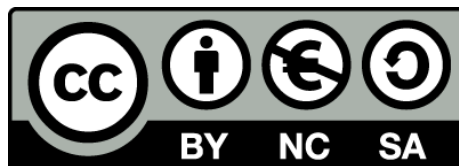




Estudio de la longitud telomérica e identificación de nuevos genes causales en el cáncer colorrectal hereditario no polipósico

Nuria Seguí Gracia



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – Compartirlqual 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – Compartirlqual 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-ShareAlike 3.0. Spain License.**



ESTUDIO DE LA LONGITUD TELOMÉRICA E IDENTIFICACIÓN DE NUEVOS GENES CAUSALES EN EL CÁNCER COLORRECTAL HEREDITARIO NO POLIPÓSICO

Memoria presentada por

Nuria Seguí Gracia

Para optar al Grado de

Doctor por la Universidad de Barcelona

Tesis realizada bajo la dirección de la

Dra. Laura Valle Velasco

en el Instituto Catalán de Oncología del

Instituto de Investigación Biomédica de Bellvitge

(ICO-IDIBELL)

Tesis adscrita a la Facultad de Biología, Universidad de Barcelona (**UB**)

Programa de Doctorado en Genética

Tutor: **Gemma Marfany Nadal**

Laura Valle Velasco

Gemma Marfany Nadal

Nuria Seguí Gracia

Barcelona, 2014

ABREVIATURAS	vii
INTRODUCCIÓN	1
1. EL CÁNCER	3
1.1. El cáncer como enfermedad genética	3
1.2. Capacidades adquiridas durante la carcinogénesis	4
2. EL CÁNCER COLORRECTAL	6
2.1. Epidemiología.....	6
2.2. Anatomía y epitelio intestinal.....	7
2.3. Estadificación del cáncer colorrectal	9
2.4. La carcinogénesis colorrectal	10
3. PREDISPOSICIÓN AL CÁNCER COLORRECTAL.....	13
3.1. Etiología: conjunción de genes y ambiente.....	13
3.2. El cáncer colorrectal familiar	16
4. SÍNDROMES POLIPÓSICOS.....	19
4.1. Poliposis adenomatosa familiar	19
4.2. Poliposis asociada a <i>MUTYH</i>	19
4.3. Poliposis asociada a la actividad reparadora de la polimerasa	21
4.4. Poliposis hereditaria mixta	21
4.5. Poliposis hamartomatosas	22
5. SÍNDROME DE LYNCH	23
5.1. Manifestaciones y variantes clínicas	23
5.2. Base genética: sistema de reparación de bases desapareadas del ADN.....	24
5.3. Identificación de familias con síndrome de Lynch	26
a) Cribado clínico.....	26
b) Cribado molecular en tumor	28
5.4. Vigilancia médica y profilaxis	29
5.5. Anticipación genética.....	30
6. CÁNCER COLORRECTAL FAMILIAR TIPO X	31
6.1. Características clínicas	31
6.2. Características moleculares	32
7. BÚSQUEDA DE GENES CAUSALES DE ENFERMEDADES MENDELIANAS	34
7.1. Análisis de ligamiento	34
7.2. Análisis directo de genes candidatos	35
7.3. Secuenciación de genomas y exomas completos.....	36
8. TELÓMEROS.....	39

8.1. Los telómeros y la telomerasa	39
8.2. Los telómeros y la telomerasa en cáncer	40
8.3. Alteraciones en la longitud telomérica y la predisposición al cáncer.....	41
HIPÓTESIS	45
OBJETIVOS	49
RESULTADOS	53
ARTÍCULO 1: “Longitud de los telómeros y anticipación genética en el síndrome de Lynch”	57
ARTÍCULO 2: “Asociación de telómeros más largos con el riesgo de cáncer en el cáncer colorrectal hereditario no polipósico sin defectos en la maquinaria de reparación de bases desapareadas”	69
ARTÍCULO 3: “GALNT12 no contribuye de manera relevante al cáncer colorrectal familiar tipo X”	75
ARTÍCULO 4: “Identificación de mutaciones en MUTYH mediante secuenciación de exomas en una familia con un fenotipo atípico”	81
ARTÍCULO 5: “Mutaciones en la línea germinal en GEN1X causan cáncer colorrectal hereditario a través de deficiencias en la reparación del ADN”	93
DISCUSIÓN	123
1. ANÁLISIS DE LA LONGITUD TELOMÉRICA EN EL CÁNCER COLORRECTAL HEREDITARIO NO POLIPÓSICO	125
1.1. Aspectos a considerar sobre la cuantificación de la longitud de los telómeros	125
1.2. Longitud telomérica en el cáncer colorrectal hereditario no polipósico	127
1.2.1. Longitud telomérica en el síndrome de Lynch	128
1.2.2. Longitud telomérica en el cáncer colorrectal familiar tipo X	129
1.2.3. Equilibrio de la longitud telomérica en la viabilidad celular	130
1.3. Anticipación genética y longitud telomérica en síndrome de Lynch	133
1.3.1. Aspectos generales de la anticipación genética.....	133
1.3.2. Valoración de la implicación de la longitud telomérica en la anticipación observada en el síndrome de Lynch	134
1.3.3. Otros posibles mecanismos que expliquen la anticipación genética en el síndrome de Lynch.....	135
2. ESTUDIO DE GALNT12 COMO GEN CAUSAL EN CÁNCER COLORRECTAL FAMILIAR TIPO X.....	137
2.1. Valoración de GALNT12 como gen candidato	137
2.2. Evaluación de las variantes de GALNT12 identificadas	138
2.3. Contribución de GALNT12 a la predisposición al CCR	141
3. LA SECUENCIACIÓN DE EXOMAS COMO HERRAMIENTA PARA DESCIFRAR LAS CAUSAS DEL CÁNCER COLORRECTAL FAMILIAR TIPO X.....	144

4. MUTACIONES EN <i>MUTYH</i> ENMASCARADAS BAJO UN FENOTIPO INUSUAL	147
4.1. Valoración de las variantes identificadas por secuenciación de exomas	147
4.1.1. Riesgo de cáncer en portadores heterocigotos de mutaciones en <i>MUTYH</i>	148
4.1.2. Estudio funcional de la variante novel	149
4.2. Características clínicas y moleculares de la familia	150
5. IDENTIFICACIÓN DE UN NUEVO GEN DE CÁNCER COLORRECTAL HEREDITARIO	153
5.1. Hallazgos de la secuenciación de exomas en una familia con cáncer colorrectal familiar tipo X.....	153
5.2. Búsqueda de mutaciones en otras familias con CCR hereditario	154
5.3. Evaluación de las mutaciones encontradas en <i>GEN1X</i>	154
5.3.1. Frecuencias poblacionales	154
5.3.2. Estudio de segregación con la enfermedad	155
5.3.3. Evidencias de afectación de la función y posible patogenicidad de las variantes identificadas.....	155
5.3.4. Ensayo de viabilidad celular en respuesta a agentes de daño celular	157
5.4. Papel del gen en la carcinogénesis colorrectal	157
5.4.1. Gen supresor de tumores: Búsqueda del segundo evento a nivel somático.....	157
5.4.2. Gen implicado en la reparación del daño genético: ¿Acumulación de errores en el ADN?	158
5.4.3. Participación de <i>GEN1X</i> en la progresión tumoral	160
5.4.4. <i>GEN1X</i> como gen de alta penetrancia en el cáncer colorrectal	161
6. IDENTIFICACIÓN DE LAS CAUSAS GENÉTICAS DEL CÁNCER COLORRECTAL FAMILIAR TIPO X.....	162
CONCLUSIONES	165
BIBLIOGRAFÍA	169
ANEXOS	201
ANEXO I: Otras publicaciones.....	203
<i>"No association between germline allele-specific expression of TGFBR1 and colorectal cancer risk in Caucasian and Ashkenazi populations"</i>	205
<i>"Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome"</i>	217
<i>"New insights into POLE and POLD1 germline mutations in familial colorectal cancer and poliposis"</i>	225
ANEXO II: Informes del director.....	237

ABREVIATURAS

3D	Tridimensional
ADN	Ácido desoxirribonucleico
ALT	<i>Alternative lengthening of telomeres</i> ; Vía alternativa de elongación de los telómeros
ARN	Ácido ribonucleico
CCR	Cáncer colorrectal
CCRF-X	Cáncer colorrectal familiar tipo X
CIMP	<i>CpG island methylator phenotype</i> ; Fenotipo metilador de islas CpG
CSCE	<i>Conformation Sensitive Capillary Electrophoresis</i> ; Electroforesis capilar sensible a conformación
EMS	Estabilidad de microsatélites
IHQ	Inmunohistoquímica
IMS	Inestabilidad de microsatélites
Mb	Megabase
MMC	Mitomicina C
MMR	<i>Mismatch repair</i> ; Reparación de bases desapareadas
PCR	<i>Polymerase chain reaction</i> ; Reacción en cadena de la polimerasa
GALNT	UDP-GalNAc:polipéptido N-acetilgalactosamina-transferasas
Q-FISH	Hibridación fluorescente in situ cuantitativa
shRNA	<i>Short hairpin RNA</i> , ARN de horquilla pequeña
SL	Síndrome de Lynch
SNP	<i>Single Nucleotide Polymorphism</i> ; Polimorfismo de una sola base
TCGA	<i>The Cancer Genome Atlas</i> ; Atlas del genoma del cáncer
NGS	<i>Next-generation sequencing</i> ; Secuenciación de nueva generación o secuenciación masiva en paralelo

GENES (nomenclatura estándar procedente del inglés)

<i>APC</i>	<i>Adenomatous Polyposis Coli</i>
<i>BMPR1A</i>	<i>Bone morphogenetic protein receptor, type IA</i>
<i>BRAF</i>	<i>V-raf murine sarcoma viral oncogene homolog B</i>
<i>BRCA1</i>	<i>Breast cancer 1, early onset</i>
<i>BUB1</i>	<i>BUB1 mitotic checkpoint serine/threonine kinase</i>
<i>BUB3</i>	<i>BUB3 mitotic checkpoint protein</i>
<i>CHEK2</i>	<i>Checkpoint kinase 2</i>

Abreviaturas

<i>DACT2</i>	<i>Dishevelled-binding antagonist of beta-catenin 2</i>
<i>DARC</i>	<i>Duffy blood group, atypical chemokine receptor</i>
<i>DCC</i>	<i>Deleted in colorectal carcinoma</i>
<i>DNMT3B</i>	<i>DNA (cytosine-5-)-methyltransferase 3 beta</i>
<i>EPCAM</i>	<i>Epithelial cell adhesion molecule</i>
<i>FAM55B</i>	<i>Neurexophilin and PC-esterase domain family, member 2</i>
<i>FANCA</i>	<i>Fanconi anemia, complementation group A</i>
<i>GABBR2</i>	<i>Gamma-aminobutyric acid B receptor, 2</i>
<i>GALNT12</i>	<i>UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 12</i>
<i>GREM1</i>	<i>Gremlin 1, DAN family BMP antagonist</i>
<i>GSTM1</i>	<i>Glutathione S-transferase mu 1</i>
<i>HABP4</i>	<i>Hyaluronan Binding Protein 4</i>
<i>KRAS</i>	<i>Kirsten rat sarcoma viral oncogene homolog</i>
<i>MCC</i>	<i>Mutated in colorectal cancers</i>
<i>MDM2</i>	<i>MDM2 oncogene, E3 ubiquitin protein ligase</i>
<i>MET</i>	<i>Met proto-oncogene</i>
<i>MGMT</i>	<i>O-6-methylguanine-DNA methyltransferase</i>
<i>MLH1</i>	<i>MutL homolog 1</i>
<i>MSH2</i>	<i>MutS homolog 2</i>
<i>MSH6</i>	<i>MutS homolog 6</i>
<i>MUTYH</i>	<i>MutY homolog</i>
<i>MYC</i>	<i>V-myc avian myelocytomatosis viral oncogene homolog</i>
<i>NUDT1</i>	<i>Nudix (nucleoside diphosphate linked moiety X)-type motif 1</i>
<i>OGG1</i>	<i>8-oxoguanine DNA glycosylase</i>
<i>P16 (CDKN2A)</i>	<i>cyclin-dependent kinase inhibitor 2A</i>
<i>PIK3CA</i>	<i>Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha</i>
<i>PKC</i>	<i>Protein kinase C</i>
<i>PMS2</i>	<i>postmeiotic segregation increased 2 (S. cerevisiae)</i>
<i>POLD1</i>	<i>Polymerase (DNA directed), delta 1, catalytic subunit</i>
<i>POLE</i>	<i>Polymerase (DNA directed), epsilon, catalytic subunit</i>
<i>PTEN</i>	<i>Phosphatase and tensin homolog</i>
<i>PTPRJ</i>	<i>Protein tyrosine phosphatase, receptor type, J</i>
<i>RACK1</i>	<i>Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1</i>
<i>RAS</i>	<i>Rat Sarcoma Viral Oncogene Homolog</i>

<i>RB1</i>	<i>Retinoblastoma 1</i>
<i>SMAD4</i>	<i>SMAD family member 4</i>
<i>STK11 (LKB1)</i>	<i>Serine/threonine kinase 11</i>
<i>TERC</i>	<i>Telomerase RNA component</i>
<i>TERT</i>	<i>Telomerase reverse transcriptase</i>
<i>TGFBR1</i>	<i>Transforming growth factor, beta receptor 1</i>
<i>TP53</i>	<i>Tumor protein p53</i>
<i>UNC5C</i>	<i>Unc-5 homolog C (C. elegans)</i>
<i>ZNF367</i>	<i>Zinc finger protein 367</i>

INTRODUCCIÓN

1. EL CÁNCER

Se denomina cáncer al conjunto de enfermedades caracterizadas por una proliferación celular descontrolada que puede derivar en una diseminación de estas células anómalas a otras zonas del cuerpo, proceso conocido como metástasis. El cáncer es la segunda causa de mortalidad en el mundo tras las enfermedades cardíacas.

1.1. El cáncer como enfermedad genética

La carcinogénesis es un proceso que se produce en varios pasos, cada uno de ellos determinado por alteraciones genéticas que dirigen la transformación progresiva de células humanas normales hasta sus derivados malignos con capacidad invasiva. Por tanto, podemos decir que el cáncer es una enfermedad genética, causada principalmente por mutaciones en las células somáticas. Aun así, si estos cambios se producen en línea germinal, estas alteraciones genéticas se pueden transmitir a la descendencia, confiriendo a los individuos portadores una predisposición al cáncer heredable (Knudson, 2002).

Desde los años setenta, cuando Knudson describió la relación entre el retinoblastoma, un tumor de la retina ocular, y las mutaciones en el gen *RB1* (Knudson et al., 1976; Friend et al., 1986), se han descrito a lo largo de las últimas décadas múltiples genes implicados en la carcinogénesis de distintos tipos de tumores, así como las vías moleculares a través de las cuales actúan. Los genes implicados en cáncer se pueden clasificar en tres categorías: Oncogenes, genes supresores de tumores y genes reparadores (Vogelstein y Kinzler, 2004). Mutaciones en las dos primeras categorías de genes actúan regulando directamente la proliferación celular, ya sea mediante la estimulación del crecimiento celular o por inhibición de la muerte celular. En el caso de los oncogenes, como por ejemplo *RAS*, *MYC* y *MET*, una mutación causa su activación constitutiva, es decir, una ganancia de función. La mutación de un solo alelo es generalmente suficiente para que un oncogen aporte una ventaja selectiva de crecimiento a una célula (Figura 1). Contrariamente, mutaciones en los genes supresores de tumores reducen su actividad, y hacen falta mutaciones en ambas copias del gen para conferir la ventaja selectiva. Se trata por tanto de una pérdida de función de naturaleza recesiva. Ejemplos de este tipo de genes son *TP53*, *APC* o *PTEN* (Figura 1).

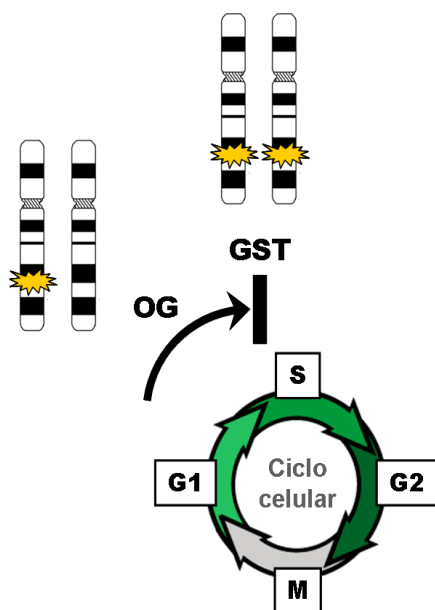


Figura 1. Alteración del ciclo celular por oncogenes (OG) y genes supresores de tumores (GST). Mutaciones en una copia de los OG altera su función, mientras que los GST requieren la deficiencia de las dos copias del gen.

Finalmente, los genes de reparación del ADN, también considerados como un subtipo de genes supresores de tumores, cuando están alterados promueven la tumorigénesis por una vía indirecta. Estos genes codifican proteínas encargadas del mantenimiento de la integridad del genoma, tales como *MLH1*, *MUTYH* o *BRCA1*. Su inactivación provoca una mayor tasa de mutación en otros genes debido a una acumulación de errores en el ADN causada por una reparación de errores deficiente. Como en el caso de los genes supresores de tumores, son necesarias mutaciones en los dos alelos del gen para inactivarse. Sin embargo, la inactivación de los dos alelos del gen reparador no es suficiente para el inicio de la carcinogénesis, sino que harán falta las consecuentes mutaciones somáticas en oncogenes o genes supresores de tumores, causadas por la reparación defectuosa del ADN, para que se inicie el proceso neoplásico.

1.2. Capacidades adquiridas durante la carcinogénesis

Hanahan y Winberg definieron hace más de una década una serie de rasgos a nivel molecular, bioquímico y celular que son compartidos por la mayoría de los tumores humanos (Hanahan y Weinberg, 2000). Propusieron una lista de seis capacidades adquiridas durante la tumorigénesis que reflejan los cambios que dirigen la transformación progresiva de células humanas normales en células cancerosas malignas: Autosuficiencia en las señales de crecimiento, insensibilidad a las señales inhibitorias del

crecimiento, evasión de la apoptosis, potencial replicativo ilimitado, mantenimiento de la angiogénesis y, finalmente, invasión de tejidos y metástasis.

Posteriormente, incorporando los avances en el conocimiento acerca de la biología del cáncer, los mismo autores propusieron un nuevo esquema más completo (Hanahan y Weinberg, 2011). En este nuevo esquema se añaden a la lista dos características emergentes funcionalmente importantes para el desarrollo del tumor: La reprogramación de la energía metabólica y la elusión a la destrucción inmune. Asimismo se añaden dos características que hacen posible la adquisición de todas las anteriores: El desarrollo de una inestabilidad genómica que produce mutaciones en las células cancerígenas, y el estado inflamatorio de las lesiones pre-malignas inducido por el sistema inmune (Figura 2).



Figura 2. Características distintivas de la progresión tumoral. Los ocho cambios esenciales en la fisiología de la célula que dictan el crecimiento maligno y las características que los posibilitan. Adaptado de Hanahan y Weinberg, 2011.

Con la unión de todos estos conceptos se ha ido construyendo paulatinamente la visión actual de los tumores, como algo más que masas aisladas de células cancerígenas en proliferación. Los tumores son complejos tejidos compuestos por múltiples tipos celulares que incluyen también la participación de células normales que conforman el estroma asociado al tumor, conocido como microambiente tumoral.

2. EL CÁNCER COLORRECTAL

2.1. Epidemiología

A nivel mundial se diagnostican más de un millón de casos de cáncer colorrectal (CCR) al año, siendo ésta una de las principales causas de mortalidad por cáncer. El CCR se sitúa entre el segundo y cuarto tumor más frecuente en los países industrializados y su incidencia varía ampliamente según regiones geográficas (Figura 3). La mayoría de casos se dan entre los 65 y 75 años de edad, aunque también se dan casos a edades más tempranas, a menudo relacionados con una predisposición genética (Ferlay et al., 2010).

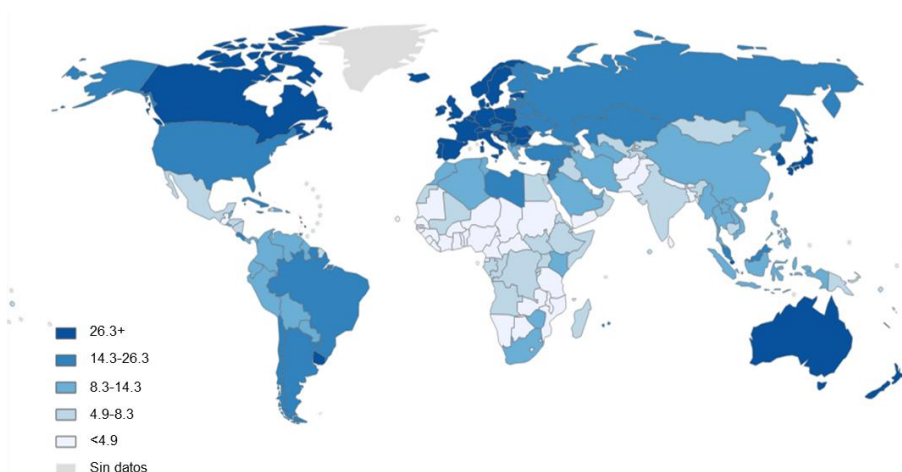


Figura 3. Mapa mundial de la incidencia del cáncer colorrectal en el año 2012. Tasa estandarizada por edad por 100.000. Imagen extraída de Globocan 2012, International Agency for Research on Cancer.

En España, con casi 30.000 casos al año, el CCR es la neoplasia más frecuente. Separando por sexos es el tercer cáncer más frecuente en hombres (15%), después del de próstata (21,7%) y de pulmón (16,9%); y el segundo en mujeres (14,9%), tras el cáncer de mama (29%). En números absolutos, en el año 2012 hubo 14.700 muertos por CCR (8.742 hombres y 5.958 mujeres). Según estas cifras el CCR representa la segunda causa de muerte por cáncer para ambos sexos en este país, después del cáncer de pulmón en hombres y de mama en mujeres (Figura 4) (Ferlay et al., 2013).

