndrome versus patients with Lynch syndrome. Human Pathology 46(11): 1616-1625

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

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

Mills AM, Sloan EA, Thomas M, Modesitt SC, Stoler MH, Atkins KA, Moskaluk CA (2016) Clinicopathologic Comparison of Lynch Syndrome–associated and “Lynch-like” Endometrial Carcinomas Identified on Universal Screening Using Mismatch Repair Protein Immunohistochemistry. The American Journal of Surgical Pathology 40(2): 155-165

Miyakura Y, Sugano K, Akasu T, Yoshida T, Maekawa M, Saitoh S, Sasaki H, Nomizu T, Konishi F, Fujita S, Moriya Y, Nagai H (2004) Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. Clinical Gastroenterology and Hepatology 2(2): 147-156

Møller P, Seppälä TT, Bernstein I, Holinski-Feder E, Sala P, Gareth Evans D, Lindblom A, Macrae F, Blanco I, Sijmons RH, Jeffries J, Vasen HFA, Burn J, Nakken S, Hovig E, Rødland EA, Tharmaratnam K, de Vos tot Nederveen Cappel WH, Hill J, Wijnen JT, Jenkins MA, Green K, Lalloo F, Sunde L, Mints M, Bertario L, Pineda M, Navarro M, Morak M, Renkonen-Sinisalo L, Valentin MD, Frayling IM, Plazzer J-P, Pylvanainen K, Genuardi M, Mecklin J-P, Moeslein G, Sampson JR, 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

Moorthie S, Mattocks CJ, Wright CF (2011) Review of massively parallel DNA sequencing technologies. The HUGO Journal 5(1): 1-12

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

Morak M, Schackert HK, 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

Moreira L, Balaguer F, Lindor N, et al. (2012) Identification of lynch syndrome among patients with colorectal cancer. JAMA 308(15): 1555-1565

Moreira L, Muñoz J, Cuatrecasas M, Quintanilla I, Leoz ML, Carballal S, Ocaña T, López-Cerón M, Pellise M, Castellví-Bel S, Jover R, Andreu M, Carracedo A, Xicola RM, Llor X, Boland CR, Goel A, Castells A, Balaguer F (2015) Prevalence of somatic mutl homolog 1 promoter hypermethylation in Lynch syndrome colorectal cancer. Cancer 121(9): 1395-1404

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

Murphy K, Darmawan H, Schultz A, da Silva EF, Reha-Krantz LJ (2006) A method to select for mutator DNA polymerase δs in Saccharomyces cerevisiae. Genome 49(4): 403-410

N

Nagasaka T, Rhees J, Kloor M, Gebert J, Naomoto Y, Boland CR, Goel A (2010) Somatic Hypermethylation of MSH2 Is a Frequent Event in Lynch Syndrome Colorectal Cancers. Cancer Research 70(8): 3098-3108

Nejadtaghi M, Jafari H, Farrokhi E, Samani KG (2017) Familial Colorectal Cancer Type X (FCCTX) and the correlation with various genes—A systematic review. Current Problems in Cancer 41(6): 388-397

Newton K, Jorgensen NM, Wallace AJ, Buchanan DD, Lalloo F, McMahon RFT, Hill J, Evans DG (2014) Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). Journal of Medical Genetics 51(12): 789-796

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

Nielsen M, Joerink - van de Beld MC, Jones N, Vogt S, Tops CM, Vasen HFA, Sampson JR, Aretz S, Hes FJ (2009) Analysis of MUTYH Genotypes and Colorectal Phenotypes in Patients With MUTYH-Associated Polyposis. Gastroenterology 136(2): 471-476

Nieminen TT, Abdel–Rahman WM, Ristimäki A, Lappalainen M, Lahermo P, Mecklin JP, Järvinen HJ, Peltomäki P (2011) BMPR1A Mutations in Hereditary Nonpolyposis Colorectal Cancer Without Mismatch Repair Deficiency. Gastroenterology 141(1): e23-e26

Nieminen TT, O'Donohue M-F, Wu Y, Lohi H, Scherer SW, Paterson AD, Ellonen P, Abdel-Rahman WM, Valo S, Mecklin J-P, Järvinen HJ, Gleizes P-E, Peltomäki P (2014) Germline Mutation of RPS20, Encoding a Ribosomal Protein, Causes