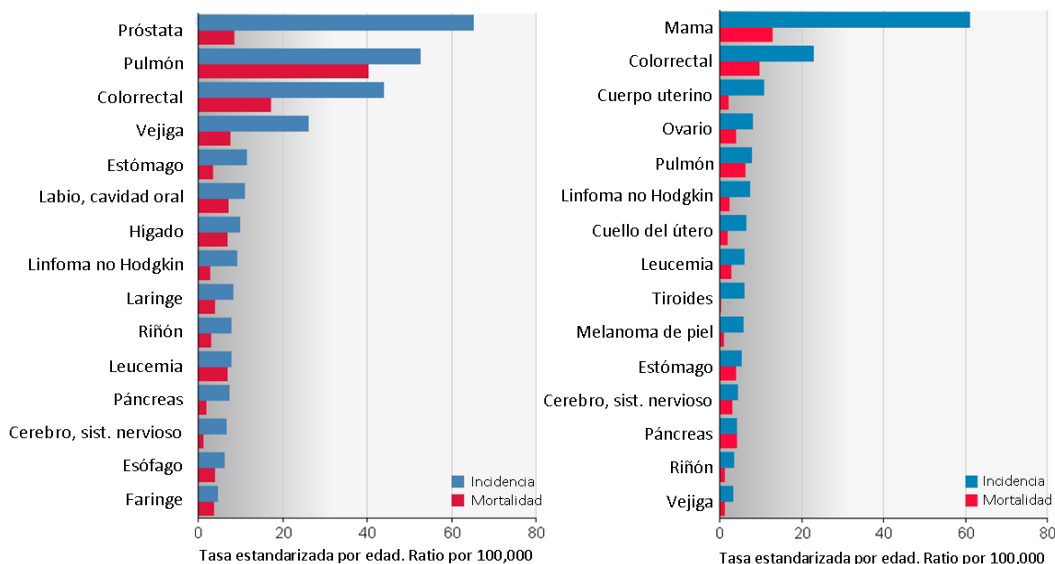


Figura 4. Incidencia y mortalidad estandarizadas por edad de las diferentes neoplasias en España en el año 2012 según sexos. Izquierda, hombres; derecha, mujeres. Imagen adaptada de GLOBOCAN 2012, International Agency for Research on Cancer.

2.2. Anatomía y epitelio intestinal

El colon y el recto forman parte del sistema digestivo o gastrointestinal. Este sistema procesa la comida ingerida y desecha el material que el cuerpo no aprovecha. Tras el esófago, estómago e intestino delgado prosigue el denominado intestino grueso, cuya primera y parte más larga corresponde al colon. A lo largo de éste órgano se absorben agua y nutrientes minerales, los desechos resultantes pasan al recto, zona final del intestino grueso, para finalmente ser expulsados por el ano. Se puede dividir el colon en cuatro secciones: colon ascendente, transversal, descendente y sigmoide; que clínicamente se pueden simplificar en dos partes: El colon proximal o derecho y el colon distal o izquierdo (Figura 5A).

Histológicamente el colon está compuesto por varias capas, diferenciando del interior hacia el exterior la mucosa, submucosa, muscularis propia, subserosa y serosa. La capa interior está a su vez formada por células epiteliales, lámina propia y muscularis mucosa (Figura 5B). Cuando se genera un tumor puede variar el número de capas que estén afectadas, factor determinante del pronóstico de la enfermedad.

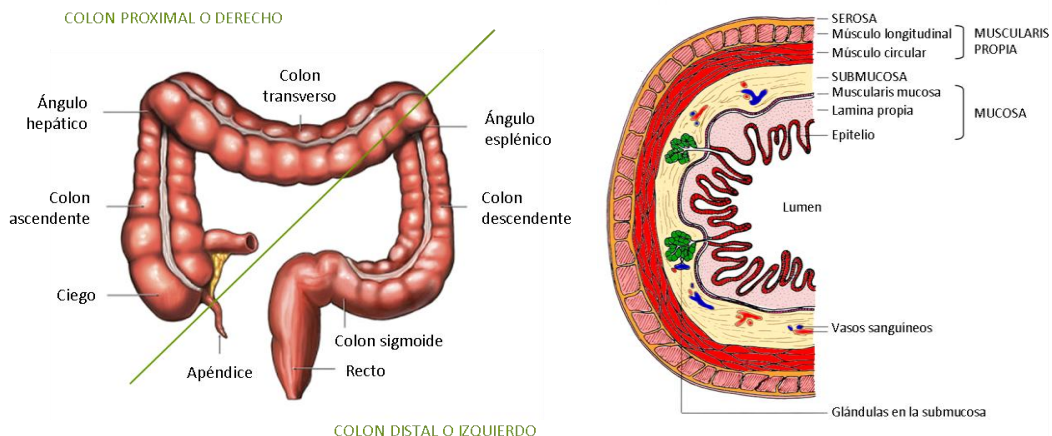


Figura 5. Esquema de la anatomía del intestino grueso. A. División anatómica y clínica del colon y recto. **B.** Capas histológicamente diferenciadas en un corte transversal del colon.

La mucosa del colon tiene la superficie plana, en contraposición al intestino delgado que presenta *villi*. La superficie del intestino grueso está organizada en estructuras tubulares llamadas criptas de Lieberkühn, cavidades epiteliales con las que gana superficie. Las células que forman esta área tienen una tasa de renovación muy elevada. En el fondo de las criptas se encuentran las células madre, que se dividen formando células pluripotentes o progenitoras no diferenciadas que migran en dirección ascendente desde la base de la cripta y que ocupan dos terceras partes de la misma. Cuando llegan a la parte superior, estas células dejan de proliferar y se diferencian en uno de los tipos celulares del colon: Colonocitos, células calciformes o células enteroendocrinas (Figura 6) (Reya y Clevers, 2005).

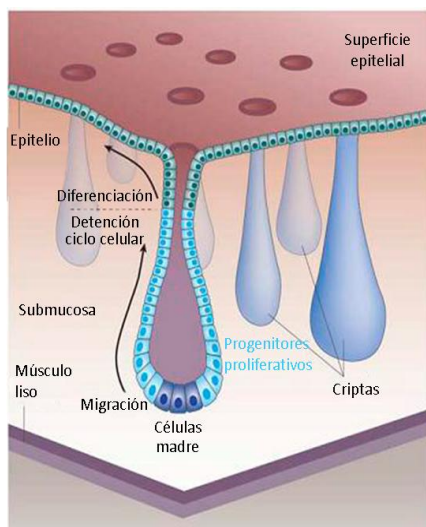


Figura 6. Esquema del epitelio colónico. Migración y etapas de diferenciación de las células epiteliales desde la base de la cripta hasta la superficie intestinal. Adaptado de Reya y Clevers, 2005.

2.3. Estadificación del cáncer colorrectal

El pronóstico de los pacientes con CCR está altamente correlacionado con el estadio en el que se diagnostica la enfermedad. Así, la supervivencia a 5 años es mayor del 90% si se detecta el tumor en estadios tempranos, pero menor del 10% si se diagnostica en estadios avanzados. Por ello, la detección precoz es el arma principal para la prevención de la mortalidad por CCR.

La estadificación del CCR se basa en el grado de invasión e infiltración del tumor primario en las diferentes capas de la pared intestinal, en la afectación de los ganglios linfáticos y en la presencia de metástasis. Dukes propuso en el año 1932 una primera clasificación en tres estadios definidos en función de estas tres características. Dada su sencillez se han realizado múltiples modificaciones de la clasificación de Dukes, siendo la más importante la de Aster y Coller en el año 1954 (Tabla II). Finalmente, con la intención de unificar sistemas, la *American Joint Committee on Cancer* propuso la clasificación TNM (Hutter y Sobin, 1986). Este sistema considera por partes separadas tumor primario (T), nódulos linfáticos (N) y metástasis (M) (Tabla I). Estas tres categorías se combinan de manera que corresponden a un estadio concreto establecido (I-IV). La estadificación en cuatro fases (I-IV) se correlaciona asimismo con la clasificación de Dukes y puede asociarse al pronóstico de los pacientes (Tabla II).

Tabla I. Descripción de las categorías del sistema TNM para el cáncer colorrectal.

T: Extensión que ocupa el tumor primario	
Tx	El tumor primario no se puede evaluar
Tis	Carcinoma in situ. Intraepitelial o invasión de la lámina propia
T1	El tumor invade la submucosa
T2	El tumor invade la muscular propia
T3	El tumor invade la subserosa sin llegar a tejidos vecinos
T4	El tumor invade otros órganos o estructuras
N: Ausencia o presencia de afectación en ganglios linfáticos	
Nx	Información incompleta, los ganglios no pueden ser evaluados
N1	Ningún ganglio afectado
N2	De uno a tres ganglios afectados
N3	Cuatro o más ganglios afectados
M: Ausencia o presencia de metástasis a distancia	
Mx	No se puede valorar la metástasis a distancia
M0	No hay metástasis a distancia
M1	Presencia de metástasis a distancia

Tabla II. Estadificación para el cáncer colorrectal según los diferentes métodos de clasificación y pronóstico de los pacientes.

Estadio	TNM	Dukes	Aster y Coller	Supervivencia a los 5 años
0	Tis	-	-	96-100%
I	T1N0M0 T2N0M0	A	A B1	80-95%
II	T3N0M0 T4N0M0	B	B2 B3	50-75%
III	T_N1M0 T_N2M0	C	C1 C2	25-50%
IV	T_N_M1	-	D	<7%

2.4. La carcinogénesis colorrectal

Como se ha comentado en apartados anteriores, la tumorigénesis constituye un proceso gradual que consiste en la sucesiva adquisición de alteraciones moleculares, desarrollo que puede ocurrir a lo largo de varios años. Los tumores colorrectales, dada su ubicación y accesibilidad, así como la existencia de varias fases bien diferenciadas de la enfermedad, han supuesto un buen sistema de estudio de los cambios genéticos involucrados en el desarrollo de la neoplasia. Fearon y Vogelstein propusieron en el año 1990 un primer modelo de carcinogénesis que explica la progresión del CCR desde sus formas benignas hasta las más agresivas. Este proceso se conoce como secuencia adenoma-carcinoma (Fearon y Vogelstein, 1990). Siguiendo este arquetipo, mutaciones en el gen *APC* se dan en fases tempranas, seguidas generalmente por mutaciones en los genes de la familia *RAS*; mientras que los cambios en *TP53* aparecen más adelante cronológicamente. Las alteraciones que promueven el paso a tumores metastáticos, en el caso de existir, aún permanecen desconocidas (Figura 7).

Las principales características de este modelo son:

- 1- Los tumores aparecen como resultado de mutaciones activadoras de oncogenes y de mutaciones inactivadoras de genes supresores de tumores.
- 2- Hacen falta mutaciones en al menos 4 o 5 genes para la formación de un tumor maligno.
- 3- Aunque los cambios genéticos se dan preferentemente en una secuencia determinada, es el total de las alteraciones acumuladas en vez de su orden de aparición, lo que determina las propiedades biológicas del tumor.

- 4- En ocasiones, algunos genes supresores de tumores mutados pueden ejercer un efecto fenotípico a nivel celular, ya en su estado heterocigoto.

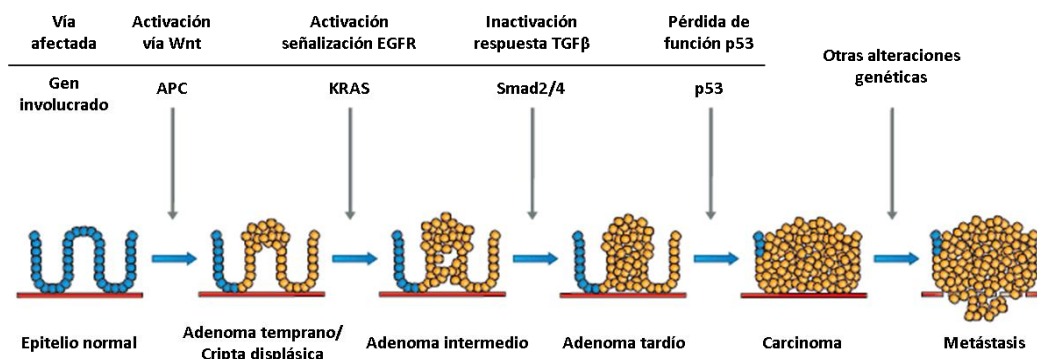


Figura 7. Secuencia adenoma-carcinoma en el cáncer colorrectal. La progresión del epitelio normal a través de adenoma a carcinoma está asociada a anomalías acumuladas en genes concretos. Adaptado de Davies et al., 2005.

La presencia de sistemas de mantenimiento y reparación del genoma hace que en condiciones normales la tasa de mutación espontánea sea muy baja a lo largo de cada generación celular. Con el fin de acumular las mutaciones génicas necesarias para avanzar en el proceso de carcinogénesis, las células tumorales incrementan su tasa de mutación, mediante distintos mecanismos, que pueden ser o no excluyentes (Salk et al., 2010). Actualmente se conocen tres mecanismos que causan inestabilidad genómica y contribuyen a la progresión tumoral: la inestabilidad cromosómica, la inestabilidad de microsatélites (IMS) y el fenotipo metilador de islas CpG (CIMP, del inglés *CpG island methylator phenotype*).

El mecanismo de la inestabilidad cromosómica es el más frecuente, presentándose en el 75-80% de todos los cánceres colorrectales, y corresponde a la presentada en el modelo de Fearon y Vogelstein. Típicamente estos tumores presentan pérdida de heterocigosidad (deleciones), amplificaciones y translocaciones cromosómicas. Estos cambios provocan la inactivación de genes supresores de tumores tales como *APC*, *TP53*, *DCC* y *SMAD4*, así como la activación de oncogenes como *KRAS* y *MYC* (Vogelstein et al., 1988; Fearon y Vogelstein, 1990).

Introducción

Los microsatélites son secuencias repetitivas del ADN, compuestas por repeticiones de uno a cinco nucleótidos. La IMS se caracteriza por un aumento o disminución de la longitud de algunas de estas secuencias en las células somáticas en comparación con la línea germinal. Esta inestabilidad resulta de la incapacidad del sistema de reparación de corregir los errores producidos durante la replicación del ADN, debido a la inactivación de los genes de reparación de bases desapareadas (MMR, del inglés *mismatch repair*) (Kinzler y Volgestein, 1996).

Finalmente, alteraciones en el patrón de metilación de las islas CpG de la región promotora de los genes causa la inactivación de su expresión. En tumores, este tipo de inactivación transcripcional ocurre en genes supresores de tumores, por ejemplo en *p16*, *MGMT*, *MLH1* o *BRCA1*. Cuando esta hipermetilación de islas CpG ocurre de manera generalizada, se origina el denominado fenotipo metilador CIMP (Toyota et al., 1999).

3. PREDISPOSICIÓN AL CÁNCER COLORRECTAL

3.1. Etiología: conjunción de genes y ambiente

Las grandes diferencias en la incidencia de CCR que hay entre países, hacen patente la participación de las costumbres y estilos de vida en el desarrollo de la enfermedad. Así mismo, estudios en grupos de poblaciones emigrantes sugieren que factores ambientales juegan un papel esencial en la etiología de la enfermedad (Boyle y Langman, 2000).

Uno de los factores ambientales que se ha asociado en estudios epidemiológicos al riesgo a desarrollar CCR es la dieta. Un alto consumo de carne roja y grasas está asociado a un mayor riesgo a padecer este tipo de cáncer, mientras que se da la correlación inversa para el consumo de fibra, vegetales y fruta. En cuanto al estilo de vida, existen evidencias que relacionan el tabaco, la ingesta de alcohol, la actividad física y el índice de masa corporal con el riesgo a desarrollar CCR. Además, también se han vinculado positiva o negativamente medicamentos como la aspirina o el tratamiento hormonal sustitutivo respectivamente (Potter et al., 1999; Bagnardi et al., 2001; Chan et al., 2010; Burn et al., 2011; Bardou et al., 2013).

Por otro lado, desde los años 50 se conocen informes de familias con una elevada agregación de tumores localizados en el intestino grueso, donde ya indicaban que los miembros de estas familias tenían un riesgo elevado de desarrollar CCR (Woolf et al., 1955; Macklin, 1960). De hecho, el riesgo a desarrollar CCR es mayor en individuos con historia familiar de cáncer de colon, sobre todo en aquellas familias con un síndrome hereditario como la poliposis adenomatosa familiar o el síndrome de Lynch (SL). Estudios caso-control sugieren que factores hereditarios contribuyen considerablemente al CCR, con razones de oportunidades (en inglés, *odds ratios*) moderadas con valores entre 2 y 4 (Johns et al., 2001). Asimismo, estudios en gemelos estiman que el efecto de los factores genéticos en el cáncer colorrectal es de hasta un 35% (Lichtenstein et al., 2000).

A nivel genético podemos clasificar los alelos de susceptibilidad según su capacidad de afectación fenotípica, lo que se denomina penetrancia (Figura 8). Las mutaciones de elevada penetrancia predisponen a desarrollar CCR con un riesgo muy alto. Suelen ser cambios únicos o muy poco frecuentes que causan los síndromes hereditarios de predisposición al CCR como la poliposis adenomatosa familiar o el SL, detallados en el siguiente apartado. Los alelos de moderada penetrancia también suelen ser variantes

genéticas poco frecuentes pero se asocian con un riesgo más bajo. Un ejemplo clásico de este tipo de alelo de susceptibilidad es la variante I1307K en el gen *APC* (Laken et al., 1997). Finalmente, también se han identificado numerosas variantes de baja penetrancia que contribuyen a la susceptibilidad del CCR de una manera aditiva. Estos son alelos comunes que se asocian a un riesgo relativo bajo. En los últimos años múltiples estudios de asociación del genoma completo o de genes candidatos en amplias series de casos y controles han localizado numerosas regiones o variantes relacionadas con el riesgo de CCR. Hasta la fecha, se han asociado al menos 41 *loci* independientes en población caucásica ($P < 1 \times 10^{-5}$ y asociación confirmada en al menos una serie caso-control independiente adicional; Tabla III) (<http://www.genome.gov/gwastudies>). Parecen seguir un modelo de predisposición poligénico, por el que individuos portadores de un número elevado de estos alelos podrían tener un riesgo tres veces mayor que aquellos con un número medio de alelos de riesgo (Houlston y COGENT, 2012). Por otro lado, las asociaciones de variantes de riesgo bajo-moderado en genes candidatos a menudo son difíciles de interpretar. Un meta-análisis realizado recientemente ha descrito asociación significativa con el CCR para un total de 62 variantes en 50 genes, 8 de ellos con una fuerte evidencia acumulativa de asociación y 2 moderada (Ma et al., 2013). Entre variantes genéticas se encuentran los genes: *MUTYH*, *APC*, *CHEK2*, *MLH1*, *DNMT3B*, *GSTM1* y *TERT*.

Tabla III. Variantes y genes identificados mediante estudios de asociación a nivel genómico asociados al CCR. Estudios con $P < 1 \times 10^{-5}$ en fases inicial y de validación. OR, *odds ratio*.

Gen mapado	SNP	OR
ATOH1 - HMGB3P15	rs13130787	1.09
C11orf93;C11orf92	rs3802842	1.11-1.14
CCND2	rs3217810	1.20
CCND2	rs3217901	1.10
CDH1	rs9929218	1.10
CHCHD3P1 - HSP90AB7P	rs10795668	1.12
DACT1 - RPL31P4	rs17094983	1.13
DCBLD1	rs2057314	1.08
DIP2B - ATF1	rs11169552	1.09
DUSP10 - QRSL1P2	rs6691170	1.06
DUSP10 - QRSL1P2	rs6687758	1.09
DUSP4 - RPL17P33	rs12548021	1.28
FGFR3P3 - BMP2	rs4813802	1.10
FGFR3P3 - BMP2	rs961253	1.12
GPR143 - HMG1P33	rs5934683	1.07
GSDMC - RPL15P12	rs2128382	1.11
HSD17B7 - C1orf110	rs1912453	1.07
HSPA12A	rs1665650	1.13
KLHL18 - PTPN23	rs8180040	1.28
KRT18P42 - MAN2A1	rs367615	1.35
LAMA5	rs4925386	1.08
LAMC1	rs10911251	1.09
LINC00536 - EIF3H	rs16892766	1.24-1.27
LOC100616530	rs3104964	1.27
LOC100996485	rs647161	1.11-1.17
MIR3925 - LAP3P2	rs1321311	1.10
MIR4291 - BARX1	rs10114408	1.37
MYNN	rs10936599	1.04
NABP1 - SDPR	rs11903757	1.16
OSBPL3 - CYCS	rs39453	1.28
POLD3	rs3824999	1.08
RHPN2	rs10411210	1.15
RPL18P9 - CCND2	rs10774214	1.09-1.17
RPS23P3 - CENPC1	rs17730929	1.47
RPS3AP46 - MIR5580	rs4444235	1.11
SALL4P5 - RPL24P7	rs4591517	1.06
SCG5 - GREM1	rs4779584	1.12-1.23
SCG5 - GREM1	rs4779584	1.18
SLC22A3	rs7758229	1.28
SMAD7	rs4939827	1.12-1.20
SRRM1P1 - POU5F1B	rs6983267	1.13-1.18
SRRM1P1 - POU5F1B	rs7014346	1.19
SRRM1P1 - POU5F1B	rs6983267	1.24-1.27
SRRM1P1 - POU5F1B	rs10505477	1.17
SRSF10P2	rs2423279	1.10-1.14
SYNJ2	rs9365723	1.27
TBX3	rs59336	1.09
TBX3 - UBA52P7	rs7315438	1.11
TPH2	rs10879357	1.25

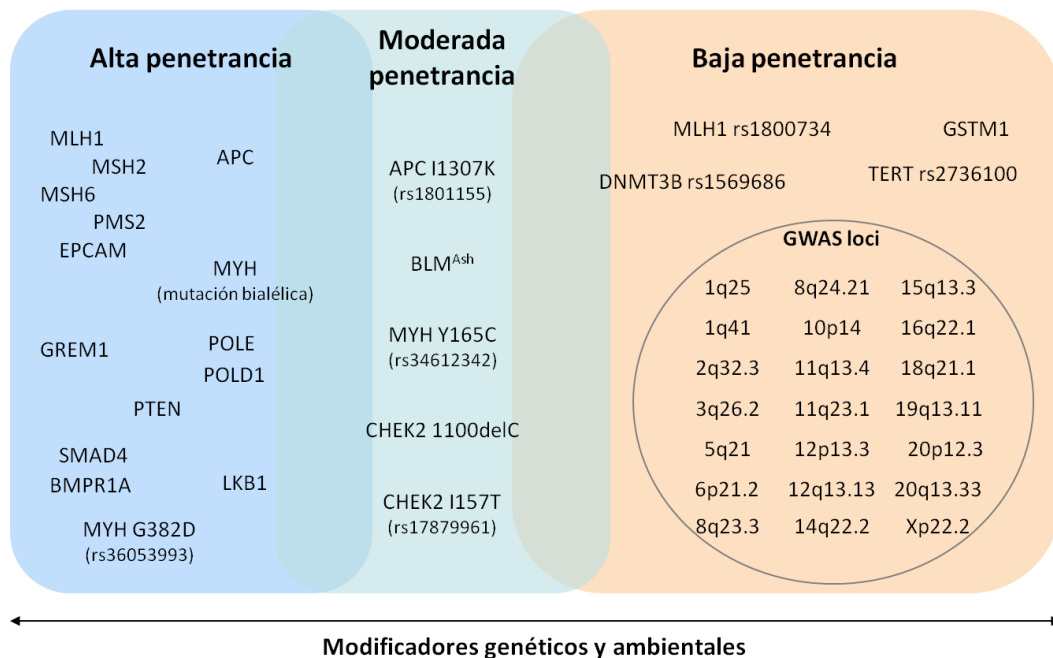


Figura 8. Contribución de los factores genéticos en la predisposición al CCR. El efecto de modificadores de riesgo, genéticos y/o ambientales, hace que los márgenes que delimitan cada grupo no queden bien definidos.

De este modo, como para la mayoría de cánceres, la distinción entre esporádico y familiar resulta un tanto borrosa, y sería conveniente visualizar la predisposición como un continuo de factores de riesgo que contribuyen al desarrollo de la neoplasia. Es por tanto el balance entre los factores genéticos y los factores ambientales lo que determina la susceptibilidad individual a desarrollar CCR, tanto en casos familiares como esporádicos (Figura 8).

3.2. El cáncer colorrectal familiar

Tradicionalmente se ha dividido el CCR en base a los antecedentes familiares, diferenciando los casos esporádicos de los familiares. De este modo, se considera CCR esporádico aquel que se da en individuos sin ningún antecedente familiar o personal a esta forma de cáncer, y supondría el resultado de mutaciones somáticas adquiridas. Se estima su proporción en un rango del 65 al 85% del total de casos (Figura 9). Por otro lado, se estima que de todos los casos de CCR un 20-25% presentan historia familiar de la enfermedad (Lichtenstein et al., 2000; de la Chapelle, 2004). Aun así, tan solo un 2-6% del total se explica por mutaciones en línea germinal de los genes de moderada-alta penetrancia conocidos (Tabla IV).

El denominado CCR familiar se caracteriza por la presencia de varios afectos en la misma familia. Estos individuos tienen un riesgo mayor al de la población general a desarrollar la enfermedad, así como a que se presente de manera más precoz. En estos grupos familiares, además de agregarse los casos explicados por las variantes de susceptibilidad conocidas, también pueden subyacer variantes de CCR hereditario que aún se desconocen. Sin embargo, la agregación familiar de algunos de estos casos también podría resultar de un ambiente o estilo de vida compartido, o incluso simplemente por azar.

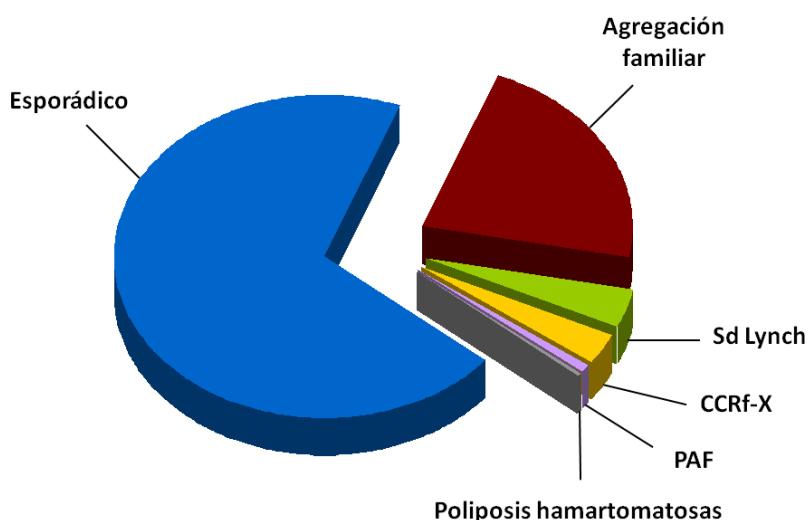


Figura 9. Clasificación de los diferentes tipos de CCR según la agregación familiar y características genéticas.

Las formas hereditarias del cáncer colorrectal pueden dividirse en polipósicas y no polipósicas. Entre los síndromes polipósicos podemos diferenciar las poliposis adenomatosas, que incluyen a la poliposis adenomatosa familiar en sus formas clásica y atenuada, la poliposis asociada a *MUTYH*, la poliposis asociada a la actividad reparadora de la polimerasa, y la poliposis hereditaria mixta; y las poliposis hamartomatosas, que incluyen al síndrome de Peutz-Jeghers, la poliposis juvenil, y el síndrome hamartoma-tumor causado por mutaciones en *PTEN*. Dentro de los síndromes no polipósicos se encuentra el SL y el cáncer colorrectal familiar tipo X (CCRF-X), donde se agrupan los casos no polipósicos cuya causa genética se desconoce (revisado por Valle, 2014b).

Introducción

Tabla IV. Síndromes de predisposición al CCR y genes asociados.

Síndrome	Gen(es)	Herencia	Función de la proteína
Poliposis adenomatosa familiar (clásica y atenuada)	<i>APC</i>	AD	Inhibición de la vía Wnt
Síndrome de Lynch	<i>MLH1</i> <i>MSH2</i> <i>MSH6</i> <i>PMS2</i> <i>EPCAM*</i>	AD	Reparación de bases desapareadas
Poliposis asociada a <i>MUTYH</i>	<i>MUTYH</i>	AR	Reparación por escisión de bases
Poliposis asociada a la actividad reparadora de la polimerasa	<i>POLE</i> <i>POLD1</i>	AD	Corrección de errores introducidos por la polimerasa
Poliposis hereditaria mixta	<i>GREM1</i>	AD	Señalización BMP
Síndrome Peutz-Jeghers	<i>LKB1</i>	AD	Serín-treonín quinasa
Poliposis juvenil	<i>SMAD4</i> <i>BMPR1A</i>	AD	Señalización TGF-beta y BMP
Síndrome de tumor hamartoma <i>PTEN**</i>	<i>PTEN</i>	AD	Inhibición de la señalización AKT

*A consecuencia de deleciones de *EPCAM* se da una metilación germinal de las islas CpG de *MSH2*.

**El síndrome de tumor hamartoma *PTEN* incluye los síndromes de Cowden, Bannayan-Riley-Ruvalcaba, Proteus y Proteus-like, asociados a mutaciones en *PTEN*.

AD, herencia autosómica dominante; AR, herencia autosómica recesiva.

4. SÍNDROMES POLIPÓSICOS

4.1. Poliposis adenomatosa familiar

La poliposis adenomatosa familiar (Nº MIM 175100) es el segundo síndrome de CCR más frecuente, representando menos del 0,2-1% del total de casos de CCR diagnosticados (de la Chapelle, 2004). Es una enfermedad autosómica dominante caracterizada en su forma clásica por la presencia de cientos a miles de pólipos adenomatosos, apareciendo generalmente en la adolescencia. Si no se tratan estos pólipos pre-cancerosos algunos derivarán inevitablemente en un carcinoma colorrectal, con una edad media de diagnóstico de 39 años (Petersen et al., 1991). Por tanto, esta enfermedad tiene penetrancia prácticamente completa, por lo que se recomienda la colectomía profiláctica tras la aparición de los adenomas.

La variante atenuada de la enfermedad se diferencia de la clásica por el número de pólipos que se desarrollan en el paciente, entre 10 y 100. Además, éstos se localizan principalmente en el colon proximal y con una edad de diagnóstico más tardía. Se estima el riesgo acumulado a los 80 años en aproximadamente el 70% (Knudsen et al., 2010).

La poliposis adenomatosa familiar está causada por mutaciones en la línea germinal en el gen supresor de tumores *APC* (*Adenomatous Polyposis Coli*), importante en la regulación de la vía de Wnt. Alrededor del 80% de los casos clásicos tienen mutaciones en *APC*, pero menos del 30% de las formas atenuadas (Giardiello et al., 2001; Sieber et al., 2002). La mayoría de las mutaciones en este gen son cambios sin sentido o de cambio de la pauta de lectura que resultan en una proteína truncada. El 90% de las mutaciones son detectables mediante secuenciación del gen, mientras que el resto son grandes deleciones. En aproximadamente el 20% de los casos la mutación en *APC* parece ser *de novo*, por lo que no va acompañada de historia familiar (Bisgaard et al., 1994). En parte de estos individuos se ha observado que es el resultado de mosaicismo a nivel somático (Hes et al., 2008).

4.2. Poliposis asociada a *MUTYH*

Mutaciones bialélicas en el gen *MUTYH* son responsables de la poliposis asociada a *MUTYH* (Nº MIM 608456), siendo el único síndrome de predisposición al CCR conocido con herencia autosómica recesiva (Al Tassan et al., 2002). La mayoría de individuos con

esta anomalía desarrollan pólipos (<100), aunque también se han descrito pacientes con CCR sin o muy pocos pólipos. El riesgo de cáncer colorrectal en individuos sin vigilancia varía entre el 43% y casi el 100% (Sampson et al., 2003; Farrington et al., 2005). Existen además pacientes con manifestaciones extracolónicas, tales como pólipos o cáncer gástrico, cáncer de ovario, vejiga, mama, endometrio, tumores benignos y malignos de piel y de tiroides, anomalías dentales o hipertrofia congénita del epitelio pigmentario de la retina (Vogt et al., 2009). La gran variabilidad fenotípica observada en los portadores de mutaciones en *MUTYH* dificulta el diagnóstico basado en las características clínicas, y por ello se han definido varios rasgos, tanto clínicos como moleculares, para establecer una sospecha de poliposis asociada a *MUTYH* (Tabla V) (Morak et al., 2010; Nielsen et al., 2011).

Tabla V. Condiciones establecidas para el estudio de *MUTYH*. Al menos una de ellas debe cumplirse.

-
- Entre uno y diez adenomas colónicos antes de los 40 años.
 - Entre diez y cientos adenomas colónicos y/o pólipos hiperplásicos.
 - Poliposis en el colon (>100 pólipos) en ausencia de mutación germinal en *APC*.
 - CCR con la mutación somática *KRAS* c.34G>T.
 - Historia familiar de cáncer de colon (con o sin pólipos) con patrón de herencia autosómico recesivo.
-

MUTYH codifica para una glicosilasa de ADN implicada en la reparación por escisión de bases, importante en la prevención de mutagénesis causada por daño oxidativo (Xie et al., 2008). Las mutaciones patogénicas descritas para este gen son predominantemente de cambio de sentido, con presencia de mutaciones fundadoras y diferencias entre etnias. En poblaciones caucásicas se ha observado dos variantes comunes, c.536A>G (p.Tyr179Cys) y c.1187G>A (p.Gly396Asp) (anotado en base a la secuencia codificante más larga, NM 001128425.1), en más del 70% de los pacientes (Nielsen et al., 2009). A nivel somático, un rasgo característico de los adenomas y tumores asociados a *MUTYH* es la elevada frecuencia de transversiones G:C>T:A. Uno de estos cambios frecuentemente encontrado en estos tumores (64%) es la mutación c.34G>T en el codón 12 de *KRAS*, por lo que se ha sugerido su análisis como prueba de cribado para el consiguiente análisis genético de *MUTYH* (Lipton et al., 2003; van Puijenbroek et al., 2008; Nielsen et al., 2011).

4.3. Poliposis asociada a la actividad reparadora de la polimerasa

Recientemente se han identificado mutaciones en los genes que codifican las ADN polimerasas ϵ (*POLE*) y δ (*POLD1*) en casos de CCR familiar y poliposis. Estas mutaciones en la línea germinal muestran herencia autosómica dominante y predisponen a desarrollar múltiples adenomas y carcinomas colorrectales (Palles et al., 2013). Asimismo, mutaciones en *POLD1* también confieren un riesgo incrementado de cáncer de endometrio en mujeres portadoras (Briggs y Tomlinson, 2013).

Todas las mutaciones patogénicas encontradas hasta la fecha en estos dos genes se localizan en el dominio exonucleasa de la respectiva polimerasa, sugiriendo que deficiencias en la actividad reparadora durante la replicación del ADN son la causa promotora de la tumorigénesis en los individuos portadores (Palles et al., 2013; Church et al., 2013). Los tumores desarrollados en este síndrome, así como los tumores colorrectales y de endometrio esporádicos con mutaciones en *POLE*, se caracterizan por ser hipermutantes y por presentar estabilidad de microsátélites (EMS). Presentan alrededor de 500 cambios genéticos somáticos en las regiones codificantes, y un espectro mutacional con un incremento de transversiones G:C>T:A y A:T>C:G (Briggs y Tomlinson, 2013).

El fenotipo de la poliposis asociada a la actividad reparadora de la polimerasa se solapa en cierta medida con fenotipos asociados con mutaciones en línea germinal en *APC*, *MUTYH* y los genes MMR (Briggs y Tomlinson, 2013). Dado esta considerable variación, es necesaria la descripción de casos adicionales para describir de manera más precisa las características clínicas y genéticas de este síndrome. Recientemente, por ejemplo, se ha descrito que, como para *APC*, se pueden dar mutaciones *de novo* en *POLE* y también se ha observado la presencia de mutaciones en familias sin poliposis, rasgos que amplían los elementos a considerar para el análisis genético (Valle et al., 2014a).

4.4. Poliposis hereditaria mixta

El síndrome de poliposis hereditaria mixta (Nº MIM 601228) muestra una herencia autosómica dominante con penetrancia variable y se caracteriza por el desarrollo de distintos tipos de lesiones colorrectales, tales como pólipos Peutz-Jeghers, pólipos juveniles, serrados, adenomas convencionales y CCR (Whitelaw et al., 1997).

Tras estudios de ligamiento en familias Ashkenazi que acotaron la región candidata (Tomlinson et al, 1999; Jaeger et al., 2003), se logró identificar duplicaciones heterocigotas en el cromosoma 15q13-q14 que causan un incremento de la expresión del gen *GREM1* como causales de la poliposis y del riesgo incrementado de CCR en estas familias (Jaeger et al., 2012).

4.5. Poliposis hamartomasas

Los síndromes de poliposis hamartomasas son formas hereditarias poco frecuentes que predisponen a desarrollar múltiples pólipos hamartomasas intestinales, tumores del sistema gastrointestinal en el caso de la poliposis juvenil (Nº MIM 174900) y el síndrome Peutz-Jeghers (Nº MIM 175200), y tumores en diversas localizaciones en los síndromes causados por mutaciones en el gen *PTEN* (Nº MIM 601728).

La poliposis juvenil, la forma de poliposis hamartomasas más común, está causada por mutaciones en la línea germinal en los genes *SMAD4* (20% de los pacientes) y en *BMPR1A* (20% de los pacientes). Ambos genes están involucrados en la vía de señalización de TGF- β (Sayed et al., 2002; Howe et al., 2004). Por otro lado, el síndrome Peutz-Jeghers está causado por deficiencias en la serina-treonina quinasa *STK11* (también conocido como *LKB1*) y clínicamente se caracteriza por la presencia de lesiones cutáneas, pólipos gastrointestinales y un riesgo incrementado de cáncer (Hemminki et al, 1998; Jenne et al., 1998).

Finalmente, el síndrome de tumor-hamartoma causado por mutaciones en *PTEN* incluye a los síndromes de Cowden, Bannayan-Riley-Ruvalcaba, Proteus y Proteus-like. De momento no se ha dilucidado una clara relación genotipo-fenotipo de las diferentes mutaciones descritas (Tan et al., 2011), variando los fenotipos asociados desde macrocefalia y retraso en el desarrollo hasta el desarrollo de múltiples neoplasias primarias (Mester et al., 2013).

5. SÍNDROME DE LYNCH

El síndrome de Lynch (Nº MIM 120435) es el trastorno hereditario de predisposición al CCR más frecuente, que puede llegar a explicar hasta un 2-3% del total de casos de CCR diagnosticados. La susceptibilidad a este tipo de cáncer se hereda de manera autosómica dominante y es causada por defectos en los genes de reparación de bases desapareadas del ADN (Lynch et al., 1999). Dada la relativamente elevada penetrancia de la enfermedad, en estas familias se observan individuos afectados en las sucesivas generaciones y pocas veces se observa un salto generacional. Aunque muy poco frecuentes, también se han descrito casos con mutaciones en la línea germinal *de novo* en los genes MMR (Morak et al., 2008).

5.1. Manifestaciones y variantes clínicas

El SL presenta una gran heterogeneidad tanto genotípica como fenotípica, pero aún así presenta rasgos clínicos específicos que facilitan su identificación. Una de las principales características es una edad temprana de aparición de los tumores. La edad media de diagnóstico está en 45 años, siendo 69 años en la población general. Asimismo, también se observa una carcinogénesis acelerada, con secuencias adenoma-carcinoma que pueden tardar menos de 3 años, comparado con los 10 años de su homólogo esporádico. Existe una preferencia por que los tumores se desarrollen en el colon derecho o proximal. Asimismo, los pacientes afectados de SL tienen un riesgo elevado a padecer cánceres colorrectales adicionales, tanto sincrónicos como metacrónicos (Lynch et al., 2009).

Además de un incremento de riesgo a desarrollar CCR, los portadores de mutación en los genes MMR también tienen mayor riesgo a desarrollar otro tipo de neoplasias como el cáncer de endometrio, ovario, estómago, intestino delgado, conducto hepático biliar, páncreas, tracto uroepitelial superior y cerebro (Figura 10). Las mujeres portadoras de una mutación germinal para SL tienen un riesgo del 20-70% de padecer cáncer de endometrio, siendo éste el tumor extracolónico más frecuente (Dunlop et al., 1997).

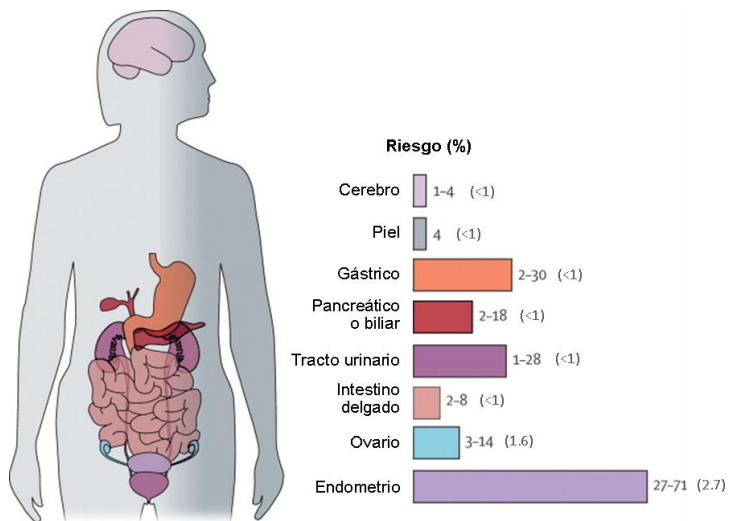


Figura 10. Riesgo de tumores extracolónicos en pacientes con síndrome de Lynch. Riesgo de la población general en paréntesis. Adaptado de Koornstra et al., 2009.

Existen también dos variantes clínicas del SL, el síndrome Muire-Torre y el síndrome de Turcot, fenotípicamente asociados con la presencia de tumores sebáceos de piel y del sistema nervioso central respectivamente (Schwartz y Torre, 1995; Hamilton et al., 1995). El síndrome de Turcot puede estar causado también por mutaciones en el gen *APC*. Estos individuos suelen presentar un mayor número de pólipos y un patrón distinto de tumores cerebrales que aquellos con síndrome de Turcot asociado al SL, es decir, con mutaciones en los genes *MMR* (Hamilton et al., 1995).

Aunque poco frecuentes, se han descrito casos con deficiencia constitucional de la reparación de bases desapareadas debido a mutaciones bialélicas en genes *MMR*. Esta anomalía provoca un fenotipo más agresivo con edades muy tempranas de aparición de los tumores y manifestando características típicas de la neurofibromatosis tipo I, como son la presencia de manchas café con leche, pecas axilares, nódulos de Lisch en los ojos o neurofibromas plexiformes (Gallinger et al., 2004).

5.2. Base genética: sistema de reparación de bases desapareadas del ADN

El síndrome de Lynch está causado por mutaciones en la línea germinal en los genes *MLH1*, *MSH2*, *MSH6* y *PMS2*, que codifican proteínas de identificación y reparación de errores de bases desapareadas en el ADN.

Los genes MMR son genes reparadores que actúan como genes supresores de tumores, de manera que, cumpliendo con el modelo de los dos eventos mutacionales de Knudson (Knudson, 1971), en individuos portadores de una mutación germinal hace falta una segunda mutación a nivel somático para que se desarrolle el cáncer. Este segundo evento puede darse por delección (pérdida de heterocigosidad), mutación o metilación de las islas CpG del promotor de *MLH1* (de la Chapelle; 2004). A consecuencia, con el fallo en la reparación de los desapareamientos del ADN resultantes de errores espontáneos que ocurren durante la replicación, se generan mutaciones deletéreas en otros genes supresores de tumores u oncogenes, promoviendo así el inicio de la carcinogénesis.

Mutaciones en *MLH1* y *MSH2* son responsables del SL en la gran mayoría de las familias con mutación identificada (más del 80%) (Figura 11). Mutaciones en *MSH6* son menos frecuentes (aproximadamente el 10%) y pueden causar una forma atenuada del SL, con menor penetrancia y una edad de desarrollo del tumor más tardía (Miyaki et al., 1997). Para *PMS2*, un gen con numerosos pseudogenes y genes parálogos que dificultan su estudio, las mutaciones encontradas son minoritarias y la penetrancia de la enfermedad también es menor que para los genes MMR principales (Senter et al., 2008). Aunque generalmente las variantes causales del SL son mutaciones nulas o de pérdida de función, se han descrito alelos hipomórficos que no muestran cambios en la secuencia, sino que resultan en una disminución de la expresión proteica (Renkonen et al., 2003). Asimismo, también hay casos de inactivación constitucional de *MLH1* mediante hipermetilación de sus islas CpG como alteración causal del SL (Ward et al., 2013a).

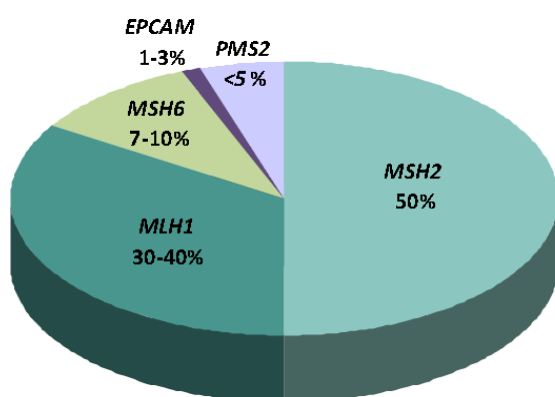


Figura 11. Frecuencia de mutaciones en los distintos genes causales del síndrome de Lynch. Estos porcentajes pueden variar según la región geográfica dependiendo de la presencia de mutaciones fundadoras o recurrentes mayoritarias.

Recientemente se ha identificado un nuevo mecanismo molecular involucrado en el SL: Deleciones germinales de los últimos exones del gen *EPCAM* provocan el silenciamiento epigenético de su gen vecino *MSH2*, causando consecuentemente el SL en los individuos portadores (Ligtenberg et al., 2009).