Predisposition to Hereditary Nonpolyposis Colorectal Carcinoma Without DNA Mismatch Repair Deficiency. Gastroenterology 147(3): 595-598.e5

Niessen RC, Hofstra RMW, Westers H, Ligtenberg MJL, Kooi K, Jager POJ, de Groote ML, Dijkhuizen T, Olderode-Berends MJW, Hollema H, Kleibeuker JH, Sijmons RH (2009) Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes, Chromosomes and Cancer 48(8): 737-744

Niu BT, Hammond RFL, Leen SLS, Faruqi AZ, Trevisan G, Gilks CB, Singh N Artefactual punctate MLH1 staining can lead to erroneous reporting of isolated PMS2 loss. Histopathology 0(ja)

Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho Y-Y, Kobayashi Y, Patil N, Thusberg J, Westbrook M, The Invitae Clinical Genomics G, Topper S (2017) Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. Genetics In Medicine 19: 1105

O

Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS (2007) Evaluation of Markers for CpG Island Methylator Phenotype (CIMP) in Colorectal Cancer by a Large Population-Based Sample. The Journal of Molecular Diagnostics 9(3): 305-314

Oliveira C, Pinto M, Duval A, Brennetot C, Domingo E, Espín E, Armengol M, Yamamoto H, Hamelin R, Seruca R, Schwartz Jr S (2003) BRAF mutations characterize colon but not gastric cancer with mismatch repair deficiency. Oncogene 22: 9192

OMS (2018) Organización Mundial de la Salud: Cáncer (http://www.who.int/es/news-room/fact-sheets/detail/cancer). (web)

Overbeek LIH, Kets CM, Hebeda KM, Bodmer D, der Looij Ev, Willems R, Goossens M, Arts N, Brunner HG, van Krieken JHJM, Hoogerbrugge N, Ligtenberg MJL (2007) Patients with an unexplained microsatellite instable tumour have a low risk of familial cancer. British Journal Of Cancer 96: 1605

P

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. Genetics In Medicine 11: 42

Palles C, Cazier J-B, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I, Sherborne A, Chubb D, Carvajal-Carmona LG, Ma Y, Kaur K, Dobbins S, Barclay E, Gorman M, Martin L, Kovac MB, Humphray S, The CC, The WGSC, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsey SE, Thomas HJW, McVean G, Houlston RS, Tomlinson I (2012) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nature Genetics 45: 136

Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB (2012) Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. Journal of Medical Genetics 49(3): 151-157

Pastrello C, Fornasarig M, Pin E, Berto E, Pivetta B, Viel A (2009) Somatic mosaicism in a patient with Lynch syndrome. American Journal of Medical Genetics Part A 149A(2): 212-215

Pedroni M, Roncari B, Maffei S, Losi L, Scarselli A, Di Gregorio C, Marino M, Roncucci L, Benatti P, Ponti G, Rossi G, Menigatti M, Viel A, Genuardi M, de Leon MP (2007) A Mononucleotide Markers Panel to Identify hMLH1/hMSH2 Germline Mutations. Disease Markers 23(3): 179-187

Peltomäki P (2001) Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 10(7): 735-40

Peltomäki P (2014) Epigenetic mechanisms in the pathogenesis of Lynch syndrome. Clinical Genetics 85(5): 403-412

Peltomäki P (2016) Update on Lynch syndrome genomics. Familial Cancer 15(3): 385-393

Pérez-Carbonell L, Alenda C, Payá A, Castillejo A, Barberá VM, Guillén C, Rojas E, Acame N, Gutiérrez-Aviñó FJ, 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. The Journal of Molecular Diagnostics 12(4): 498-504

Perucho M (1999) Correspondence re: C. R. Boland et al. A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. Cancer Res., 58:5248–5257, 1998. Cancer Research 59(1): 249-256

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 (2015) Detailed characterization of MLH1 p.D41H and p.N710D variants coexisting in a Lynch syndrome family with conserved MLH1 expression tumors. Clinical Genetics 87(6): 543-548

Pineda M, González S, Lázaro C, Blanco I, Capellá G (2010) Detection of genetic alterations in hereditary colorectal cancer screening. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 693(1): 19-31

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

Pino MS, Chung DC (2010) The Chromosomal Instability Pathway in Colon Cancer. Gastroenterology 138(6): 2059-2072

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

Pritchard CC, Grady WM (2011) Colorectal cancer molecular biology moves into clinical practice. Gut 60(1): 116-129

Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, Gulden C, Kupfer SS, Swisher EM, Bennett RL, Novetsky AP, Jarvik GP, Olopade OI, Goodfellow PJ, King M-C, Tait JF, Walsh T (2012) ColoSeq Provides Comprehensive Lynch and

Polyposis Syndrome Mutational Analysis Using Massively Parallel Sequencing. The Journal of Molecular Diagnostics 14(4): 357-366

Puccini A, Berger MD, Naseem M, Tokunaga R, Battaglin F, Cao S, Hanna DL, McSkane M, Soni S, Zhang W, Lenz H-J (2017) Colorectal cancer: epigenetic alterations and their clinical implications. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1868(2): 439-448

Q

Quintana I, Mejías-Luque R, Terradas M, Navarro M, Piñol V, Mur P, Belhadj S, Grau E, Darder E, Solanes A, Brunet J, Capellá G, Gerhard M, Valle L (2018) Evidence suggests that germline RNF43 mutations are a rare cause of serrated polyposis. Gut

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

R

Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE (2002) RAF/RAS oncogenes and mismatch-repair status. Nature 418: 934

Raskin L, Schwenter F, Freytsis M, Tischkowitz M, Wong N, Chong G, Narod SA, Levine DA, Bogomolniy F, Aronson M, Thibodeau SN, Hunt KS, Rennert G, Gallinger S, Gruber SB, Foulkes WD (2011) Characterization of two Ashkenazi Jewish founder mutations in MSH6 gene causing Lynch syndrome. Clinical Genetics 79(6): 512-522

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

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E (2013) ACMG clinical laboratory standards for next-generation sequencing. Genetics In Medicine 15: 733

Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, Aittomäki K, Järvinen HJ, Mecklin J-P, Lindblom A, Peltomäki P (2003) Altered Expression of MLH1, MSH2, and MSH6 in Predisposition to Hereditary Nonpolyposis Colorectal Cancer. Journal of Clinical Oncology 21(19): 3629-3637

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

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL (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. Genetics In Medicine 17: 405

Risio M, Reato G, di Celle PF, Fizzotti M, Rossini FP, Foà R (1996) Microsatellite Instability Is Associated with the Histological Features of the Tumor in Nonfamilial Colorectal Cancer. Cancer Research 56(23): 5470-5474

Rivera B, Castellsagué E, Bah I, van Kempen LC, Foulkes WD (2015) Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. New England Journal of Medicine 373(20): 1985-1986

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

Rohlin A, Eiengård F, Lundstam U, Zagoras T, Nilsson S, Edsjö A, Pedersen J, Svensson J, Skullman S, Karlsson BG, Björk J, Nordling M (2016) GREM1 and POLE variants in hereditary colorectal cancer syndromes. Genes, Chromosomes and Cancer 55(1): 95-106

Rohlin A, Rambech E, Kvist A, Törngren T, Eiengård F, Lundstam U, Zagoras T, Gebre-Medhin S, Borg Å, Björk J, Nilbert M, Nordling M (2017) Expanding the genotype–phenotype spectrum in hereditary colorectal cancer by gene panel testing. Familial Cancer 16(2): 195-203

Rohlin A, Zagoras, T., Nilsson, S., Lundstam, U., Wahlström, J., Hultén, L., Martinsson, T., Karlsson, G. B., Nordling, M (2014) A mutation in POLE predisposing to a multi-tumour phenotype. In International Journal of Oncology Vol. 45. pp 77-81

Rosner G, Bercovich D, Daniel YE, Strul H, Fliss-Isakov N, Ben-Yehoiada M, Santo E, Halpern Z, Kariv R (2015) Increased risk for colorectal adenomas and cancer in mono-allelic MUTYH mutation carriers: results from a cohort of North-African Jews. Familial Cancer 14(3): 427-436

Rumilla K, Schowalter KV, Lindor NM, Thomas BC, Mensink KA, Gallinger S, Holter S, Newcomb PA, Potter JD, Jenkins MA, Hopper JL, Long TI, Weisenberger DJ, Haile RW, Casey G, Laird PW, Le Marchand L, Thibodeau SN (2011) Frequency of Deletions of EPCAM (TACSTD1) in MSH2-Associated Lynch Syndrome Cases. The Journal of Molecular Diagnostics 13(1): 93-99