Entre el conjunto de mutaciones identificadas existen algunas que son recurrentes. Éstas pueden ser mutaciones que se producen repetidamente *de novo*, como la c.942+3A>T localizada en un sitio de *splicing* (mecanismo de corte y empalme) de *MSH2* (Desai et al., 2000), o mutaciones fundadoras presentes en algunas poblaciones (Nyström-Lahti et al., 1995; Clendenning et al., 2008; Pineda et al., 2010; Tomsic et al., 2013).

5.3. Identificación de familias con síndrome de Lynch

a) Cribado clínico

Para distinguir a nivel clínico las formas hereditarias de CCR de las esporádicas, se han utilizado la agregación familiar de neoplasias y una edad temprana de aparición del cáncer como fuertes indicadores del SL. Basándose en estas premisas, en 1990, antes de la identificación de los genes causales del trastorno, se describieron los denominados criterios de Ámsterdam (Vasen et al., 1991) (Tabla VIa). Ocho años después el mismo grupo internacional de investigadores revisó estos criterios para incluir los tumores extracolónicos, dando lugar a los criterios de Ámsterdam II (Vasen et al., 1999) (Tabla VIb).

En 1996, una vez descubiertos los genes responsables del síndrome, se definieron unos criterios clínicos menos estrictos, los criterios de Bethesda, con el objetivo de aumentar la sospecha de casos con SL y no perder individuos para los estudios genéticos (véase siguiente apartado, cribado molecular en tumor) (Rodríguez-Bigas et al., 1997). Estos criterios describen casi todas las condiciones clínicas en las que puede haber sospecha de un caso de SL. Una familia poco extensa, causas de muerte desconocidas o tempranas, penetrancia incompleta de mutaciones en la línea germinal y casos de no paternidad pueden ser causas de una historia familiar negativa. Posteriormente, estos criterios fueron revisados y simplificados (Umar et al., 2004) (Tabla VII).

Los criterios de Bethesda son los más sensibles, más del 90% de los afectados con SL cumplen una de sus condiciones. Pero por el contrario son los menos específicos (25%).

Son los criterios de Ámsterdam I los que diagnostican como portadores de mutaciones MMR una mayor proporción de individuos seleccionados (67% de especificidad) (Syngal et al., 2000).

Tabla VI. Criterios de Ámsterdam I y II. Deben cumplirse todos los criterios.

A) CRITERIOS DE ÁMSTERDAM I
<ul style="list-style-type: none"> - Tres familiares afectados de CCR, uno de ellos en primer grado de parentesco con los otros dos. - Al menos dos generaciones afectadas. - Al menos un CCR diagnosticado antes de los 50 años. - Poliposis excluida.
B) CRITERIOS DE ÁMSTERDAM II
<ul style="list-style-type: none"> - Tres familiares con cáncer asociado a CCHNP*, uno de ellos en primer grado de parentesco con los otros dos. - Al menos dos generaciones afectadas. - Al menos un CCR diagnosticado antes de los 50 años. - Poliposis excluida.

* colon, recto, endometrio, intestino delgado, uréter o pelvis renal.

Abreviaturas: CCHNP, CCR hereditario no polipósico.

Tabla VII. Criterios de Bethesda revisados. Debe cumplirse al menos uno de los criterios.

CRITERIOS DE BETHESDA
<ul style="list-style-type: none"> - CCR diagnosticado antes de los 50 años. - CCR u otro cáncer asociado a CCHNP* sincrónico o metacrónico, sin tener en cuenta la edad. - CCR con características patológicas asociadas a una alta IMS (presencia de linfocitos infiltrantes de tumor, carcinoma con diferenciación mucinosa o en anillo de sello, reacción linfocitaria peritumoral tipo Crohn-like, patrón de crecimiento medular) antes de los 60 años. - CCR con uno o más familiares en primer grado con cáncer colorrectal u otro cáncer relacionado con CCHNP*, uno de los cánceres diagnosticado antes de los 50 años. - CCR con dos o más familiares con CCR o cánceres asociado a CCHNP*, sin tener en cuenta la edad.

* Incluye endometrio, ovario, gástrico, intestino delgado, tracto urinario, tracto biliar, páncreas, cerebro y glándula sebácea.

Abreviaturas: CCHNP, CCR hereditario no polipósico.

b) Cribado molecular en tumor

Como ya se ha comentado en apartados anteriores, la deficiencia del sistema de reparación de bases desapareadas del ADN causa IMS, rasgo que emerge como muestra representativa de la inestabilidad genómica general existente. Este distintivo supone un muy buen marcador de la deficiencia de los genes MMR, y por ello el análisis de microsatélites suele ser la primera prueba molecular que se realiza en el laboratorio tras la sospecha de un caso con SL (Boland et al., 1998).

La inestabilidad de microsatélites es un rasgo característico pero no específico del SL. Tan solo un 20-25% de todos los tumores colorrectales con IMS está asociado a mutaciones en la línea germinal en un gen MMR; el 75-80% restante corresponde a casos esporádicos (Tabla VIII). Además, la IMS no se detecta en aproximadamente un 5% de todos los individuos SL portadores de mutación. Por ello la inmunohistoquímica (IHQ) para detectar la presencia o ausencia de los productos proteicos que codifican los genes MMR supone un valor añadido que complementa el test de IMS para el diagnóstico del SL. Esta técnica es sensible para alrededor del 95% de las deficiencias de genes MMR, pero requiere experiencia por parte del patólogo para su interpretación. (Lynch et al., 2009). Para la aplicación de esta prueba molecular hay que tener en cuenta que las proteínas MMR “principales”, MSH2 y MLH1, dimerizan con MSH6 y PMS2, respectivamente, logrando así un dímero proteico estable. Por ello, cuando en un tumor hay pérdida de expresión de MSH2 o MLH1, habrá pérdida de expresión concomitante de MSH6 y PMS2 respectivamente.

Además de estas dos pruebas esenciales para el diagnóstico del SL, se puede estudiar la presencia somática de metilación del promotor de *MLH1* y la mutación *BRAF* V600E para detectar aquellos tumores con IMS que presumiblemente son esporádicos (Parsons et al., 2012) (Tabla VIII).

Por tanto, una alta IMS y la pérdida de expresión de una o dos de las proteínas MMR en el tumor son los distintivos moleculares de esta patología, resultando indicadores positivos para efectuar el análisis mutacional de los genes implicados.

Sin embargo, dada la prevalencia de la enfermedad en la población general y los claros beneficios de las estrategias de vigilancia para la prevención del cáncer y la detección precoz en pacientes con SL, recientemente se ha propuesto realizar el cribado universal mediante IMS o IHQ en todos los casos de CCR diagnosticados (Teutsch et al., 2009).

Esto permitiría un incremento de la detección de los individuos con riesgo de cáncer elevado, resultando en una disminución de la mortalidad. Actualmente ya se han realizado varios estudios que apoyan la viabilidad de la estrategia (Hampel et al., 2008; Heald et al., 2013; Ward et al., 2013b).

Tabla VIII. Diferencias clínicas y moleculares entre los tumores IMS esporádicos y hereditarios (SL).

	Tumores IMS	
	Hereditarios	Esporádicos
% del total casos CCR	0,2-5%	7-18%
Edad media diagnóstico	45	69
Historia familiar	Positiva	Negativa
Tumores primarios	Múltiples sincrónicos	Individuales
Causa de la IMS	Mutación germinal gen MMR	Metilación somática <i>MLH1</i>
Característica molecular	IMS	Mutación <i>BRAF V600E</i>

5.4. Vigilancia médica y profilaxis

Existen varios estudios que han demostrado que las colonoscopias regulares con la extirpación de pólipos precancerosos reducen la incidencia de CCR y la mortalidad asociada en individuos con SL (Järvinen et al., 1995; de Vos tot Nederveen Cappel et al., 2002; Järvinen et al., 2009; Engel et al., 2010). Así, en aquellos individuos en los que se ha detectado una mutación en un gen de reparación de bases desapareadas se recomienda realizar colonoscopias anuales o bianuales, empezando entre los 20 y 25 años o bien 10 años antes del cáncer diagnosticado a menor edad en la familia. Para los portadores de mutación en los genes *MSH6* y *PMS2* el riesgo de CCR es menor, por lo que el inicio de vigilancia mediante colonoscopias se puede retrasar hasta los 30 años de edad (Senter et al., 2008; Baglietto et al., 2010).

Dado que la colonoscopia de rutina es una medida preventiva eficaz para el CCR, la colectomía profiláctica (resección del colon) no es una estrategia generalmente recomendada para individuos con SL. Por el contrario, la resección profiláctica en mujeres del útero y los ovarios sí que se considera más frecuentemente una vez sobrepasada la edad reproductiva.

5.5. Anticipación genética

La anticipación genética se define como el fenómeno de reducción de la edad de diagnóstico de una enfermedad y/o el aumento de la severidad de su fenotipo en sucesivas generaciones.

El síndrome de Lynch presenta cierta heterogeneidad fenotípica inter- e intrafamiliar, de manera que las características clínicas pueden variar considerablemente entre individuos de una misma familia (Vasen et al., 1994). En este sentido, aunque con cierta controversia, hay estudios que sugieren que se da un adelanto de la edad de aparición del cáncer al avanzar generaciones en familias con SL (Figura 12) (Tsai et al., 1997; Westphalen et al., 2005; Stella et al., 2007; Nilbert et al., 2009; Stupart et al., 2013). Además de las grandes diferencias clínicas que hay entre familias, existe una clara complicación analítica de los datos que ha hecho patente que el método estadístico utilizado sea determinante (Larsen et al., 2009; Boonstra et al., 2010). Los estudios más recientes que tienen en cuenta variables relevantes como son tendencias seculares, cribado molecular y la práctica médica, concluyen que sí existe anticipación en la edad de aparición del cáncer en sucesivas generaciones.

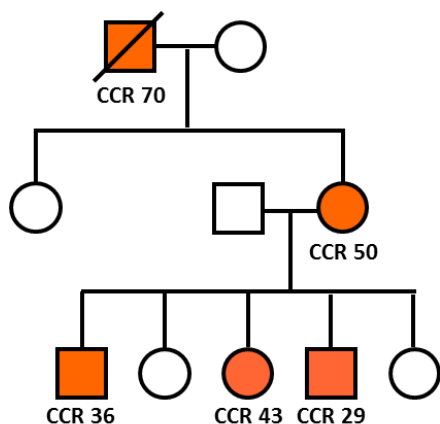


Figura 12. Anticipación genética en el síndrome de Lynch. Ejemplo de árbol genealógico. La edad de diagnóstico de los cánceres es menor en las sucesivas generaciones.

En algunas enfermedades donde se ha observado anticipación genética la base molecular se ha explicado por diversos mecanismos: Expansión de una repetición trinucleotídica (enfermedades neurodegenerativas como la enfermedad de Huntington), acortamiento telomérico (síndrome Li-Fraumeni y Dyskeratosis Congenita) o degradación proteosómica de p53 (síndrome Li-Fraumeni); sin embargo, para el SL la causa molecular subyacente permanece desconocida (Gruber y Mukherjee, 2009).

6. CÁNCER COLORRECTAL FAMILIAR TIPO X

Clásicamente el término CCR hereditario no polipósico se había utilizado como sinónimo de la deficiencia hereditaria del sistema de reparación del ADN MMR (SL). Sin embargo, éste es un término basado en las características clínicas de la genealogía y ya se ha discutido largamente que, dada la heterogeneidad de familias que abarca, puede llevar a confusión (Jass; 2006). Existen familias con agregación de cáncer colorrectal no polipósico que, aún cumpliendo los criterios clínicos de Ámsterdam, no presentan evidencias de IMS ni se detectan alteraciones en los genes MMR. Como se ha comentado en el apartado anterior, éstos son los distintivos principales del SL. Este conjunto de individuos, lejos de resultar una minoría, comprenden aproximadamente el 40% de todas las familias que cumplen los criterios de Ámsterdam (I). Descritas por primera vez en 1995 (Jass et al., 1995), en la última década numerosos grupos se han centrado en su caracterización, pero la etiología genética permanece aún desconocida. Actualmente, estas familias se agrupan bajo de denominación de CCRf-X (Lindor et al., 2005).

6.1. Características clínicas

Dado que las familias con CCRf-X cumplen los criterios de Ámsterdam, todas presentan un patrón de herencia autosómico dominante. Por otro lado, se han descrito ciertas características clínicas en estos pacientes que difieren de aquellos individuos afectados de SL. En primera instancia, destacar que las familias con CCRf-X presentan un riesgo menor de padecer CCR que las familias Lynch. Además, la edad media de diagnóstico es mayor para estos individuos, pero se sitúa alrededor de 10 años antes que en los casos esporádicos. Otra de las principales diferencias es la localización de los cánceres colorrectales, mientras en el SL existe una tendencia a situarse en la zona proximal del colon, en el grupo de casos CCRf-X la mayoría de tumores se desarrollan en el lado izquierdo o distal, principalmente en el colon sigmoide y el recto. Además, estos tumores no presentan las características de infiltración linfocítica que se han visto en los CCR de pacientes con SL y tienen un ritmo más lento de progresión adenoma-carcinoma. Cabe destacar también que estas familias tienen una menor incidencia de tumores extra-colónicos y de múltiples tumores primarios (Jass et al., 1995; Lindor et al., 2005; Llor et al., 2005; Mueller-Koch et al., 2005; Valle et al., 2007).

Por todo ello, los pacientes con CCRf-X no deben ser tratados ni asesorados bajo el mismo patrón que los pacientes con SL. La definición de esta nueva entidad y su diferenciación del SL ha sido importante para la implantación de un nuevo protocolo de vigilancia para estas familias. Se recomienda seguir protocolos de vigilancia menos estrictos con colonoscopias más espaciadas (cada 5 años) y con un inicio a edad más avanzada. Asimismo, se recomienda no hacer la evaluación endometrial, a menos que este tumor ya haya sido diagnosticado en algún miembro de la familia.

6.2. Características moleculares

Partiendo de la definición de CCRf-X, se agrupan en esta entidad todas aquellas familias Ámsterdam-positivas que no presentan deficiencias en el sistema de reparación de bases desapareadas del ADN. Por ello, al tener este sistema de reparación intacto, estos pacientes se caracterizan por tener tumores con EMS, al contrario que los individuos con mutaciones en la línea germinal en los genes MMR (tumores IMS). Hasta la fecha, éste es el rasgo distintivo molecular más característico que diferencia claramente los afectos de CCRf-X de aquellos individuos con SL, sugiriendo consecuentemente que son otras las vías de tumorigénesis implicadas (Dominguez-Valentin et al., 2013).

Varios estudios han analizado en tumores CCRf-X el patrón mutacional de genes típicamente alterados en la carcinogénesis colorrectal, tales como *KRAS*, *BRAF*, *APC* o *TP53*. Se han descrito similitudes entre estos tumores y los de casos esporádicos con EMS, aunque el patrón molecular en general se caracteriza por pocos cambios en las vías de carcinogénesis colorrectal conocidas (Abdel-Rahman et al., 2005; Sánchez-de-Abajo et al., 2007; Francisco et al., 2011). Analizando la presencia de inestabilidad cromosómica en estas neoplasias, parece que existen dos grupos bien diferenciados molecularmente en función de si presentan o no inestabilidad cromosómica.

El análisis de alteraciones del número de copias en el genoma de tumores CCRf-X ha detectado cambios que resultan también recurrentes en casos de CCR esporádico (ganancias en 7p, 7q, 8q, 13q, 20p y 20q; pérdidas de 17p, 18p y 18q), aunque con un incremento significativo en la frecuencia de la ganancia de 20q (Middeldorp et al., 2012; Therkildsen et al., 2013).

Por otro lado, en cuanto al nivel de alteraciones epigenéticas, se ha sugerido que en los casos de CCRf-X existe una hipometilación global comparado con los tumores del SL y los

esporádicos (tanto con EMS como IMS) que podría estar provocando una reactivación de proto-oncogenes que de otra manera estarían silenciados (Goel et al., 2010).

Por tanto, aun teniendo en cuenta los numerosos estudios realizados en la última década, poco se sabe a nivel molecular sobre estas familias, y aún se desconocen las causas genéticas que aportan la susceptibilidad aumentada a padecer CCR. Se cree que bajo la denominación de CCRf-X podrían estar agrupados distintos síndromes hereditarios de predisposición al CCR, formando así un grupo heterogéneo de familias. De hecho, probablemente engloba algunas que se explicarían con un modelo monogénico y otras que encajan con un componente poligénico, incluso algunas donde un estilo de vida compartido pudiera ser el responsable de la agregación familiar. Este hecho complica la decisión de qué técnica o estrategia es la adecuada para interrogar estas familias (Ku et al., 2012; Valle, 2014b). Por ejemplo, mediante la aproximación de gen candidato, Garre y colaboradores han propuesto los genes de la vía de reparación por escisión de bases, *NUDT1*, *OGG1* y *MUTYH*, como posibles responsables de la predisposición al CCR en algunas familias CCRf-X (Garre et al., 2011). Incluso se han asociado genes relacionados con síndromes polipósicos como es *BMPR1A* (Nieminen et al., 2011). Además, en los últimos años también se han relacionado con el CCR familiar mutaciones germinales en diversos genes como *UNC5C*, *GALNT12*, *BUB1* y *BUB3* (Gylfe et al., 2010; Coissieux et al., 2011; Clarke et al., 2012; de Voer et al., 2013), que podrían explicar la predisposición en algunas familias con CCRf-X.

7. BÚSQUEDA DE GENES CAUSALES DE ENFERMEDADES MENDELIANAS

Se denomina enfermedad mendeliana o monogénica aquella que está causada por mutaciones en un solo gen que sigue un patrón de herencia de acuerdo con las leyes de Mendel, ya sea autosómico dominante, autosómico recesivo o ligado al sexo. Existen numerosos métodos de búsqueda de genes causales para este tipo de trastornos (Tabla IX), siendo los más importantes los que se comentan a continuación.

7.1. Análisis de ligamiento

El análisis de ligamiento a nivel genómico es la técnica clásicamente más utilizada para identificar nuevos genes asociados a enfermedades monogénicas. El estudio de ligamiento consiste en el mapeo de *loci* genéticos en varios individuos de una familia o en varios individuos/familias con el mismo fenotipo clínico, mediante el análisis de numerosos marcadores, generalmente microsatélites o polimorfismos de una sola base (SNP, del inglés *Single Nucleotide Polymorphism*) situados a lo largo de todo el genoma. Con ello se intenta definir una región cromosómica de interés que cosegregue con la enfermedad (Figura 11).

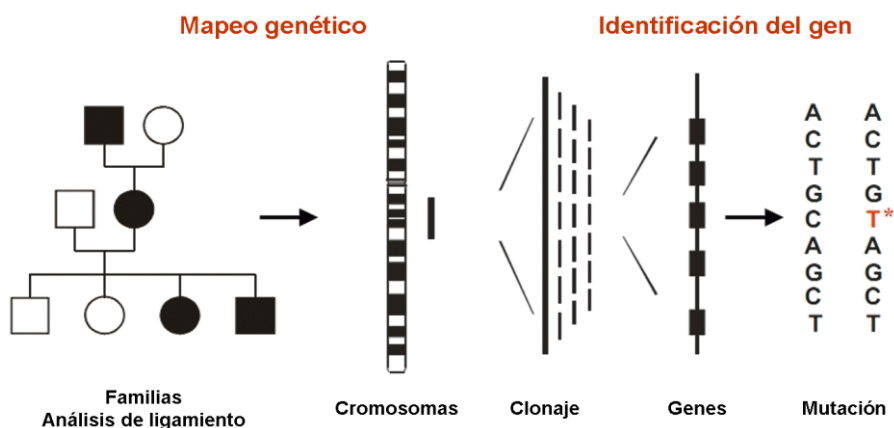


Figura 11. Esquema del análisis de ligamiento y clonaje posicional para la identificación de genes en enfermedades mendelianas. Tras delimitar la región presente en los individuos afectados y no en los no afectados, se ha de estudiar individualmente los genes situados en la región candidata.

La mayor parte de los genes de cáncer hereditario que hoy conocemos se han identificado mediante estudios de ligamiento en familias. Sin embargo, se ha llegado a un punto en que esta estrategia no es capaz de identificar nuevos genes causales. En las últimas

décadas, en el ámbito del CCR familiar se han reportado varios *loci* utilizando el análisis de ligamiento, incluyendo las regiones cromosómicas 3q21-q24, 7q31, 9q22.2-31.2, 11q23.2, 11q13.4, 14q24.2 y 22q12.1 (Kemp et al., 2006a; Neklason et al., 2008; Wiesner et al., 2003; Gray-McGuire et al., 2010; Djureinovic et al., 2006); pero no se han identificado las mutaciones causales responsables del riesgo incrementado en estas familias dentro de esas regiones (van Wezel et al., 2012).

7.2. Análisis directo de genes candidatos

Si se posee suficiente información acerca de los procesos bioquímicos y metabólicos involucrados en la enfermedad se pueden estudiar directamente diversos genes candidatos seleccionados según su función. Esta estrategia inicialmente no comporta grandes complicaciones, pero el número de genes que permite testar es reducido y suele obtener escasos resultados. Siguiendo esta aproximación, recientemente dos estudios han presentado evidencias de la implicación del gen *N-acetilgalactosaminiltransferasa 12* (*GALNT12*) en el riesgo a padecer cáncer colorrectal (Guda et al., 2009; Clarke et al., 2012). La enzima codificada por este gen pertenece a una familia de glicosiltransferasas importantes para el inicio de la O-glicosilación de la mucina (Ten Hagen et al., 2003). Las sospechas partían de la implicación de otro gen de la misma familia en el cáncer de mama (Wood et al., 2007), junto con el patrón de expresión de *GALNT12* en colon normal y tumoral (Guo et al., 2002; 2004) y la importancia del papel de la glicosilación en el cáncer (Brockhausen, 2006). Además, *GALNT12* está situado en 9q21-33, muy cerca de un pico de asociación (9q22-31) recurrentemente identificado en estudios de ligamiento de CCR familiar (Wiesner et al., 2003; Skoglund et al., 2006; Kemp et al., 2006b; Gray-McGuire et al., 2010).

Otros genes de predisposición al cáncer colorrectal familiar que se han identificado mediante la estrategia del análisis directo de genes candidatos son *UNC5C* (Coissieux et al., 2011), *NUDT1* y *OGG1* (Garre et al., 2011). La actividad pro-apoptótica de *UNC5C* junto con su asociación con la progresión de tumores intestinales en ratones, lo convertían en un buen gen candidato para ser testado en pacientes con CCR. Por otro lado, *NUDT1* y *OGG1* forman parte junto con *MUTYH* de la vía de reparación por escisión de bases. Esta vía ya es ampliamente conocida en la susceptibilidad al CCR por la poliposis asociada a mutaciones en *MUTYH* (véase apartado 4.2), mientras que no se había descrito nada para los otros dos genes.

7.3. Secuenciación de genomas y exomas completos

Más recientemente, la evolución en la tecnología de secuenciación masiva y los recientes avances en la captura de secuencias de alto rendimiento ha permitido que la secuenciación de exomas emerja como estrategia efectiva y asequible, tanto desde un punto de vista técnico como económico, para la identificación de variantes raras que confieren un riesgo elevado a desarrollar una enfermedad (Teer y Mullikin, 2010). En los últimos años la secuenciación de exomas ha propiciado un crecimiento exponencial en la identificación de nuevos genes causales de numerosas enfermedades mendelianas previamente sin resolver, tales como el síndrome Miller, el síndrome Kabuki o el síndrome Fowler entre otros (Ku et al., 2011) (Figura 12). En el área del cáncer familiar se han realizado algunos estudios que han permitido localizar genes de susceptibilidad con resultados positivos para formas familiares del cáncer de páncreas, feocromocitoma, cáncer de mama o pulmón (Jones et al., 2009a; Comino-Mendez et al., 2011; Thompson et al., 2012; Yamamoto et al., 2013).

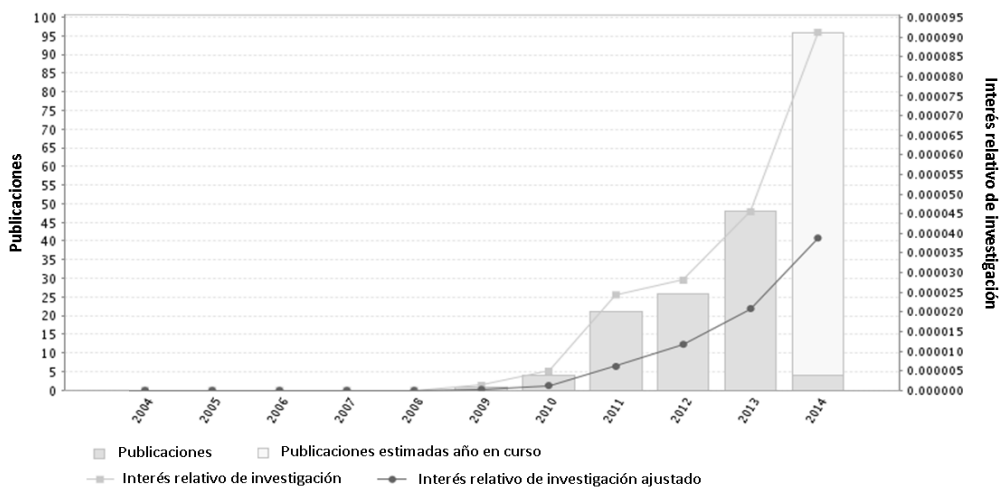


Figura 12. Evolución de las publicaciones e interés de investigación para la secuenciación de exomas en enfermedades mendelianas en la última década. La trama gris más oscura corresponde al interés relativo de investigación ajustado por el crecimiento general de PubMed.

Todos estos hallazgos han supuesto un avance considerable en el conocimiento de las enfermedades monogénicas, dadas las limitaciones de las técnicas previamente existentes. Aun así, los métodos tradicionales que reducen la búsqueda a un *locus* genómico específico se han unido en numerosos casos con la secuenciación masiva en

paralelo o de nueva generación (NGS, del inglés *next-generation sequencing*) para identificar el gen causal subyacente en enfermedades monogénicas. Es el caso del estudio realizado por Palles et al., en el que con la combinación de análisis de ligamiento y la secuenciación masiva de todo el genoma han conseguido identificar dos genes de predisposición al CCR y poliposis, *POLE* y *POLD1* (Palles et al., 2013); o el de los genes *BUB1* y *BUB3*, identificados mediante el análisis de número de copia junto con la secuenciación de exomas (de Voer et al, 2013).

A pesar de que la secuenciación masiva de exomas tan solo estudia el 1-2% del genoma humano, el enriquecimiento de las regiones codificantes permite que se obtenga una mayor profundidad de secuencia que cuando se realiza la secuenciación del genoma completo. Además, la secuenciación de exomas es más asequible a nivel de coste y su análisis bioinformático comporta una menor complejidad (Tabla IX).

Introducción

Tabla IX. Estrategias de identificación de genes para enfermedades Mendelianas. Adaptado de Gilissen et al., 2011.

Técnica	Ventajas	Desventajas
Gen candidato	Fácil de realizar para un número reducido de genes No requiere mapeo Identificación directa de la variante causal	Dependencia del conocimiento biológico existente Tasa de éxito muy baja
Mapeo genético por cariotipado	Fácil de realizar No requiere casos familiares Detecta eventos balanceados (grandes)	Baja resolución, solo detecta grandes anomalías cromosómicas Detección de la mutación requiere un segundo paso
Mapeo genético por análisis de ligamiento	Fácil de realizar	Requiere familias grandes Generalmente identifica <i>loci</i> amplios Detección de la mutación requiere un segundo paso
Mapeo genético por mapeo de homocigosidad	Permite utilizar familias pequeñas	Utilidad limitada a familias consanguíneas Generalmente identifica <i>loci</i> amplios Detección de la mutación requiere un segundo paso
Mapeo genético por análisis de variantes de número de copia	Cribado de alta resolución de variantes de número de copia No requiere casos familiares Potencialmente puede identificar <i>loci</i> pequeños	Solo estudia variantes de número de copia No puede detectar eventos balanceados Detección de la mutación requiere un segundo paso
Secuenciación de todo el exoma	Resolución nucleotídica a lo largo de todo el exoma Detecta la mayoría de variaciones génicas Identificación directa de la variante causal	Capacidad limitada de detección de variantes no codificantes Resolución limitada para las variantes de número de copia y otros cambios estructurales Variabilidad de la cobertura debido al proceso de enriquecimiento
Secuenciación de todo el genoma	Resolución nucleotídica a lo largo de todo el genoma Detecta la mayoría de variaciones génicas Identificación directa de la variante causal	Análisis de datos complejo Coste elevado Almacenamiento de datos

Abreviaturas: CNV, del inglés *Copy Number Variation*, cambio de número de copia.

8. TELÓMEROS

8.1. Los telómeros y la telomerasa

Los telómeros son secuencias repetitivas (5' TTAGGG 3') de ADN no codificante que están localizadas en los extremos de los cromosomas, protegiéndolos así de la degradación nucleotídica, de las fusiones entre extremos cromosómicos y de recombinaciones irregulares. Son esenciales para la estabilidad e integridad genómica y, por tanto, para la viabilidad celular (Blackburn, 1991). Las secuencias teloméricas sirven como plataforma para la unión de un complejo multiproteico, la shelterina o telosoma, que evita que el extremo final del ADN telomérico sea reconocido por la maquinaria de reparación del ADN (de Lange, 2005).

Los telómeros pierden entre 50 y 200 pares de bases con cada ciclo de división celular como consecuencia de la incapacidad de las ADN polimerasas para completar la replicación en los extremos del ADN. La inflamación y el estrés oxidativo pueden acelerar este acortamiento telomérico (von Zglinicki, 2002). La telomerasa es la enzima capaz de añadir repeticiones teloméricas *de novo* en el extremo 3' de los cromosomas, contrarrestando así las pérdidas. Esta polimerasa está formada por dos unidades principales: la subunidad catalítica TERT y el componente de ARN TERC, que se usa como plantilla para la síntesis del ADN telomérico (Figura 13). Sin embargo, la expresión de esta enzima está altamente regulada y la mayoría de las células somáticas no poseen actividad telomerasa, por lo que los telómeros se van acortando progresivamente con las sucesivas divisiones celulares. Cuando los telómeros alcanzan una longitud crítica las células entran en apoptosis, senescencia o adquieren anomalías en su estructura cromosómica (Londono-Vallejo, 2008). En este sentido, los telómeros actúan como mecanismo de control contra la proliferación ilimitada, y se les ha denominado "reloj mitótico" (Harley, 1991).

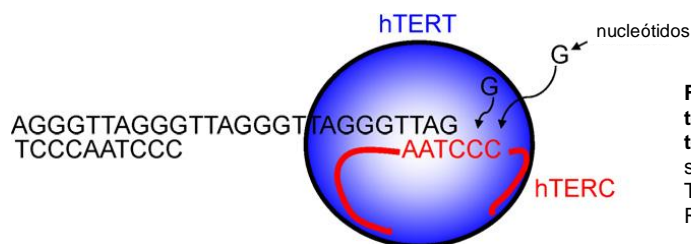


Figura 13. Actividad de la telomerasa en un extremo telomérico. Se diferencian las dos subunidades que forman la enzima: TERT y TERC. Modificado de Ryaunaud (2008).

8.2. Los telómeros y la telomerasa en cáncer

La biología de los telómeros juega un papel crítico y complejo en la iniciación y progresión del cáncer, situación en la que el equilibrio natural entre telómeros y telomerasa está alterado. Inicialmente, las anomalías en la longitud telomérica pueden actuar como inhibidores o promotores de la tumorigénesis dependiendo del ambiente celular. En células normales esta deficiencia conduce a la senescencia o apoptosis, pero en células en las que los puntos de control del ciclo celular están alterados, la presencia de telómeros cortos o anómalos promueve la inestabilidad cromosómica, de manera que facilitan que la célula acumule cambios genéticos potencialmente oncogénicos (Figura 14) (Xu et al., 2013). Cabe destacar que irregularidades en la longitud telomérica parecen ser una de las primeras y más prevalentes alteraciones genéticas en el proceso de la transformación tumoral (Meeker et al., 2004). Además, estas alteraciones en los telómeros se han asociado con la iniciación y la progresión de neoplasias intestinales e incluso se han propuesto como un factor de elevado potencial pronóstico en tumores sólidos (Bisoffi et al., 2006; Raynaud et al., 2008).

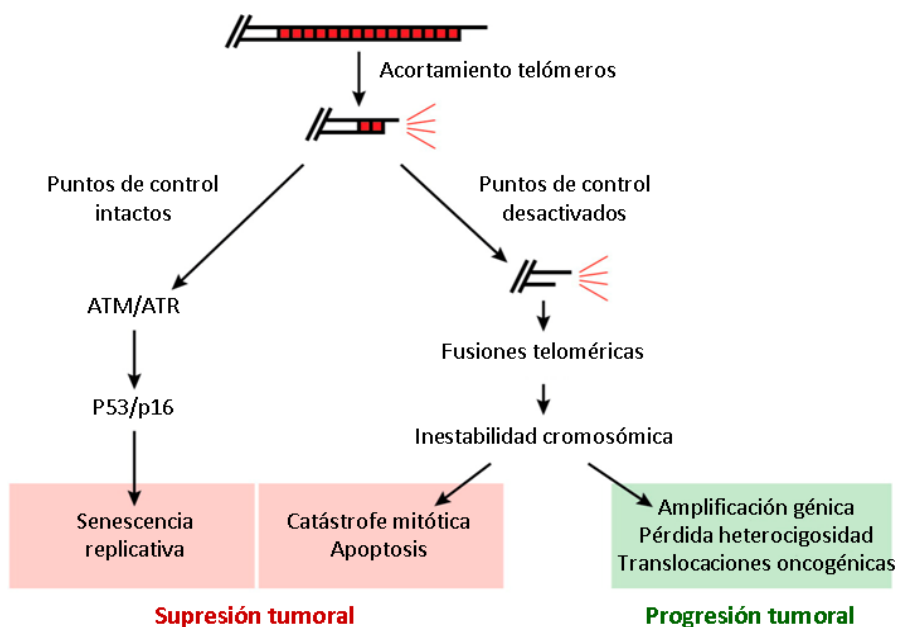


Figura 14. Papel del acortamiento telomérico en la supresión o formación de tumores. Adaptado de Xu et al., 2013.

Por otro lado, se ha observado que en más del 80% de tumores humanos existe una activación de la telomerasa que permite la estabilización de la longitud de los telómeros. Con ello las células cancerígenas son capaces de escapar de la apoptosis y conseguir la inmortalización celular, siendo éste uno de los rasgos característicos de las neoplasias (Hanahan y Weinberg, 2011). Además, en los tejidos tumorales puede activarse una vía alternativa de elongación de los telómeros (ALT, del inglés *alternative lengthening of telomeres*) independiente de la telomerasa. Este mecanismo está basado en la recombinación homóloga, utilizando el ADN como plantilla para preservar la longitud telomérica (Cesare y Reddel, 2010). En consonancia con estas observaciones, el modelo de ratón deficiente para la telomerasa evidencia la supresión de la progresión tumoral con telómeros cortos (Blasco, 2005).

8.3. Alteraciones en la longitud telomérica y la predisposición al cáncer

Existen numerosos ejemplos en la literatura que asocian anomalías en la longitud de los telómeros con la susceptibilidad a desarrollar cáncer. Mutaciones en el complejo de la telomerasa causan disqueratosis congénita, enfermedad en la que los telómeros cortos provocan una insuficiencia de la médula ósea y un riesgo incrementado de cáncer (Vulliamy et al., 2001). Además, se ha demostrado mediante modelos de ratones inducibles para la telomerasa, que alteraciones en la longitud de los telómeros con la posterior reactivación de la telomerasa predisponen a desarrollar cáncer (Artandi et al., 2000; Ding et al., 2012).

En los últimos años se han publicado numerosos estudios epidemiológicos que han evaluado la longitud telomérica en sangre periférica o en células bucales como marcador biológico del riesgo a padecer cáncer. Muchos estudios han descrito asociaciones de telómeros cortos (la mayoría) o largos con riesgo de cáncer, aunque los resultados son inconsistentes entre los estudios y los tipos de neoplasias (revisado por Prescott et al., 2011 y Hou et al., 2012). En el caso de CCR, se han obtenido resultados contradictorios en los diferentes estudios que se han realizado (Tabla X). Se ha relacionado tanto la presencia de telómeros cortos como largos con el riesgo de CCR, mientras que en algunos estudios no se ha encontrado asociación. Cabe destacar que existen diferencias considerables en las poblaciones estudiadas, diseño experimental y tamaño muestral (Zee et al., 2009; Pooley et al., 2010; Lee et al., 2010; Bozzao et al., 2011a; Maxwell et al., 2011; Jones et al., 2012; Cui et al., 2012; Pellatt et al., 2012; Qin et al., 2014). En cáncer familiar se han descrito las asociaciones más claras para el cáncer de mama, ovario y

Introducción

papilar de tiroides (Capezzone et al., 2008; Martínez-Delgado et al., 2012; Martínez-Delgado et al., 2013).

Tabla X. Estudios de longitud telomérica como factor de riesgo realizados hasta la fecha en pacientes de cáncer colorrectal.

Estudio	Diseño del estudio	Tipo de muestra	Tamaño muestral Casos/Controles	Asociación con incremento de riesgo a CCR
Zee et al., 2009	Prospectivo	Sangre periférica	357 / 134	Nula
Lee et al., 2010	Prospectivo	Sangre periférica	191 / 306	Nula
Pooley et al., 2010	Prospectivo y Retrospectivo	Sangre periférica	2249 / 2161	Nula (prospectivo) Telómeros cortos (retrospectivo)
Bozzao et al., 2011a	Retrospectivo	Sangre periférica (SL)	43 / 81	Nula
Maxwell et al., 2011	Retrospectivo	Sangre periférica	64 / 1348	Telómeros cortos
Jones et al., 2012	Retrospectivo	Sangre periférica	2157 / 3912	Telómeros largos
Cui et al., 2012	Prospectivo	Sangre periférica	441 / 549	Nula (linear) / Telómeros cortos y largos
Pellatt et al., 2012	Retrospectivo	Línea celular inmortalizada (colon) Sangre periférica (recto)	246 / 371 (colon) 276 / 372 (recto)	Colon: Tel. Cortos Recto: Nula (linear) / Telómeros cortos y largos
Qin et al., 2014	Retrospectivo	Sangre periférica	628 / 1256	Telómeros cortos

HIPÓTESIS

Aunque se estima la proporción de casos de CCR atribuibles a factores genéticos entre el 20 y 35%, gran parte de la herencia genética al CCR queda por descubrir. Dentro de ese conglomerado de casos de CCR con agregación familiar sin una causa subyacente clara, se encuentra el CCRf-X. Estas familias cumplen los criterios clínicos más estrictos para el CCR hereditario no polipósico (criterios de Ámsterdam) pero no presentan defectos en el sistema de reparación de bases desapareadas, rasgo definitorio de los pacientes con SL. Dado el patrón dominante de herencia que muestran estas familias, se espera que el gen causal sea de alta penetrancia. Sin embargo, la heterogeneidad presupuesta de esta agrupación hace que estrategias clásicas para la identificación de genes mediante la búsqueda de un gen común que explique una fracción relativamente elevada de casos, no obtengan resultados. Por ello hipotetizamos que el estudio individual de familias con CCRf-X mediante secuenciación de exomas, permitirá identificar su causa genética, si realmente se trata de una herencia autosómica dominante causada por una mutación en la línea germinal de un gen codificante de alta/moderada penetrancia.

Una manera más directa y específica de interrogar una enfermedad monogénica es mediante el estudio de un gen candidato. Se ha sugerido que mutaciones en el gen que codifica la glicosilasa GALNT12 confieren un riesgo incrementado a padecer CCR. Asumiendo estos indicios, hipotetizamos que este gen podría ser el responsable del riesgo incrementado en una fracción de las familias con CCRf-X.

Por otro lado, pese a que se conoce la causa genética del SL, existe una considerable heterogeneidad fenotípica inter- e intra-familiar que sugiere la participación de factores genéticos modificadores de la susceptibilidad y de las características clínicas de esta enfermedad. Teniendo en cuenta la importancia del papel de los telómeros en el proceso de iniciación y progresión tumoral en los cánceres humanos y su descripción como factor de riesgo a cáncer en numerosos tipos de tumores, hipotetizamos que la longitud telomérica puede ejercer un papel modulador del riesgo a desarrollar CCR en las formas de CCR hereditario no polipósico, tanto en SL como en aquellas clasificadas como CCRf-X. En el SL, además, podría explicar la anticipación de la edad de diagnóstico observada en las sucesivas generaciones de algunas familias.

OBJETIVOS

Objetivo general

Identificación de nuevos genes causales y mecanismos moduladores de riesgo a desarrollar cáncer en el CCR hereditario no polipósico.

Como aplicación a la práctica clínica, el objetivo final de este trabajo será disminuir el riesgo (prevenir) a desarrollar cáncer o facilitar su detección precoz para reducir la mortalidad por cáncer en individuos de alto/moderado riesgo. Esto será posible mediante la implementación en el diagnóstico genético de rutina del estudio de los genes identificados y de los factores modificadores de riesgo para conseguir la personalización de los protocolos de vigilancia en base al riesgo individual.

Objetivos específicos

- Estudiar el papel de la longitud de los telómeros como factor modificador de riesgo a cáncer en los distintos grupos de pacientes de CCR hereditario no polipósico, tanto SL como CCRf-X, y en controles sanos.
- Evaluar la participación de la longitud telomérica en la anticipación genética descrita en familias con SL.
- Evaluar el papel de las mutaciones en gen *GALNT12* como causantes de la agregación familiar de CCR en familias de alto riesgo (CCRf-X).
- Estudio de dos familias de alto riesgo sin causa genética conocida mediante secuenciación de exomas para la identificación de nuevos genes causales.
- Evaluación de los posibles genes candidatos resultantes de la secuenciación de exomas en su participación en la predisposición al CCR en la familia y como gen de susceptibilidad en general.

RESULTADOS

Los resultados del trabajo experimental de esta tesis están incluidos en varios artículos científicos, cuatro de ellos ya publicados y un quinto preparado para enviar a publicar. A continuación se incluye un breve resumen de cada uno de los trabajos seguido de la publicación.

Artículo 1

Longitud de los telómeros y anticipación genética en el síndrome de Lynch.

Artículo 2

Asociación de telómeros más largos con el riesgo de cáncer en el cáncer colorrectal hereditario no polipósico sin defectos en la maquinaria de reparación de bases desapareadas.

Artículo 3

GALNT12 no contribuye de manera relevante al cáncer colorrectal familiar tipo X.

Artículo 4

Identificación de mutaciones en *MUTYH* mediante secuenciación de exomas en una familia con un fenotipo atípico.

Artículo 5

Mutaciones en la línea germinal en *GEN1X* causan cáncer colorrectal hereditario a través de deficiencias en la reparación del ADN.

ARTÍCULO 1**Telomere length and genetic anticipation in Lynch syndrome**

Nuria Seguí, Marta Pineda, Elisabet Guinó, Ester Borràs, Matilde Navarro, Fernando Bellido, Víctor Moreno, Conxi Lázaro, Ignacio Blanco, Gabriel Capellá, Laura Valle.

PLoS One 23;8(4):e61286; 2013. doi: 10.1371/journal.pone.0061286.

Resumen:

Se han asociado alteraciones en la longitud de los telómeros con el riesgo incrementado a padecer varios tipos de tumores. Además, el acortamiento telomérico explica la anticipación genética de algunas enfermedades hereditarias. En este trabajo evaluamos la longitud de los telómeros en el SL tanto como factor modificador del riesgo a desarrollar cáncer, así como mecanismo asociado a la anticipación en la edad de aparición del cáncer en generaciones sucesivas. Estudiamos la longitud telomérica relativa en ADN extraído de sangre periférica de 244 individuos portadores de mutación en los genes MMR pertenecientes a 96 familias con SL y de 234 controles sanos. La cuantificación de la longitud telomérica relativa la realizamos mediante PCR cuantitativa monocroma multiplex. Observamos que los portadores de mutación en genes MMR que han desarrollado cáncer tienen los telómeros significativamente más cortos que sus familiares portadores no afectados. Además, el acortamiento telomérico progresivo que se da con la edad (fenómeno fisiológico normal) es más pronunciado en estos individuos. Asimismo, los portadores de mutación en genes MMR que no han desarrollado cáncer muestran los telómeros más largos que los controles no portadores sanos.

Por otra parte, mediante el estudio de 59 familias con SL, observamos que no existe una asociación entre el acortamiento telomérico y la anticipación genética. Sin embargo, a pesar del reducido tamaño muestral, observamos relación con la anticipación en todas las parejas madre-hijo, sugiriendo que el acortamiento telomérico podría estar relacionado con el fenómeno de la dependencia del riesgo de desarrollar cáncer según el origen paterno/materno de la mutación en el gen MMR.