S

Sampson JR, Dolwani S, Jones S, Eccles D, Ellis A, Evans DG, Frayling I, Jordan S, Maher ER, Mak T, Maynard J, Pigatto F, Shaw J, Cheadle JP (2003) Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. The Lancet 362(9377): 39-41

Schulz E, Klampfl P, Holzapfel S, Janecke AR, Ulz P, Renner W, Kashofer K, Nojima S, Leitner A, Zebisch A, Wölfler A, Hofer S, Gerger A, Lax S, Beham-Schmid C, Steinke V, Heitzer E, Geigl JB, Windpassinger C, Hoefler G, Speicher MR, Richard Boland C, Kumanogoh A, Sill H (2014) Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. Nature Communications 5: 5191

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 (2015) Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. Gastroenterology 149(3): 563-566

SEOM (2018) Sociedad Española de Oncología Médica: Las cifras del cáncer en España (Disponible en: https://seom.org/seomcms/images/stories/recursos/Las_Cifras_del_cancer_en_Espana2018.pdf).

Shia J, Zhang L, Shike M, Guo M, Stadler Z, Xiong X, Tang LH, Vakiani E, Katabi N, Wang H, Bacares R, Ruggeri J, Boland CR, Ladanyi M, Klimstra DS (2012) Secondary mutation in a coding mononucleotide tract in MSH6 causes loss of immunoexpression of MSH6 in colorectal carcinomas with MLH1/PMS2 deficiency. Modern Pathology 26: 131

Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JKV, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2006) The Consensus Coding Sequences of Human Breast and Colorectal Cancers. Science 314(5797): 268-274

Sloane MA, Nunez AC, Packham D, et al. (2015) Mosaic epigenetic inheritance as a cause of early-onset colorectal cancer. JAMA Oncology 1(7): 953-957

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

Smith JS, Costello JF (2006) A broad band of silence. Nature Genetics 38: 504

Snowsill T, Coelho H, Huxley N, Jones-Hughes T, Briscoe S, Frayling I, Hyde C (2017) Molecular testing for Lynch syndrome in people with colorectal cancer: systematic reviews and economic evaluation. Health Technol Assess 21(51)

Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A (2016) Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. PLOS Genetics 12(1): e1005756

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

Spier I, Holzapfel S, Altmüller J, Zhao B, Horpaopan S, Vogt S, Chen S, Morak M, Raeder S, Kayser K, Stienen D, Adam R, Nürnberg P, Plotz G, Holinski-Feder E, Lifton RP, Thiele H, Hoffmann P, Steinke V, Aretz S (2015) Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. International Journal of Cancer 137(2): 320-331

Stoffel EM, Koeppe E, Everett J, Ulintz P, Kiel M, Osborne J, Williams L, Hanson K, Gruber SB, Rozek LS (2018) Germline Genetic Features of Young Individuals With Colorectal Cancer. Gastroenterology 154(4): 897-905.e1

Stupart DA, Goldberg PA, Algar U, Ramesar R (2009) Surveillance colonoscopy improves survival in a cohort of subjects with a single mismatch repair gene mutation. Colorectal Disease 11(2): 126-130

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

Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L, Vaccari EM, Bissonnette J, Booker JK, Cremona ML, Gibellini F, Murphy PD, Pineda-Alvarez DE, Pollevick GD, Xu Z, Richard G, Bale S, Klein RT, Hruska KS, Chung WK (2015) Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genetics In Medicine 18: 823

Suter CM, Martin DIK, Ward RL (2004a) Germline epimutation of MLH1 in individuals with multiple cancers. Nature Genetics 36: 497

Suter CM, Martin DIK, Ward RL (2004b) Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 36(5): 497-501

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. The American Journal Of Gastroenterology 110: 223

Syngal S, Fox EA, Li C, et al. (1999) Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: Implications for clinical predisposition testing. JAMA 282(3): 247-253

T

Takahashi Y, Wu J, Suzuki K, Martinez-Redondo P, Li M, Liao H-K, Wu M-Z, Hernández-Benítez R, Hishida T, Shokhirev MN, Esteban CR, Sancho-Martinez I, Belmonte JCI (2017) Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells. Science 356(6337): 503-508

Takeda T, Banno K, Yanokura M, Adachi M, Iijima M, Kunitomi H, Nakamura K, Iida M, Nogami Y, Umene K, Masuda K, Kobayashi Y, Yamagami W, Hirasawa A, Tominaga E, Susumu N, Aoki D (2016) Methylation Analysis of DNA Mismatch Repair Genes Using DNA Derived from the Peripheral Blood of Patients with Endometrial Cancer: Epimutation in Endometrial Carcinogenesis. Genes 7(10): 86

Taupin D, Lam W, Rangiah D, McCallum L, Whittle B, Zhang Y, Andrews D, Field M, Goodnow CC, Cook MC (2015) A deleterious RNF43 germline mutation in a severely affected serrated polyposis kindred. Human Genome Variation 2: 15013

The Cancer Genome Atlas N (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487: 330

Theodoratou E, Campbell H, Tenesa A, Houlston R, Webb E, Lubbe S, Broderick P, Gallinger S, Croitoru EM, Jenkins MA, Win AK, Cleary SP, Koessler T, Pharoah PD, Küry S, Bézieau S, Buecher B, Ellis NA, Peterlongo P, Offit K, Aaltonen LA, Enholm S, Lindblom A, Zhou XL, Tomlinson IP, Moreno V, Blanco I, Capellà G, Barnetson R, Porteous ME, Dunlop MG, Farrington SM (2010) A large-scale meta-analysis to refine colorectal cancer risk estimates associated with MUTYH variants. British Journal Of Cancer 103: 1875

Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B, Bernstein I, Capellá G, den Dunnen JT, du Sart D, Fabre A, Farrell MP, Farrington SM, Frayling IM, Frebourg T, Goldgar DE, Heinen CD, Holinski-Feder E, Kohonen-Corish M, Robinson KL, Leung SY, Martins A, Moller P, Morak M, Nystrom M, Peltomaki P, Pineda M, Qi M, Ramesar R, Rasmussen LJ, Royer-Pokora B, Scott RJ, Sijmons R, Tavtigian SV, Tops CM, Weber T, Wijnen J, Woods MO, Macrae F, Genuardi M, on behalf of InSi GHT (2013) Application of a 5-tiered scheme for

standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nature Genetics 46: 107

Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine MP, Soret J, Tazi J, Frébourg T, Tosi M (2008) A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Human Mutation 29(12): 1412-1424

U

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HFA, Hawk ET, Barrett JC, Freedman AN, Srivastava S (2004) Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. Journal of the National Cancer Institute 96(4): 261-268

V

Valle L (2017) Recent Discoveries in the Genetics of Familial Colorectal Cancer and Polyposis. Clinical Gastroenterology and Hepatology 15(6): 809-819

Valle L, Carbonell P, Fernandez V, Dotor AM, Sanz M, Benitez J, Urioste M (2007) MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. Clinical Genetics 71(3): 232-237

Valle L, Hernández-Illán E, Bellido F, Aiza G, Castillejo A, Castillejo M-I, Navarro M, Seguí N, Vargas G, Guarinos C, Juarez M, Sanjuán X, Iglesias S, Alenda C, Egoavil C, Segura Á, Juan M-J, Rodriguez-Soler M, Brunet J, González S, Jover R, Lázaro C, Capellá G, Pineda M, Soto JL, Blanco I (2014) New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Human Molecular Genetics 23(13): 3506-3512

van Roon EH, van Puijenbroek M, Middeldorp A, van Eijk R, de Meijer EJ, Erasmus D, Wouters KA, van Engeland M, Oosting J, Hes FJ, Tops CM, van Wezel T, Boer JM, Morreau H (2010) Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? BMC Cancer 10(1): 180

Vargas-Parra G, Navarro M, Pineda M, Capellá G (2018) The Molecular Basis of Lynch-like Syndrome. In Hereditary Colorectal Cancer Genetic Basis and Clinical Implications, Valle L, Gruber SB, Capellá G (eds), Chapter 2, pp 21-41. Springer

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

Vasen H, Mecklin J, Khan P, Lynch H (1991) The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 34(5): 424-5

Vasen H, Watson P, Mecklin J, Lynch H (1999) New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 116(6): 1453-6