Telomere Length and Genetic Anticipation in Lynch Syndrome

Nuria Seguí¹, Marta Pineda¹, Elisabet Guinó², Ester Borràs¹, Matilde Navarro¹, Fernando Bellido¹, Victor Moreno^{2,3}, Conxi Lázaro¹, Ignacio Blanco¹, Gabriel Capellá¹, Laura Valle^{1*}

1 Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain, **2** Unit of Biomarkers and Susceptibility, Catalan Institute of Oncology, IDIBELL and CIBERESP, Hospitalet de Llobregat, Barcelona, Spain, **3** Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Hospitalet de Llobregat, Barcelona, Spain

Abstract

Telomere length variation has been associated with increased risk of several types of tumors, and telomere shortening, with genetic anticipation in a number of genetic diseases including hereditary cancer syndromes. No conclusive studies have been performed for Lynch syndrome, a hereditary colorectal cancer syndrome caused by germline mutations in the DNA mismatch repair genes. Here we evaluate telomere length in Lynch syndrome, both as a cancer risk factor and as a mechanism associated with anticipation in the age of cancer onset observed in successive generations of Lynch syndrome families. Leukocyte telomere length was measured in 244 mismatch repair gene mutation carriers from 96 Lynch syndrome families and in 234 controls using a monochrome multiplex quantitative PCR method. Cancer-affected mutation carriers showed significantly shorter telomeres than cancer-free mutation carriers. In addition, cancer-affected carriers showed the most pronounced shortening of telomere length with age, compared with unaffected carriers. The anticipation in the age of cancer onset observed in successive generations was not associated with telomere shortening, although, interestingly, all mother-son pairs showed telomere shortening. In conclusion, cancer-affected mismatch repair gene mutation carriers have distinct telomere-length pattern and dynamics. However, anticipation in the age of onset is not explained by telomere shortening. Pending further study, our findings suggest that telomere attrition might explain the previously reported dependence of cancer risk on the parent-of-origin of mismatch repair gene mutations.

Citation: Seguí N, Pineda M, Guinó E, Borràs E, Navarro M, et al. (2013) Telomere Length and Genetic Anticipation in Lynch Syndrome. PLoS ONE 8(4): e61286. doi:10.1371/journal.pone.0061286

Editor: Paolo Peterlongo, IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Italy

Received: January 15, 2013; **Accepted:** March 8, 2013; **Published:** April 23, 2013

Copyright: © 2013 Seguí et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the Spanish Ministry of Economy (State Secretariat for Research, Development and Innovation) (BFU2009-10281), Carlos III Health Institute (FIS PI08/1635, FIS PI08/1359 and RETIC RD06/0020/1050 and RD06/0020/1051), the Scientific Foundation Asociación Española Contra el Cáncer, the Catalan Government (2009SGR290), and CIBERESP (CB07/02/2005). Nuria Seguí is supported by a fellowship from the Carlos III Health Institute and Laura Valle holds a contract "Ramón y Cajal". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lvalle@iconcologia.net

Introduction

Lynch syndrome (LS) (MIM 120435) is the most common hereditary syndrome that predisposes to colorectal cancer (CRC) and other extracolonic tumors, accounting for 2%–5% of the total burden of CRC. It is caused by germline mutations in the DNA mismatch repair (*MMR*) genes *MLH1* (MIM 120436), *MSH2* (MIM 609309), *MSH6* (MIM 600678) and *PMS2* (MIM 600259), as well as in *EPCAM* (MIM 185535). There is large phenotypic variation in the age of onset among and within families [1,2]. Progressively earlier age of cancer onset in successive generations has been reported [3–10]. However, the underlying molecular mechanisms of both the overall phenotypic variation and the anticipation in the age of onset are still unknown.

Telomere shortening has gained considerable interest as a potential biomarker of cancer risk and as a mechanism associated with genetic anticipation. Moreover, we have recently shown that a common genetic variant located in the telomerase gene (*hTERT*; MIM 187270) increases cancer risk in LS patients at early ages, and that this SNP is associated with shortened telomere length in cancer-affected *MMR* gene mutation carriers [11]. Telomeres consist of multiple short repeats (TTTAGG) at the ends of

chromosomes and protect them against large-scale genomic rearrangements. In most somatic cells, telomeres shorten with each replication cycle, eventually leading to cellular senescence or apoptosis. Interestingly, telomere length anomaly appears to be one of the earliest and most prevalent genetic alterations in the process of malignant transformation [12,13]. Given the proposed role of telomere shortening in early processes of carcinogenesis, attention has been directed to its potential role as a cancer risk biomarker. With this purpose, a number of retrospective and prospective observational studies have been conducted, although results are inconsistent among studies and tumor types [14–16]. With respect to CRC, despite the conflicting results obtained in retrospective studies, prospective studies have not found a convincing association between leukocyte telomere length and CRC risk [16–18].

Telomere length attrition has also been proposed as a mechanism of anticipation in different inherited diseases, being associated with early onset and/or severity of disease in genetic disorders such as dyskeratosis congenita [19–21], Li-Fraumeni [22,23], and familial breast and ovarian cancer [24,25].

Here we aim to elucidate the role of telomere length in LS-associated cancer risk and to study its implication in anticipation in the age of onset observed in successive generations of LS families.

Materials and Methods

Ethics statement

Written informed consent was obtained from all subjects. The study received the approval of the IDIBELL Ethics Committee (ref. PR221/09).

Study participants

A total of 388 individuals, 244 *MMR* gene mutation carriers and 144 non-carriers, from 96 LS families were included in the study. They were assessed through the Hereditary Cancer Program at the Catalan Institute of Oncology, ICO-IDIBELL, from 1998 to 2012. *MMR* gene mutation analysis was performed on genomic DNA extracted from peripheral blood. Large genomic alterations in the genes were studied using Multiplex Ligation-dependent Probe Amplification (MLPA) (SALSA MLPA Kits, MRC-Holland, The Netherlands). Mutation screening was carried out by direct sequencing after PCR amplification (primers and conditions available upon request). Ninety cancer-free individuals were included as controls. All individuals are of Caucasian origin. Informed consent was obtained from all subjects, although data were analyzed anonymously. Table 1 shows a summary of the genetic and clinical characteristics of the groups studied.

Cases and controls were recruited from the same homogeneous population, and storage and DNA extraction from peripheral blood were performed at the same facility and using the same extraction methods. We have observed that blood DNA samples from different sources (i.e., extracted at different laboratories, using different extraction methods, etc.) show large non-genetic variation in telomere length (data not shown). Thus, only samples (cases and controls) that underwent the exact same DNA extraction protocol (Flexigene DNA kit, Qiagen, Hilden, Germany) performed by the personnel of the Unit of Molecular Diagnostics of the Hereditary Cancer Program at ICO-IDIBELL,

and that were stored under the same conditions, were included in the study, as recommended by Prescott et al. for retrospective studies [16].

Telomere length assessment

Telomere length quantification was performed using the monochrome multiplex quantitative PCR method described by Cawthon et al. [26], which has been found to provide greater consistency than other methods used to measure telomere length [17]. The assays were performed using the Quantifast SybrGreen PCR Master Mix (Qiagen, Hamburg, Germany), beta-globin as single copy gene, and the LightCycler 480 real-time PCR detection system (Roche Diagnostics GmbH, Mannheim, Germany). All samples were assayed in triplicate. A standard curve with 7 concentrations spanning an 81-fold range (60 ng, 40 ng, 20 ng, 6.7 ng, 4 ng, 2.2 ng and 0.74 ng), also in triplicate, of an anonymous standard DNA (healthy and cancer-free 45 year-old individual) was included in every 384-well plate. Good replicates of the standard curves (for telomeres (T) and single copy gene (S)) and subsequent fitted linear regression lines were obtained among plates. Whenever possible, equal numbers of samples from different clinical groups were run in each 384-well plate. The relative telomere length (RTL) value for each sample was calculated according to the standard lines (for T and S) of the corresponding plate. Therefore, the value obtained was relative to the value of the anonymous standard DNA in the same plate (inter-run calibration). The average telomere length of each sample is expected to be proportional to its RTL.

Statistical analyses

Shortening of telomere length is observed with increasing age in cancer-free individuals [27]. In our control set (controls plus non-carriers, $n=234$), RTL was inversely correlated with age (Pearson's correlation $r=-0.187$; $p=0.004$). As previously described, RTL measurements were adjusted for age using the line of best fit for controls [25]. Thus, the difference between the observed and the predicted value was calculated for each sample and is represented as "age-adjusted RTL". There were no

Table 1. Genetic and clinical characteristics of the studied groups.

	Lynch syndrome families (n = 96)		Controls (n = 234)	
	<i>MMR</i> gene mutation carriers (244)		Non-carriers	Unrelated controls
	*Cancer	Cancer-free		
N	144	100	144	90
<i>MMR</i> gene mutated: n (%)	<i>MLH1</i> : 87 (60.4)	<i>MLH1</i> : 62 (62.0)	-	-
	<i>MSH2</i> : 42 (29.2)	<i>MSH2</i> : 29 (29.0)	-	-
	<i>MSH6</i> : 9 (6.3)	<i>MSH6</i> : 9 (9.0)	-	-
	<i>PMS2</i> : 5 (3.5)	<i>PMS2</i> : 0	-	-
	<i>EpCAM</i> : 1 (0.7)	<i>EpCAM</i> : 0	-	-
Median age at blood draw (\pm SD)	51.5 (\pm 13.4)	35.0 (\pm 11.5)	42.0 (\pm 14.9)	41.5 (\pm 15.1)
Sex: n (%)	M: 74 (51.4)	M: 40 (40.0)	M: 64 (44.4)	M: 28 (31.1)
	F: 70 (48.6)	F: 60 (60.0)	F: 80 (55.6)	F: 62 (68.9)
Median age at cancer diagnosis (\pm SD)	43.0 (\pm 12.9)			

N: number of subjects; SD: standard deviation; M: male; F: female.

*Cancer: Individual affected with a LS-associated cancer: CRC and/or cancer of the endometrium, ovary, stomach, small bowel, hepatobiliary tract, pancreas, upper uro-epithelial tract or brain.

doi:10.1371/journal.pone.0061286.t001

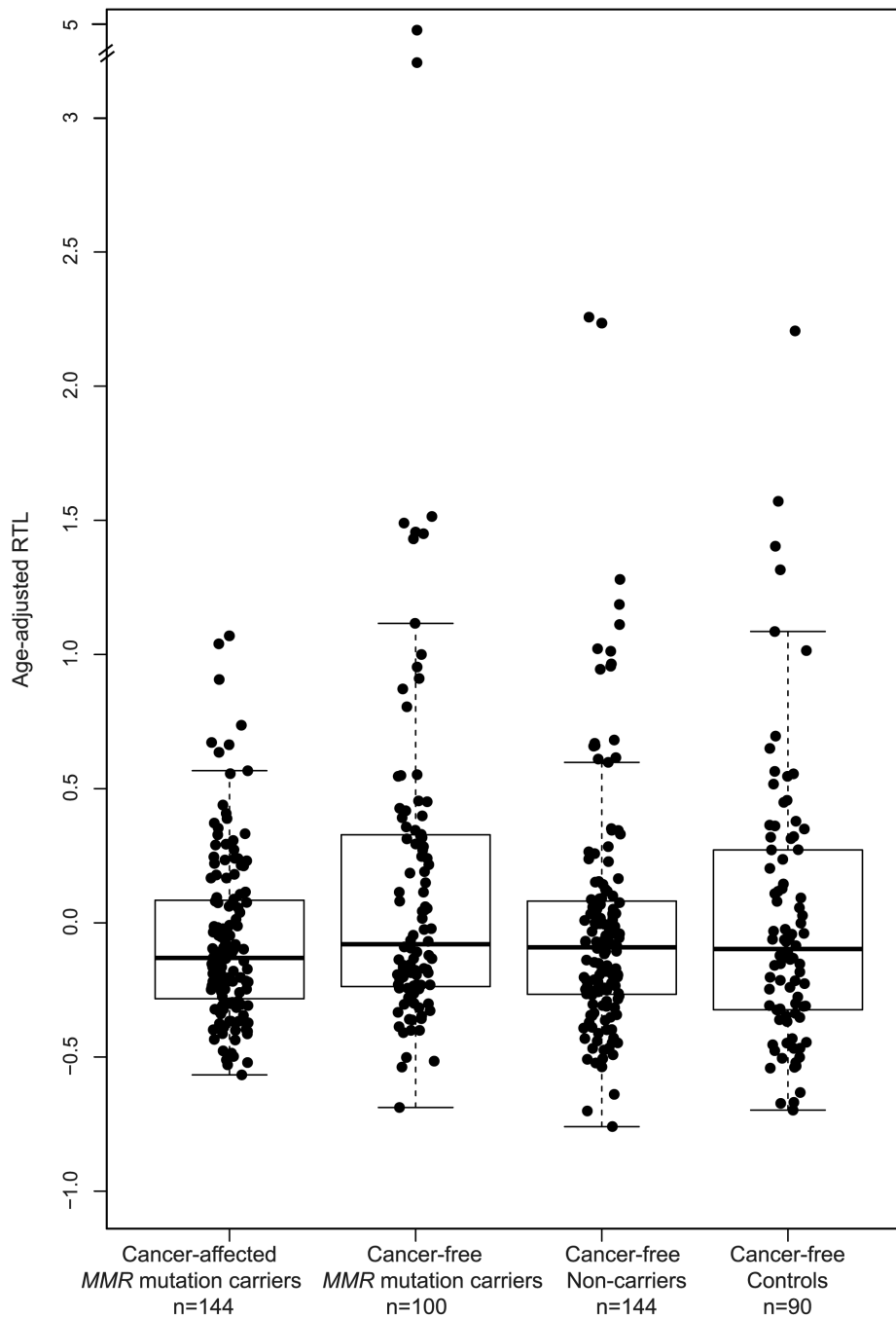


Figure 1. Age-adjusted relative telomere length. In subjects belonging to LS families: *MMR* gene mutation carriers affected with cancer (median: -0.131), *MMR* gene mutation carriers with no diagnosed cancer (median: -0.079) and cancer-free non-carriers (median: -0.091); and in cancer-free controls (median: -0.097). The boxes represent the interquartile range of distributions (25th and 75th percentiles); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles. doi:10.1371/journal.pone.0061286.g001

telomere length differences between sexes (data not shown); therefore, sex was not included as a confounding factor in the statistical tests.

Differences in age-adjusted telomere lengths were analyzed using the Wilcoxon rank sum test (Mann-Whitney U). Pearson's test was used to assess the correlation between telomere length and age (telomere length dynamics). To measure the significance of the difference between two correlation coefficients (Pearson's r),

Fisher's r -to- z transformation was performed. Anticipation was represented by Kaplan-Meier curves and the differences between survival curves were studied using a log-rank test. Proportions were compared by Fisher's exact test (expected cell count <5) or Chi-square. All tests were two-sided and p -values below 0.05 were considered statistically significant. The analyses were performed using R.

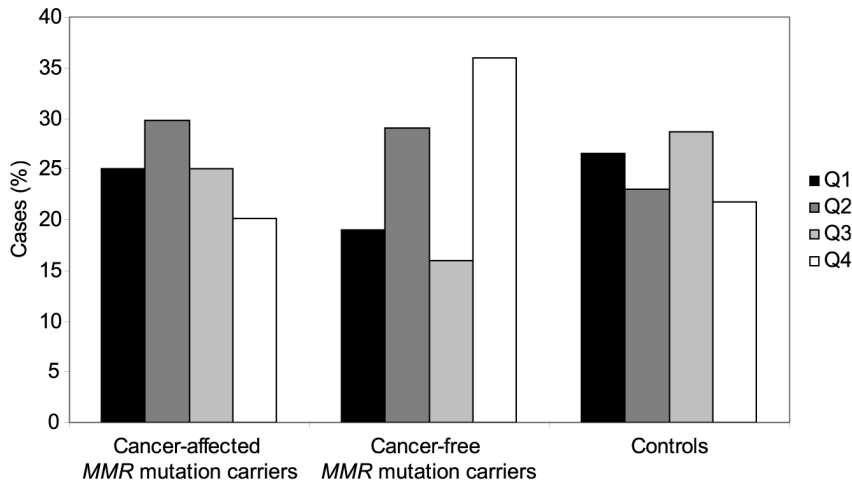


Figure 2. Distribution of cancer-affected and cancer-free *MMR* gene mutation cases and controls by quartiles of age-adjusted telomere length.

doi:10.1371/journal.pone.0061286.g002

Results

Telomere length and shortening dynamics

Relative telomere length was assessed in 244 *MMR* gene mutation carriers, both cancer-affected ($n = 144$) and unaffected ($n = 100$), and in 234 controls, including non-carriers from the same LS families ($n = 144$) and unrelated cancer-free controls ($n = 90$) (Figure 1). Relative telomere length values were adjusted for age as indicated in the Materials and Methods section (see statistical analyses). We found that among mutation carriers, cancer cases had significantly shorter age-adjusted telomeres than unaffected individuals ($p = 0.032$). It is worth noting that unaffected mutation carriers have longer telomeres than unaffected controls: a significant increase of the fourth quartile (longest telomeres) is observed in cancer-free *MMR* mutation carriers

(36.0%) compared with controls (21.8%; $p = 0.010$) and cancer-affected carriers (20.1%; $p = 0.009$) (Figure 2).

As expected, the negative slope of the linear regression line that best fit the RTL data for the controls indicated that telomere length shortened with age (Pearson's correlation coefficient $r = -0.187$; $p = 0.004$); the same was observed for the *MMR* gene mutation carriers ($r = -0.297$; $p = 2.37 \times 10^{-6}$). No differences were observed between both groups ($p = 0.204$) (Figure 3A). Among *MMR* gene mutation carriers, cancer-affected individuals showed faster telomere attrition with age ($r = -0.344$; $p = 2.4 \times 10^{-5}$) than cancer-free carriers ($r = -0.094$; $p = 0.351$) (for the difference, $p = 0.045$) (Figure 3B). No differences were observed between *MSH2* ($r = -0.446$; $p = 0.003$) and *MLH1* ($r = -0.318$; $p = 0.003$) mutation carriers (for the difference, $p = 0.435$).

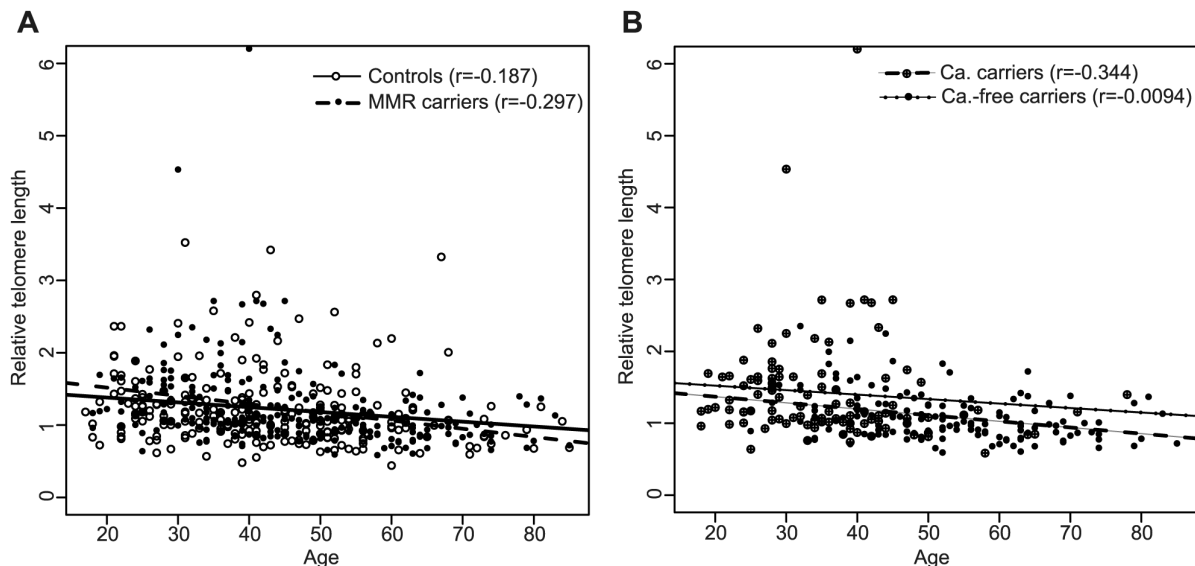


Figure 3. Correlation of relative telomere length (RTL) with age at blood draw. (A) RTL distribution as a function of age for controls ($n = 234$; white circles) and all *MMR* gene mutation carriers ($n = 244$; black circles). (B) RTL distribution as a function of age for cancer-affected *MMR* gene mutation carriers ($n = 144$; black circles) and cancer-free *MMR* gene mutation carriers ($n = 100$; crossed circles). r : Pearson's correlation coefficient. doi:10.1371/journal.pone.0061286.g003

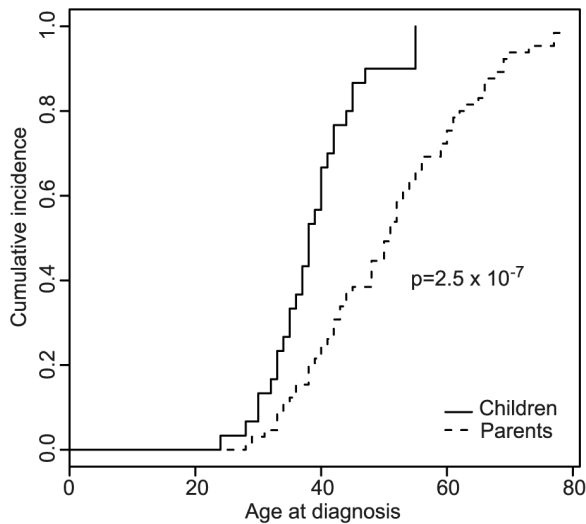


Figure 4. Anticipation in age of cancer onset in LS families. Kaplan-Meier curves and associated p-values showing the differences in age of cancer onset between parents and children, all of them cancer-affected *MMR* gene mutation carriers. doi:10.1371/journal.pone.0061286.g004

Telomere length and anticipation in the age of cancer onset

The occurrence of age anticipation in cancer-affected mutation carriers was analyzed in 59 LS families, harbouring *MLH1* (32 families), *MSH2* (20 families) and *MSH6* (7 families) mutations. The distribution of age at cancer diagnosis in parents and children showed a consistent shift to early ages in children ($p = 2.5 \times 10^{-7}$) (Figure 4). On average, cancer was diagnosed 12 years earlier in children (Table 2). Five out of 65 parents and 3 out of 30 children had been diagnosed with cancer (1st LS-related cancer diagnosed) during the follow-up period or clinical surveillance of mutation carriers. Their exclusion from the analysis did not alter the results (data not shown). The information available from polyp removal and surveillance time across generations was incomplete and was not included in the study.