Vasen HFA, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, Bernstein I, Bertario L, Burn J, Capella G, Colas C, Engel C, Frayling IM, Genuardi M, Heinimann K, Hes FJ, Hodgson SV, Karagiannis JA, Lalloo F, Lindblom A, Mecklin J-P, Møller P, Myrhoj T, Nagengast FM, Parc Y, Ponz de Leon M, Renkonen-Sinisalo L, Sampson JR, Stormorken A, Sijmons RH, Tejpar S, Thomas HJW, Rahner N, Wijnen JT, Järvinen HJ, Möslein G (2013) Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 62(6): 812-823

Venkatachalam R, Ligtenberg MJL, Hoogerbrugge N, Schackert HK, Görgens H, Hahn M-M, Kamping EJ, Vreede L, Hoenselaar E, van der Looij E, Goossens M, Churchman M, Carvajal-Carmona L, Tomlinson IPM, de Bruijn DRH, Van Kessel AG, Kuiper RP (2010) Germline Epigenetic Silencing of the Tumor Suppressor Gene PTPRJ in Early-Onset Familial Colorectal Cancer. Gastroenterology 139(6): 2221-2224

Venkatachalam R, Verwiel ETP, Kamping EJ, Hoenselaar E, Görgens H, Schackert HK, van Krieken JHJM, Ligtenberg MJL, Hoogerbrugge N, van Kessel AG, Kuiper RP (2011) Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. International Journal of Cancer 129(7): 1635-1642

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Smits AMM, Bos JL (1988) Genetic Alterations during Colorectal-Tumor Development. New England Journal of Medicine 319(9): 525-532

Volikos E, Robinson J, Aittomäki K, Mecklin J-P, Järvinen H, Westerman AM, de Rooji FWM, Vogel T, Moeslein G, Launonen V, Tomlinson IPM, Silver ARJ, Aaltonen LA (2006) LKB1 exonic and whole gene deletions are a common cause of Peutz-Jeghers syndrome. Journal of Medical Genetics 43(5): e18-e18

W

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

Walsh MD, Buchanan DD, Pearson S-A, Clendenning M, Jenkins MA, Win AK, Walters RJ, Spring KJ, Nagler B, Pavluk E, Arnold ST, Goldblatt J, George J, Suthers GK, Phillips K, Hopper JL, Jass JR, Baron JA, Ahnen DJ, Thibodeau SN, Lindor N, Parry S, Walker NI, Rosty C, Young JP (2012) Immunohistochemical testing of conventional adenomas for loss of expression of mismatch repair proteins in Lynch syndrome mutation carriers: a case series from the Australasian site of the colon cancer family registry. Modern Pathology 25: 722

Walsh MD, Buchanan DD, Walters R, Roberts A, Arnold S, McKeone D, Clendenning M, Ruszkiewicz AR, Jenkins MA, Hopper JL, Goldblatt J, George J, Suthers GK, Phillips K, Young GP, Macrae F, Drini M, Woods MO, Parry S, Jass JR, Young JP (2009) Analysis of families with Lynch syndrome complicated by advanced serrated neoplasia: the importance of pathology review and pedigree analysis. Familial Cancer 8(4): 313

Wang X, Fan J, Liu D, Fu S, Ingvarsson S, Chen H (2011) Spreading of Alu Methylation to the Promoter of the MLH1 Gene in Gastrointestinal Cancer. PLOS ONE 6(10): e25913

Ward RL, Dobbins T, Lindor NM, Rapkins RW, Hitchins MP (2013a) Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. Genet Med 15(1): 25-35

Ward RL, Hicks S, Hawkins NJ (2013b) Population-Based Molecular Screening for Lynch Syndrome: Implications for Personalized Medicine. Journal of Clinical Oncology 31(20): 2554-2562

Warthin AS (1985) Classicsin oncology: 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. CA: A Cancer Journal for Clinicians 35(6): 348-359

Watson AJM, Collins PD (2011) Colon Cancer: A Civilization Disorder. Digestive Diseases 29(2): 222-228

Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, French AJ, Thibodeau SN, Jass J, Haile R, Laird PW (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nature Genetics 38: 787

Weren RDA, Ligtenberg MJL, Kets CM, de Voer RM, Verwiel ETP, Spruijt L, van Zelst-Stams WAG, Jongmans MC, Gilissen C, Hehir-Kwa JY, Hoischen A, Shendure J, Boyle EA, Kamping EJ, Nagtegaal ID, Tops BBJ, Nagengast FM, Geurts van Kessel A, van Krieken JHJM, Kuiper RP, Hoogerbrugge N (2015) A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nature Genetics 47: 668

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 (2014) Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'Care for CMMRD' (C4CMMRD). Journal of Medical Genetics 51(6): 355-365

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

Win AK, Dowty JG, Cleary SP, Kim H, Buchanan DD, Young JP, Clendenning M, Rosty C, MacInnis RJ, Giles GG, Boussioutas A, Macrae FA, Parry S, Goldblatt J, Baron JA,

Burnett T, Le Marchand L, Newcomb PA, Haile RW, Hopper JL, Cotterchio M, Gallinger S, Lindor NM, Tucker KM, Winship IM, Jenkins MA (2014) Risk of Colorectal Cancer for Carriers of Mutations in MUTYH, With and Without a Family History of Cancer. Gastroenterology 146(5): 1208-1211.e5

Win AK, Jenkins MA, Dowty JG, Antoniou AC, Lee A, Giles GG, Buchanan DD, Clendenning M, Rosty C, Ahnen DJ, Thibodeau SN, Casey G, Gallinger S, Le Marchand L, Haile RW, Potter JD, Zheng Y, Lindor NM, Newcomb PA, Hopper JL, MacInnis RJ (2017) Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. Cancer Epidemiology Biomarkers & Prevention 26(3): 404-412

X

Xiao Y, Freeman GJ (2015) The Microsatellite Instable Subset of Colorectal Cancer Is a Particularly Good Candidate for Checkpoint Blockade Immunotherapy. Cancer Discovery 5(1): 16-18

Y

Yamamoto H, Imai K (2015) Microsatellite instability: an update. Archives of Toxicology 89(6): 899-921

Yan HHN, Lai JCW, Ho SL, Leung WK, Law WL, Lee JFY, Chan AKW, Tsui WY, Chan ASY, Lee BCH, Yue SSK, Man AHY, Clevers H, Yuen ST, Leung SY (2017) RNF43 germline and somatic mutation in serrated neoplasia pathway and its association with BRAF mutation. Gut 66(9): 1645-1656

Yoshida R, Miyashita K, Inoue M, Shimamoto A, Yan Z, Egashira A, Oki E, Kakeji Y, Oda S, Maehara Y (2010) Concurrent genetic alterations in DNA polymerase proofreading and mismatch repair in human colorectal cancer. European Journal Of Human Genetics 19: 320

You JF, Buhard O, Ligtenberg MJL, Kets CM, Niessen RC, Hofstra RMW, Wagner A, Dinjens WNM, Colas C, Lascols O, Collura A, Flejou JF, Duval A, Hamelin R (2010) Tumours with loss of MSH6 expression are MSI-H when screened with a pentaplex of five mononucleotide repeats. British Journal of Cancer 103(12): 1840-1845

Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kaushik P, Roa BB, Wenstrup RJ, 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.e20

Yurgelun MB, Goel A, Hornick JL, Sen A, Turgeon DK, Ruffin MT, Marcon NE, Baron JA, Bresalier RS, Syngal S, Brenner DE, Boland CR, Stoffel EM (2012) Microsatellite Instability and DNA Mismatch Repair Protein Deficiency in Lynch Syndrome Colorectal Polyps. Cancer Prevention Research 5(4): 574-582

Z

Zhang Y, Jeltsch A (2010) The Application of Next Generation Sequencing in DNA Methylation Analysis. Genes 1(1): 85

Zhang Y, Shu J, Si J, Shen L, Estecio MRH, Issa J-PJ (2012) Repetitive elements and enforced transcriptional repression co-operate to enhance DNA methylation spreading into a promoter CpG-island. Nucleic Acids Research 40(15): 7257-7268

ANEXO

INFORME DE LOS DIRECTORES SOBRE EL FACTOR DE IMPACTO DE LOS ARTÍCULOS PUBLICADOS

Gabriel Capellá y Marta Pineda, en calidad de directores de la tesis doctoral de Estela Dámaso Riquelme, titulada: "Identificación y caracterización de alteraciones epigenéticas responsables de síndrome de Lynch y Lynch-like",hacen constar que la doctoranda ha participado activamente en el diseño y realización experimental de los trabajos que se incluyen en esta tesis, el análisis de resultados, su discusión y en la preparación de los manuscritos finales. La contribución concreta de cada trabajo s, así como el factor de impacto de los artículos científicos publicados se especifican a continuación.