The relationship between telomere length and anticipation in LS was next investigated. No differences between parents and children were observed in cancer-affected *MMR* gene mutation carriers ($p = 0.867$) (Figure 5). Similarly, no association was found when *MLH1* and *MSH2* gene mutation cases were analyzed

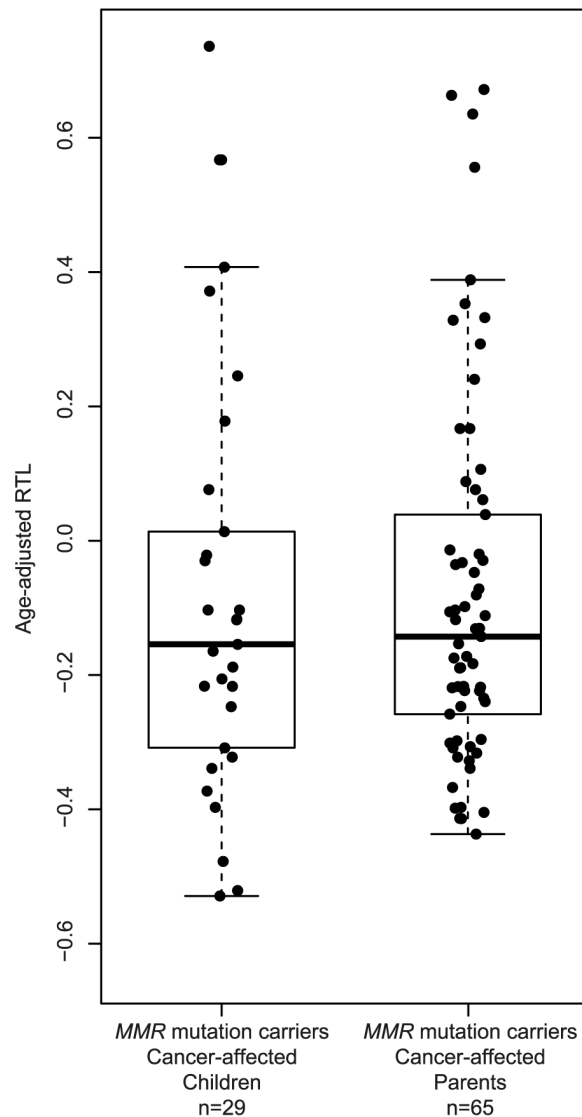


Figure 5. Changes in age-adjusted RTL between parents and children. Box plots representing age-adjusted RTL distributions in cancer-affected parents and children with *MMR* gene mutations. The boxes represent the interquartile range of distributions (25th and 75th percentiles); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles. doi:10.1371/journal.pone.0061286.g005

Table 2. Differences in average age of onset of cancer (1st LS-related tumor diagnosed) between parents and children with Lynch syndrome.

	Parents			Children			Mean age difference	t-statistic	p-value		
	N	Age of onset		N	Age of onset						
		Median	Mean		SD	Median				Mean	SD
All	65	51.0	50.9	12.8	30	38.0	38.9	7.6	12.0	4.74	7.5×10^{-6}
<i>MLH1</i>	36	50.5	50.9	11.7	20	38.0	38.0	6.8	12.9	4.53	3.3×10^{-5}
<i>MSH2</i>	21	53.0	51.4	15.4	7	37.0	41.4	10.4	10	1.59	0.124
<i>MSH6</i>	8	44.0	49.6	11.5	2	37.5	37.5	6.4	12.1	1.40	0.200

N: number of subjects; SD: standard deviation. doi:10.1371/journal.pone.0061286.t002

Table 3. Age-adjusted RTL and anticipation in age at cancer diagnosis in parent-child pairs.

Family	MMR gene	PARENT			CHILD			Age-adjusted RTL	**Tel _{child} <Tel _{parent}	
		Tumor	*Age	Sex	Age-adjusted RTL	Tumor	*Age			Sex
<i>Cancer-affected carriers</i>										
Fam-1	MSH2	CRC	73	F	-0.172	CRC	45	M	-0.216	yes
Fam-2	MSH2	EC	78	F	0.167	CRC	55	F	-0.339	yes
Fam-2	MSH2	CRC	55	F	-0.339	CRC	33	F	-0.188	no
Fam-2	MSH2	EC	78	F	0.167	CRC	55	F	-0.217	yes
Fam-3	MSH2	CRC	60	F	-0.098	CRC	35	F	0.407	no
Fam-4	MLH1	CRC	53	F	-0.143	CRC	40 (in surv)	F	0.372	no
Fam-4	MLH1	CRC	53	F	-0.143	CRC	32	M	-0.521	yes
Fam-5	MLH1	CRC	48	M	0.328	CRC	35	M	0.567	no
Fam-6	MLH1	CRC	52	F	0.635	CRC	34	F	0.736	no
Fam-7	MLH1	CRC	50	F	0.388	CRC	55 (in surv)	F	-0.103	yes
Fam-8	MLH1	CRC	51	M	-0.240	CRC	28	F	-0.022	no
Fam-9	MLH1	CRC	51	M	0.332	CRC	45	F	0.178	yes
Fam-10	MLH1	CRC	33	M	-0.398	CRC	24	F	-0.154	no
Fam-11	MLH1	CRC	69 (in surv)	M	-0.298	CRC	37	F	-0.206	no
Fam-11	MLH1	CRC	69 (in surv)	M	-0.298	CRC	41 (in surv)	F	-0.247	no
Fam-12	MLH1	CRC	66	F	-0.014	CRC	47	M	-0.373	yes
Fam-12	MLH1	CRC	66	F	-0.014	CRC	39	M	-0.322	yes
Fam-12	MLH1	CRC	66	F	-0.014	CRC	40	M	-0.103	yes
Fam-13	MLH1	CRC	60	F	-0.081	CRC	38	F	-0.030	no
Fam-14	MLH1	CRC	77	F	-0.235	CRC	38	F	-0.076	no
Fam-15	MSH6	UC	69	M	-0.032	CRC	33	M	-0.478	yes
<i>Cancer-free carriers</i>										
Fam-16	MLH1	-	48	M	-0.347	-	24	M	-0.327	no
Fam-16	MLH1	-	48	M	-0.347	-	25	F	-0.688	yes
Fam-17	MLH1	-	47	M	-0.121	-	27	F	-0.387	yes
Fam-18	MLH1	-	64	F	-0.244	-	36	M	-0.095	no
Fam-18	MLH1	-	64	F	-0.244	-	39	M	-0.134	no
Fam-19	MLH1	-	45	F	0.048	-	18	M	-0.202	yes
Fam-20	MLH1	-	78	M	0.399	-	37	F	-0.186	yes
Fam-21	MSH2	-	47	F	0.552	-	22	M	-0.021	yes
Fam-22	MSH6	-	65	F	-0.230	-	30	M	-0.300	yes
<i>***Cancer-free non-carriers</i>										
Fam-23	-	-	71	F	-0.336	-	45	M	0.000	no
Fam-20	-	-	74	F	-0.284	-	50	M	-0.181	no
Fam-24	-	-	85	F	-0.268	-	53	F	-0.432	yes

MMR: DNA mismatch repair; RTL: relative telomere length; CRC: colorectal cancer; EC: endometrial cancer; UC: urinary tract cancer; M: male; F: female; in surv: cancer diagnosed as a consequence of extensive clinical surveillance of LS families.

*Age: Age at cancer diagnosis for cancer-affected carriers, and age at blood draw for cancer-free carriers and non-carriers

**Tel_{child}<Tel_{parent}, age-adjusted telomere length is shorter in the child than in the parent. "Yes" indicates that age-adjusted RTL is smaller in the child than in the parent, therefore indicating association with anticipation when cancer-affected carriers are studied.

***Cancer-free non-carriers are individuals who belong to Lynch syndrome families but do not carry the pathogenic MMR gene mutation.

doi:10.1371/journal.pone.0061286.t003

separately (Figure S1). Next, we compared changes in telomere length in individual parent-child pairs (Table 3). In all instances, the child was diagnosed with cancer at an earlier age than the corresponding LS parent. Children showed shorter telomeres than their parents in 10 out of 21 (47.6%) cancer-affected carrier pairs, in 6 out of 9 (66.7%) cancer-free carrier pairs, and in 1 out of 3

(33.3%) cancer-free non-carriers evaluated, although these differences were not statistically significant ($p = 0.562$).

Anticipation was associated with shorter telomeres in all (5/5) mother-son cancer-affected MMR gene mutation carriers, whereas this association was only observed in 31% (5/16) of the other cancer-affected carrier pair combinations ($p = 0.012$) (Table 3).

Discussion

In this study we found that cancer-affected *MMR* gene mutation carriers have shorter telomeres and faster telomere attrition rates (accelerated telomere shortening with age) than unaffected mutation carriers and controls, and that telomere shortening is not associated with anticipation in the age of cancer onset in successive generations.

In line of recent retrospective and prospective studies on sporadic CRC risk and leukocyte telomere length, it has been suggested that the shorter telomere length observed in retrospectively collected samples from CRC patients is not a cancer risk factor, but rather a consequence of the disease (from treatment or disease burden) [16,17]. However, a recent prospective study has shown that both extremely long and extremely short telomeres increase CRC risk [18]. In hereditary cases, we observe shorter telomeres in cancer-affected *MMR* gene mutation carriers than in cancer-free carriers, suggesting either that the shortened telomeres are a result of the disease, or that shorter telomere length is an additional risk factor for LS patients, i.e. a risk modifier. This is supported by our recent observation that a variant in the telomerase gene, *hTERT*, increases cancer risk in young LS patients (<45 years of age), as it is associated with shortened telomeres in cancer-affected *MMR* gene mutation carriers [11]. Further studies assessing telomere length before and after cancer diagnosis in CRC patients will provide a definitive answer regarding the suggested effect of cancer on telomere length in blood cells. Unfortunately, samples from sporadic CRC cases that underwent the same DNA extraction and storage than the hereditary cases and controls herein studied were not available, precluding a direct comparison. This study will be critical to clarify whether telomere shortening is specific to cancer-affected LS patients or if it occurs in all CRC-affected patients.

Our data show longer telomeres in cancer-free *MMR* gene mutation carriers compared to cancer-free controls. A plausible explanation for this observation, and consistent with the fact that short telomeres act as factors of increased cancer risk, is that longer telomeres protect *MMR* gene mutation carriers from developing cancer. On the other hand, Jones et al. proposed that certain genetic factors that increase CRC risk may cause longer telomeres [17]. Alternatively, and in line with this model, our results might suggest that *MMR* haploinsufficiency, as a genetic factor of CRC risk, can exert a lengthening effect on the telomeres. Additional studies to assess the effect of *MMR* haploinsufficiency on telomere lengthening are needed to validate this hypothesis.

We also observed that leukocyte telomere dynamics differ between cancer-free mutation carriers and cancer-affected carriers, the latter group showing faster telomere shortening with age. Bozzao et al. found that affected carriers showed accelerated attrition rate compared to controls [28]. Based on previously published data on *MMR*-deficiency and telomeres [29–33], they suggested that *MMR* haploinsufficiency in LS individuals may lead to tolerance of short telomeres, causing faster telomere shortening over the course of life. However, our results show that this increased shortening rate does not occur in cancer-free carriers, thus invalidating the hypothesis. Therefore, this effect may be either the result of telomere shortening in cancer-affected individuals, or a marker of increased cancer risk among *MMR* gene mutation carriers, supporting the idea that short telomere length or accelerated telomere attrition rate are factors of increased cancer risk in LS. Identifying the dynamics of leukocyte telomere length with age in sporadic cases from retrospective and prospective studies may resolve this question.

No association was found between the anticipation in the age of onset observed in successive generations and telomere length. For years, and despite the numerous reports identifying anticipation in LS, it was uncertain whether true genetic anticipation contributed to the early diagnosis age observed in LS. Recently, a Bayesian method that corrects for random effects, isolating the confounding effect of changes in secular trends, screening, and medical practices, and adjusting for changes in age-specific incidence across birth cohorts, confirmed the anticipation in the age of onset between successive generations of LS families [10]. Nevertheless, the molecular mechanism underlying this observation has not been yet identified. Due to the lack of information on surveillance time in parents and children, polyp removal, and other confounding factors, as well as the insufficient sample size for an accurate assessment of genetic anticipation, it is certain that the measurement obtained in our series (12 years) is clearly overestimated. Nevertheless, we accept the results of previous reports on this matter in which accurate statistical methodologies were used and anticipation was still confirmed [8–10], and our series was used to assess the involvement of telomere attrition in anticipation. Our data rule out the role of telomere shortening in anticipation in LS. As previously suggested, other molecular mechanisms such as the accumulation of mismatch repair slippage events through generations or other genetic or environmental factors might explain age anticipation in successive generations of LS families [34,35]. Considering the limitations of our study regarding sample size, and the inconclusive results obtained when studying telomere length in cancer patients, further studies on larger LS series should be performed to validate these results.

Van Vliet et al. recently suggested that CRC risk in carriers of *MMR* gene mutations depends on the parent-of-origin of the mutation. In particular, they suggested a maternally transmitted mechanism modifying cancer risk in male *MMR* gene mutation carriers [36]. Here we observe anticipation associated with shorter telomeres in all mother-son cancer-affected *MMR* gene mutation carriers, whereas the same association is only found in 31% of the other cancer-affected carrier pair combinations. Although the sample size of parent-child pairs is very limited, this observation might suggest that the maternally transmitted risk mechanism is shortened telomeres. Studies assessing whether telomere shortening occurs in the female germline, as recently demonstrated in other mammals [37], together with the analysis of larger samples of LS parent-child pairs, are required to verify this hypothesis.

In conclusion, our findings indicate that cancer-affected *MMR* gene mutation carriers show distinct features and dynamics of telomere length measured in blood DNA, to controls and unaffected mutation carriers. While we rule out telomere length attrition as the common cause of anticipation in LS, it may account for the dependence of cancer risk on the parent-of-origin of *MMR* gene mutations previously observed.

Supporting Information

Figure S1 Age-adjusted RTL distributions in cancer-affected parents and children with *MLH1* and *MSH2* gene mutations. The boxes represent the interquartile range of distributions (25th and 75th percentiles); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles. Pairwise comparisons using Wilcoxon rank sum test showed no differences in RTL distributions between cancer-affected parents and children with germline mutations in *MLH1* ($p = 0.65$), and between parents and children with mutations in *MSH2* ($p = 0.67$). (TIF)

Acknowledgments

We thank all those responsible for genetic counselling and genetic testing as part of the ICO-IDIBELL Hereditary Cancer Program.

Author Contributions

Conceived and designed the experiments: LV NS GC. Performed the experiments: NS LV FB. Analyzed the data: LV EG VM NS. Contributed reagents/materials/analysis tools: MP EG EB MN CL IB VM. Wrote the paper: LV. Revised the article critically for important intellectual content: GC IB MP VM MN NS EB EG CL FB. Final approval of the article: GC IB MP VM MN NS EB EG CL FB.

References

- Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, et al. (1993) Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. *Gastroenterology* 104: 1535–1549.
- Scott RJ, McPhillips M, Meldrum CJ, Fitzgerald PE, Adams K, et al. (2001) Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. *Am J Hum Genet* 68: 118–127.
- Vasen HF, Taal BG, Griffioen G, Nagengast FM, Cats A, et al. (1994) Clinical heterogeneity of familial colorectal cancer and its influence on screening protocols. *Gut* 35: 1262–1266.
- Rodriguez-Bigas MA, Lee PH, O'Malley L, Weber TK, Suh O, et al. (1996) Establishment of a hereditary nonpolyposis colorectal cancer registry. *Dis Colon Rectum* 39: 649–653.
- Westphalen AA, Russell AM, Buser M, Berthod CR, Hutter P, et al. (2005) Evidence for genetic anticipation in hereditary non-polyposis colorectal cancer. *Hum Genet* 116: 461–465.
- Stella A, Surdo NC, Lastella P, Barana D, Oliani C, et al. (2007) Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. *Clin Genet* 71: 130–139.
- Nilbert M, Timshel S, Bernstein I, Larsen K (2009) Role for genetic anticipation in Lynch syndrome. *J Clin Oncol* 27: 360–364.
- Larsen K, Petersen J, Bernstein I, Nilbert M (2009) A parametric model for analyzing anticipation in genetically predisposed families. *Stat Appl Genet Mol Biol* 8: Article26.
- Boonstra PS, Gruber SB, Raymond VM, Huang SC, Timshel S, et al. (2010) A review of statistical methods for testing genetic anticipation: looking for an answer in Lynch syndrome. *Genet Epidemiol* 34: 756–768.
- Boonstra PS, Mukherjee B, Taylor JM, Nilbert M, Moreno V, et al. (2011) Bayesian modeling for genetic anticipation in presence of mutational heterogeneity: a case study in Lynch syndrome. *Biometrics* 67: 1627–1637.
- Bellido F, Guino E, Jagmohan-Changur S, Segui N, Pineda M, et al. (2012) Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome. *Eur J Hum Genet*.
- Londono-Vallejo JA (2004) Telomere length heterogeneity and chromosome instability. *Cancer Lett* 212: 135–144.
- Meeker AK, Hicks JL, Iacobuzio-Donahue CA, Montgomery EA, Westra WH, et al. (2004) Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. *Clin Cancer Res* 10: 3317–3326.
- Ma H, Zhou Z, Wei S, Liu Z, Pooley KA, et al. (2011) Shortened telomere length is associated with increased risk of cancer: a meta-analysis. *PLoS One* 6: e20466.
- Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA (2011) The association of telomere length and cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 20: 1238–1250.
- Prescott J, Wentzensen IM, Savage SA, De Vivo I (2012) Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutat Res* 730: 75–84.
- Jones AM, Beggs AD, Carvajal-Carmona L, Farrington S, Tenesa A, et al. (2012) TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. *Gut* 61: 248–254.
- Cui Y, Cai Q, Qu S, Chow WH, Wen W, et al. (2012) Association of Leukocyte Telomere Length with Colorectal Cancer Risk: Nested Case-Control Findings from the Shanghai Women's Health Study. *Cancer Epidemiol Biomarkers Prev* 21: 1807–1813.
- Vulliamy T, Marrone A, Szydlo R, Walne A, Mason PJ, et al. (2004) Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. *Nat Genet* 36: 447–449.
- Marrone A, Walne A, Dokal I (2005) Dyskeratosis congenita: telomerase, telomeres and anticipation. *Curr Opin Genet Dev* 15: 249–257.
- Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, et al. (2005) Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc Natl Acad Sci U S A* 102: 15960–15964.
- Tabori U, Nanda S, Druker H, Lees J, Malkin D (2007) Younger age of cancer initiation is associated with shorter telomere length in Li-Fraumeni syndrome. *Cancer Res* 67: 1415–1418.
- Trkova M, Prochazkova K, Krutilkova V, Sumerauer D, Sedlacek Z (2007) Telomere length in peripheral blood cells of germline TP53 mutation carriers is shorter than that of normal individuals of corresponding age. *Cancer* 110: 694–702.
- Gramatges MM, Telli ML, Balise R, Ford JM (2010) Longer relative telomere length in blood from women with sporadic and familial breast cancer compared with healthy controls. *Cancer Epidemiol Biomarkers Prev* 19: 605–613.
- Martinez-Delgado B, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M, et al. (2011) Genetic anticipation is associated with telomere shortening in hereditary breast cancer. *PLoS Genet* 7: e1002182.
- Cawthon RM (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 37: e21.
- Iwama H, Ohyashiki K, Ohyashiki JH, Hayashi S, Yahata N, et al. (1998) Telomeric length and telomerase activity vary with age in peripheral blood cells obtained from normal individuals. *Hum Genet* 102: 397–402.
- Bozzao C, Lastella P, Ponz de Leon M, Pedroni M, Di Gregorio C, et al. (2011) Analysis of telomere dynamics in peripheral blood cells from patients with Lynch syndrome. *Cancer* 117: 4325–4335.
- Pickett HA, Baird DM, Hoff-Olsen P, Meling GI, Rognum TO, et al. (2004) Telomere instability detected in sporadic colon cancers, some showing mutations in a mismatch repair gene. *Oncogene* 23: 3434–3443.
- Mendez-Bermudez A, Hills M, Pickett HA, Phan AT, Mergny JL, et al. (2009) Human telomeres that contain (CTAGGG)_n repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res* 37: 6225–6238.
- Martinez P, Siegl-Cachedenier I, Flores JM, Blasco MA (2009) MSH2 deficiency abolishes the anticancer and pro-aging activity of short telomeres. *Aging Cell* 8: 2–17.
- Rampazzo E, Bertorelle R, Serra L, Terrin L, Candiotti C, et al. (2010) Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers. *Br J Cancer* 102: 1300–1305.
- Mendez-Bermudez A, Royle NJ (2011) Deficiency in DNA mismatch repair increases the rate of telomere shortening in normal human cells. *Hum Mutat* 32: 939–946.
- Gruber SB, Mukherjee B (2009) Anticipation in lynch syndrome: still waiting for the answer. *J Clin Oncol* 27: 326–327.
- Coolbaugh-Murphy MI, Xu JP, Ramagli LS, Ramagli BC, Brown BW, et al. (2010) Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. *Hum Mutat* 31: 317–324.
- Van Vliet CM, Dowty JG, van Vliet JL, Smith L, Mead IJ, et al. (2011) Dependence of colorectal cancer risk on the parent-of-origin of mutations in DNA mismatch repair genes. *Hum Mutat* 32: 207–212.
- Bender HS, Murchison EP, Pickett HA, Deakin JE, Strong MA, et al. (2012) Extreme telomere length dimorphism in the tasmanian devil and related marsupials suggests parental control of telomere length. *PLoS One* 7: e46195.