<u>ARTÍCULO 1</u>

Elucidating the molecular basis of MSH2-deficient tumors in suspected Lynch Syndrome cases.

*Gardenia M. Vargas-Parra, 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 Holinski-Feder, Joan Brunet, Lídia Feliubadaló, Conxi Lázaro, Matilde Navarro, Marta Pineda, Gabriel Capellá.*

- **International Journal of Cancer,** 141: 1365-1380
- **Factor de impacto (2017):** 6,513

Contribución del doctorando: Cultivo de linfocitos, extracción de RNA y retrotranscripción. Gestión de los bloques de parafina, corte y tinción de laminillas, macrodisección y extracción del DNA del tejido parafinado. Diseño y validación de los resultados obtenidos mediante el panel de NGS. Revisión del manuscrito.

Un borrador previo de este artículo fue incluido en la tesis doctoral de la Dra. Gardenia Maria Vargas Parra (Programa de doctorado en Genética, Universidad de Barcelona, 2015). La contribución de la Dra. Vargas y la doctorando en el estudio fue diferente

ARTÍCULO 2

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.*

- **Manuscrito en preparación**

Contribución del doctorando: Coordinación y recogida de muestras de centros externos. Gestión de los bloques de parafina, corte y tinción de laminillas, macrodisección y extracción del DNA del tejido parafinado. Enriquecimiento y preparación de librerías del panel NGS. Llamada de variantes, filtrado y anotación de resultados NGS. Predicción in-silico de las variantes encontradas por el panel NGS. Diseño y validación de los resultados obtenidos mediante el panel de NGS. Estudio de MLPA e interpretación de resultados. Elaboración de las matrices de comparación del estudio del metiloma. Análisis, interpretación y discusión de resultados. Preparación de tablas y figuras. Escritura de la primera versión delmanuscrito.

ARTÍCULO 3

Primary constitutional *MLH1* epimutations: a focal epigenetic event

Estela Dámaso*, Adela Castillejo, María del Mar Arias, Julia Canet-Hermida, Matilde Navarro, Jesús del Valle, Olga Campos, Anna Fernández, Fatima Marín, Daniela Turchetti, Juan de Dios García-Díaz,Conxi Lázaro, Maurizio Genuardi, Daniel Rueda, Ángel Alonso, Jose Luis Soto, Megan Hitchins, Marta Pineda*, Gabriel Capellá*.*
(*) Estos autores contribuyeron por igual en este trabajo y comparten la autoría.

- **British Journal of Cancer,** *in press*
- **Factor de impacto (2017):** 6,176

Contribución del doctorando: Coordinación y recogida de muestras de centros externos. Gestión de los bloques de parafina, corte y tinción de laminillas, macrodisección y extracción del DNA del tejido parafinado. Diseño y puesta a punto de ensayos. Secuenciación y estudios de metilación (secuenciación con bisulfito, MS-MLPA). Secuenciación clonal con bisulfito. Inmortalización de linfocitos. Realización de ensayos de impacto funcional. Estudio de cosegregación en una familia. Análisis e interpretación de resultados de array CGH. Elaboración de las matrices de comparación

del estudio del metiloma. Análisis, interpretación y discusión de resultados. Preparación de tablas y figuras. Escritura de la primera versión del manuscrito.

ARTÍCULO 4

Highly sensitive *MLH1* methylation analysis in blood allows the identification of low-level epigenetic mosaicism

Estela Dámaso*, *Julia Canet-Hermida*, Gardenia Vargas-Parra, Àngela Velasco, Esther Darder, Anna Fernández, Fátima Marin, Ángel Izquierdo, Bernat Queralt, V. Piñol, H. Uchima, José Luis Soto, Megan Hitchins, Conxi Lázaro, Joan Brunet, Marta Pineda*, Gabriel Capellá**
(*) Estos autores contribuyeron por igual en este trabajo y comparten la autoría.

- **Manuscrito en preparación**

Contribución del doctorando: Propuesta de la hipótesis. Diseño de experimentos. Obtención, gestión y tratamiento de las muestras. Cultivo primario de linfocitos y fibroblastos. Extracción de RNA. Parte de la secuenciación clonal con bisulfito. Análisis e interpretación de resultados de array CGH y SNP array. Estudio de tejidos mediante Infinium 450K. Análisis, interpretación y discusión de resultados. Preparación de figuras. Escritura de la primera versión del manuscrito.

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