SUPPORTING INFORMATION

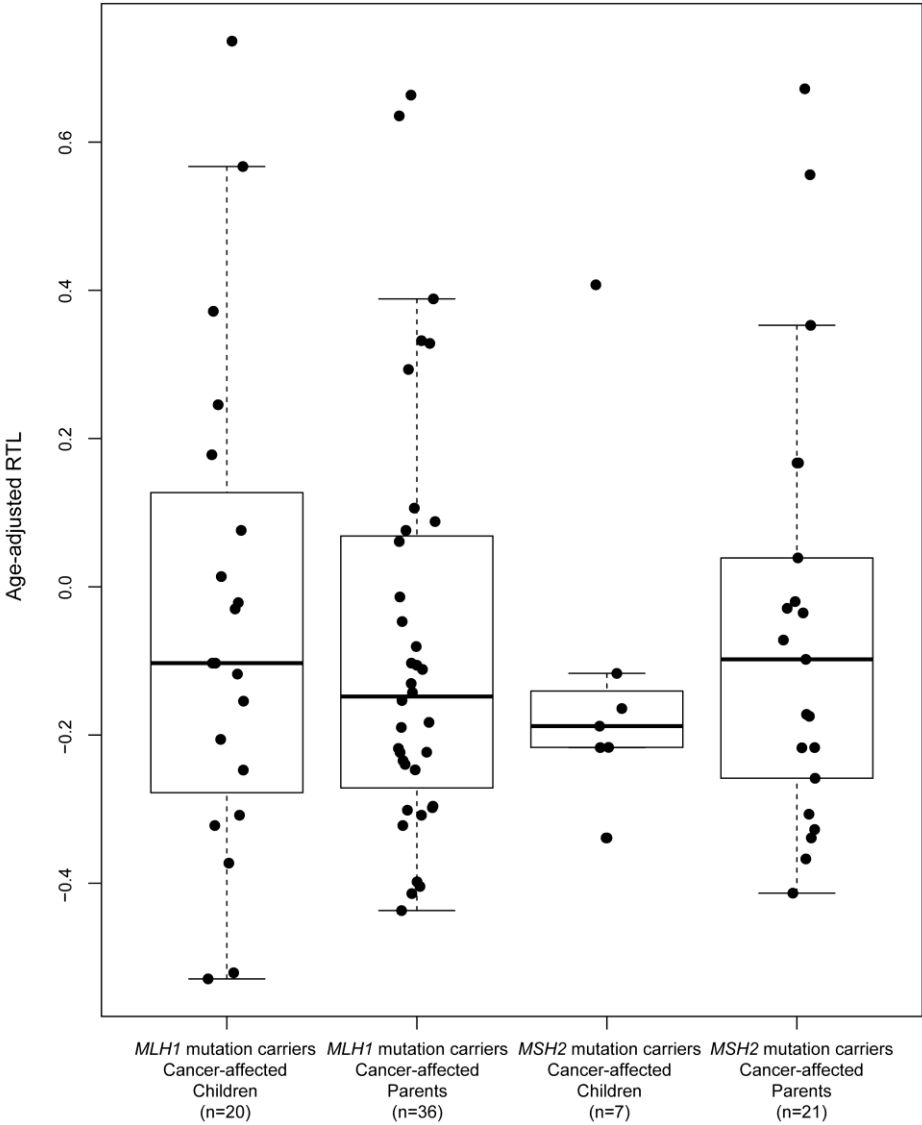


Figure S1. Age-adjusted RTL distributions in cancer-affected parents and children with *MLH1* and *MSH2* gene mutations. The boxes represent the interquartile range of distributions (25th and 75th percentile); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles. Pairwise comparisons using Wilcoxon rank sum test showed no differences in RTL distributions between cancer-affected parents and children with germline mutations in *MLH1* ($p=0.65$), and between parents and children with mutations in *MSH2* ($p=0.67$).

ARTÍCULO 2**Longer Telomeres Are Associated with Cancer Risk in MMR-Proficient Hereditary Non-Polyposis Colorectal Cancer**

Nuria Seguí, Elisabet Guinó, Marta Pineda, Matilde Navarro, Fernando Bellido, Conxi Lázaro, Ignacio Blanco, Víctor Moreno, Gabriel Capellá, Laura Valle.

PLoS One 3;9(2):e86063; 2014. doi: 10.1371/journal.pone.0086063.

Resumen:

Este estudio está estrechamente relacionado con el artículo anterior, en el que se evaluó la longitud de los telómeros en SL. En este trabajo cuantificamos la longitud telomérica relativa en sangre periférica de 114 individuos de familias CCRf-X: 57 afectados de CCR y 57 no afectados, y valoramos los resultados en relación a los obtenidos previamente. Se definen como CCRf-X aquellas familias que cumplen los criterios clínicos de Ámsterdam pero que, a diferencia del SL, no presentan defectos en el sistema de reparación de bases desapareadas. En claro contraste con los resultados anteriores, los individuos afectados de estas familias CCRf-X presentan telómeros más largos que sus familiares sanos, así como que los controles libres de cáncer. Aun teniendo en cuenta el carácter retrospectivo del estudio, nuestros resultados sugieren que telómeros largos pueden actuar como un factor de riesgo en el CCR. Asimismo, deja patente que la asociación del acortamiento o del alargamiento de la longitud telomérica con el riesgo a desarrollar CCR en los casos hereditarios depende de forma directa del estado del sistema de reparación MMR.

Longer Telomeres Are Associated with Cancer Risk in MMR-Proficient Hereditary Non-Polyposis Colorectal Cancer

Nuria Seguí¹, Elisabet Guinó², Marta Pineda¹, Matilde Navarro¹, Fernando Bellido¹, Conxi Lázaro¹, Ignacio Blanco¹, Victor Moreno^{2,3}, Gabriel Capellá¹, Laura Valle^{1*}

1 Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain, **2** Unit of Biomarkers and Susceptibility, Catalan Institute of Oncology, IDIBELL and CIBERESP, Hospitalet de Llobregat, Barcelona, Spain, **3** Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Hospitalet de Llobregat, Barcelona, Spain

Abstract

Aberrant telomere length measured in blood has been associated with increased risk of several cancer types. In the field of hereditary non-polyposis colorectal cancer (CRC), and more particularly in Lynch syndrome, caused by germline mutations in the mismatch repair (*MMR*) genes, we recently found that cancer-affected *MMR* gene mutation carriers had shorter telomeres and more pronounced shortening of telomere length with age than controls and unaffected *MMR* gene mutation carriers. Here we evaluate blood telomere length in *MMR*-proficient hereditary non-polyposis CRC, i.e. familial CRC type X (fCRC-X). A total of 57 cancer-affected and 57 cancer-free individuals from 34 Amsterdam-positive fCRC-X families were analyzed and compared to the data previously published on 144 cancer-affected and 100 cancer-free *MMR* gene mutation carriers, and 234 controls. Relative telomere length was measured using a monochrome multiplex quantitative PCR method, following strict measures to avoid sources of bias and adjusting by age. Despite the retrospective nature of our study, the results show that longer telomeres associate with cancer risk in fCRC-X, thus identifying different patterns of telomere length according to the status of the *MMR* system.

Citation: Seguí N, Guinó E, Pineda M, Navarro M, Bellido F, et al. (2014) Longer Telomeres Are Associated with Cancer Risk in MMR-Proficient Hereditary Non-Polyposis Colorectal Cancer. PLoS ONE 9(2): e86063. doi:10.1371/journal.pone.0086063

Editor: Amanda Ewart Toland, Ohio State University Medical Center, United States of America

Received: October 8, 2013; **Accepted:** December 9, 2013; **Published:** February 3, 2014

Copyright: © 2014 Seguí et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work has been funded by the Spanish Ministry of Economy (State Secretariat for Research, Development and Innovation) (grants SAF2012-38885 to LV and SAF2012-33636 to GC; and Ramón y Cajal contract to LV); L'Oréal-UNESCO "For Women in Science"; the Scientific Foundation Asociación Española Contra el Cáncer; the Catalan Government (grant 2009SGR290); and Carlos III Health Institute (fellowship to NS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lvalle@iconcologia.net

Introduction

Family history is one of the strongest risk factors for the development of colorectal cancer (CRC) and is involved in approximately 20% of all CRC cases. However, only 2–6% of all CRCs are explained by germline mutations in known high-penetrance CRC genes. The Amsterdam criteria were defined to identify hereditary non-polyposis CRC cases, considering young age (<50 years) at cancer diagnosis and high familial aggregation of CRC (Amsterdam I) or other related tumors (Amsterdam II). Approximately 60% of the families that fulfill the Amsterdam criteria show DNA mismatch repair (*MMR*) deficiency as a consequence of a germline mutation or epimutation in a *MMR* gene, i.e. *MLH1*, *MSH2*, *MSH6* or *PMS2* (Lynch syndrome; LS). The remaining 40% do not show *MMR* defects and the genetic cause of the familial CRC aggregation is still unknown, having been grouped as familial CRC type X (fCRC-X) [1,2].

Chromosome telomeres consist of multiple short repeats (TTTAGG) that protect against large-scale genomic rearrangements. Telomeres shorten with cell division, eventually leading to cellular senescence. On rare occasions, cells that aberrantly bypass replicative senescence with critically short telomeres may develop genomic instability and potentially become tumorigenic. In cancer

cells, however, as in stem cells, telomerase, the enzyme that adds telomeric repeats to the chromosome ends, is expressed, compensating for telomere erosion and preventing senescence/apoptosis [3–5].

Germline mutations in the components of the telomerase complex cause dyskeratosis congenita. Patients with this disorder have short telomeres, which lead to bone marrow failure and increased cancer risk [6]. Likewise, mouse models with telomerase deficiency and short telomeres have high risk of cancer [7]. Recent epidemiological studies have evaluated telomere length measured in peripheral blood DNA as a potential biomarker of cancer risk. Several studies have reported associations between telomere length and cancer risk, although the data are inconsistent among studies and tumor types [8]. In CRC studies, contradictory results have been observed, apparently due to differences in study population, study design, analytical approach, sample size, or exposure to environmental factors [9–15].

Regarding hereditary CRC, our group recently reported that cancer-affected *MMR* gene mutation carriers had shorter telomeres and showed faster telomere attrition with age, measured in blood, than controls and cancer-free *MMR* gene mutation carriers [16]. Nevertheless, the role of telomere length as cancer

risk modifier in LS could not be asserted since it had been argued that the shortened telomeres observed in retrospectively collected samples from cancer-affected individuals might be a consequence of the disease [8,14]. However, the fact that cancer-free mutation carriers had longer telomeres than cancer-free controls provided additional evidence in support to the hypothesis that telomere length might act as a cancer risk modifier in LS [16].

Here we report the first study of the behavior of blood telomere length in MMR-proficient hereditary non-polyposis CRC, i.e. fCRC-X, and compare it to the behavior observed in controls and in hereditary non-polyposis CRC with a MMR defect, i.e. LS (previously published [16]).

Materials and Methods

Ethics Statement

Written informed consent was obtained from all subjects. The study was approved by the Ethics Committee of IDIBELL (ref. PR221/09).

Study Participants

A total of 114 individuals, 57 cancer-affected and 57 cancer-free, from 34 fCRC-X families were included in the study. These families fulfilled the Amsterdam criteria but did not show MMR defects (microsatellite instability or loss of expression of the MMR proteins MLH1, MSH2, MSH6 and PMS2). In all, 76.5% (26/34) of the families fulfilled Amsterdam I criteria and 23.5% (8/34) Amsterdam II. For comparative purposes, previously published telomere length data from individuals belonging to LS families and from controls were included in the analyses [16]. Familial CRC-X, LS and control samples were all of Caucasian origin and recruited from the same homogeneous population (the Spanish region of Catalonia) through the Hereditary Cancer Program of the Catalan Institute of Oncology, IDIBELL. Table 1 shows the characteristics of the fCRC-X cases studied, comparing them to the LS individuals and controls previously analyzed [16].

Relative Telomere Length Measurement

Relative telomere length (RTL) was assessed using a mono-chrome multiplex quantitative PCR method [17], under the same conditions as in Segui et al. [16]. Moreover, MMR-proficient hereditary cases were run together with LS cases and controls. In

order to rule out artifacts or technical sources of variation, a standard curve was included in each run (384-well plate), each sample was assayed in triplicate, and whenever possible, equal numbers of samples from different clinical groups were run in the same plate. Storage and DNA extraction from peripheral blood were performed at the same facility and using the same extraction methods, as recommended for retrospective telomere length studies [8].

Statistical Analyses

RTL was adjusted by age based on the telomere length shortening occurring with age in the general population (controls), as previously reported [16,18]. Differences in age-adjusted telomere lengths were analyzed using the Wilcoxon rank sum test (Mann-Whitney U). The Kruskal-Wallis rank sum test was used to compare the demographic characteristics among groups. All tests were two-sided and p-values below 0.05 were considered statistically significant. The analyses were performed using R statistical software.

Results

Age-adjusted RTL was evaluated in 57 cancer-affected fCRC-X patients and 57 unaffected individuals from 34 fCRC-X families. The results were compared to the results we published in a previous study of 144 cancer-affected *MMR* gene mutation carriers, 100 unaffected *MMR* gene mutation carriers and 234 controls [16].

Cancer-affected fCRC-X cases showed significantly longer telomeres than unaffected fCRC-X individuals ($p = 0.009$) and cancer-free controls ($p = 0.013$) (Fig. 1). These results were in sharp contrast to those previously observed in LS, where telomere lengths detected in cancer-affected *MMR* gene mutation carriers were shorter than those in unaffected mutation carriers ($p = 0.032$) [16]. On the other hand, cancer-free fCRC-X cases showed shorter telomeres than cancer-free *MMR* gene mutation carriers ($p = 0.015$), but of similar length to controls (Fig. 1).

Discussion

We found that longer telomeres are associated with cancer in fCRC-X. Moreover, based on data previously published by our

Table 1. Characteristics of the hereditary non-polyposis CRC and control groups studied. Data on Lynch syndrome families and controls were published previously [16].

	fCRC-X families (n = 34)		Lynch syndrome families (n = 96) <i>MMR</i> gene mutation carriers		² Controls	³ P-value
	¹ Cancer	Cancer-free	¹ Cancer	Cancer-free		
N	57	57	144	100	234	
Median age at blood draw (\pm SD)	56.0 (\pm 13.9)	50.0 (\pm 20.3)	51.5 (\pm 13.4)	35.0 (\pm 11.5)	42.0 (\pm 14.9)	0.037
Sex: n (%)	M: 27 (47.4) F: 30 (52.6)	M: 28 (49.1) F: 29 (50.9)	M: 74 (51.4) F: 70 (48.6)	M: 40 (40.0) F: 60 (60.0)	M: 92 (39.3) F: 142 (60.7)	0.968
Median age at cancer diagnosis (\pm SD)	49.0 (\pm 13.4)	–	43.0 (\pm 12.9)	–	–	0.278

N, number of subjects; SD, standard deviation; M, male; F, female.

¹Cancer: Individuals affected with a LS-associated cancer: CRC and/or cancer of the endometrium, ovary, stomach, small bowel, hepatobiliary tract, pancreas, upper uro-epithelial tract or brain.

²Controls include non-carriers from LS families (n = 144) and unrelated cancer-free controls (n = 90).

³Kruskal-Wallis rank sum test.

doi:10.1371/journal.pone.0086063.t001

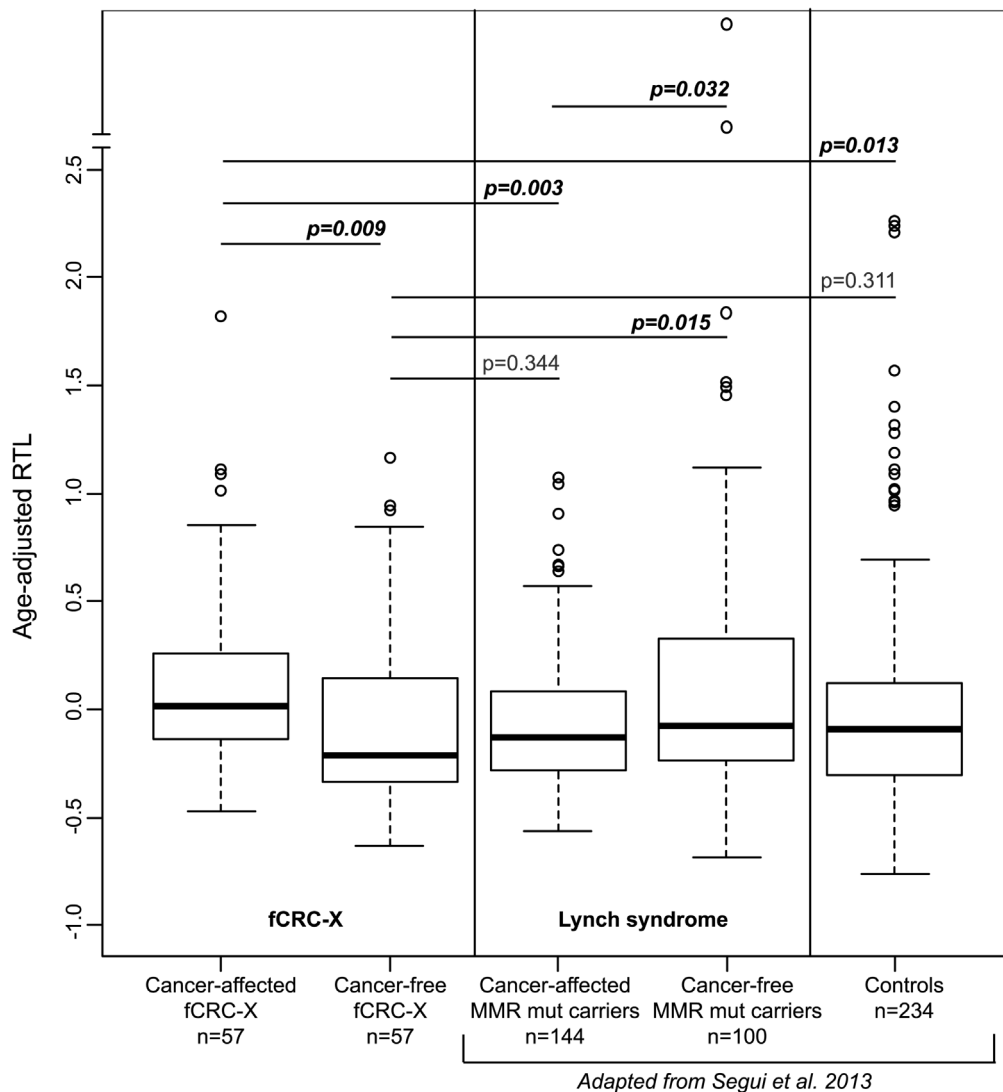


Figure 1. Age-adjusted RTL in subjects belonging to fCRC-X and LS families. The different groups correspond to: cancer-affected fCRC-X cases (median age-adjusted RTL: 0.017); cancer-free individuals from fCRC-X families (median: -0.215); cancer-affected *MMR* gene mutation carriers (median: -0.131); cancer-free *MMR* gene mutation carriers (median: -0.079); and cancer-free controls (median: -0.092). Differences in age-adjusted RTL were analyzed using the Wilcoxon rank sum test (Mann-Whitney U). The boxes represent the interquartile range of distributions (25th and 75th percentiles); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles. Data from LS families and controls were published previously [16].

doi:10.1371/journal.pone.0086063.g001

group [16], *MMR*-deficient and -proficient non-polyposis hereditary cases show distinct patterns of blood telomere length, suggesting that the status of the *MMR* system is key in defining telomere length in hereditary cases.

Several epidemiological studies have investigated the association of telomere length with CRC risk in the general population, producing conflicting results [19]. Evidence from retrospective studies indicates that telomeres in peripheral blood cells are shorter in CRC cases than controls. However, this association has not been replicated in prospective studies, suggesting that the association between short telomeres and CRC in retrospective studies is somehow the result of disease, treatment or differential survival rather than the cause (effect of reverse causation) [14]. Regarding prospective studies, no association was found in three relatively small studies (134–191 cases vs. 306–406 controls) [9–11]. However, more recently, larger prospective studies found that

longer telomeres were associated with higher CRC risk (cases/controls = 2,157/3,921) [14], or that both long and short telomeres increased CRC risk (cases/controls = 441/549) [15].

Despite the retrospective nature of our study, we found that cancer-affected fCRC-X individuals had longer telomeres than cancer-free members of the same type X families, controls, and cancer-affected members of LS families (*MMR* gene mutation carriers). Therefore, longer telomeres are associated with cancer risk in *MMR*-proficient hereditary non-polyposis CRC, as observed in large prospective population-based CRC series [14,15]. Further studies assessing telomere length before and after cancer diagnosis in fCRC-X cases will be crucial to demonstrate the precise effect of cancer on blood telomere length in this group of patients.

Previous evidence supports the hypothesis that long telomeres might increase cancer risk: first, long telomeres may delay cellular

senescence and apoptosis, increasing the chance that genetic abnormalities will accumulate [5,20]; secondly, a subset of colorectal tumors has longer telomeres than the adjacent non-tumor colon mucosa [21]. Telomere length assessment in fCRC-X tumors will provide additional insight into whether MMR-proficient hereditary tumors also have elongated telomeres. Similarly, knowing the levels of telomerase expression or of enzymatic activity in peripheral blood and tumor samples might aid the design of specific therapeutic and/or preventive approaches for fCRC-X patients in the future [22].

Our findings, together with those observed in large population-based CRC series [14,15], suggest that longer telomeres are a risk factor for hereditary and sporadic MMR-proficient CRC. Following this observation, it is plausible to hypothesize that

longer telomeres also act as modifiers of the age of onset of cancer and/or polyposis in other MMR-proficient CRC syndromes.

In conclusion, in contrast to in the observations for LS but in line with the results of large CRC prospective studies, elongated telomeres are associated with increased cancer risk in Amsterdam-positive MMR-proficient hereditary non-polyposis CRC.

Author Contributions

Conceived and designed the experiments: LV NS. Performed the experiments: NS LV. Analyzed the data: EG NS LV VM. Contributed reagents/materials/analysis tools: MP MN FB CL IB GC. Wrote the paper: LV NS. Critically revised the manuscript: MP IB GC VM.

References

- Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, et al. (2009) Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 76: 1–18.
- Ku CS, Cooper DN, Wu M, Roukos DH, Pawitan Y, et al. (2012) Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. *Mod Pathol* 25: 1055–1068.
- Blasco MA (2005) Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet* 6: 611–622.
- Stewart JA, Chaiken MF, Wang F, Price CM (2012) Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. *Mutat Res* 730: 12–19.
- Hou L, Zhang X, Gawron AJ, Liu J (2012) Surrogate tissue telomere length and cancer risk: shorter or longer? *Cancer Lett* 319: 130–135.
- Alter BP, Giri N, Savage SA, Rosenberg PS (2009) Cancer in dyskeratosis congenita. *Blood* 113: 6549–6557.
- Blasco MA (2005) Mice with bad ends: mouse models for the study of telomeres and telomerase in cancer and aging. *Embo J* 24: 1095–1103.
- Prescott J, Wentzensen IM, Savage SA, De Vivo I (2012) Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutat Res* 730: 75–84.
- Zee RY, Castonguay AJ, Barton NS, Buring JE (2009) Mean telomere length and risk of incident colorectal carcinoma: a prospective, nested case-control approach. *Cancer Epidemiol Biomarkers Prev* 18: 2280–2282.
- Pooley KA, Sandhu MS, Tyrer J, Shah M, Driver KE, et al. (2010) Telomere length in prospective and retrospective cancer case-control studies. *Cancer Res* 70: 3170–3176.
- Lee IM, Lin J, Castonguay AJ, Barton NS, Buring JE, et al. (2010) Mean leukocyte telomere length and risk of incident colorectal carcinoma in women: a prospective, nested case-control study. *Clin Chem Lab Med* 48: 259–262.
- Bozzao C, Lastella P, Ponz de Leon M, Pedroni M, Di Gregorio C, et al. (2011) Analysis of telomere dynamics in peripheral blood cells from patients with Lynch syndrome. *Cancer* 117: 4325–4335.
- Maxwell F, McGlynn LM, Muir HC, Talwar D, Benzeval M, et al. (2011) Telomere attrition and decreased fetuin-A levels indicate accelerated biological aging and are implicated in the pathogenesis of colorectal cancer. *Clin Cancer Res* 17: 5573–5581.
- Jones AM, Beggs AD, Carvajal-Carmona L, Farrington S, Tenesa A, et al. (2012) TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. *Gut* 61: 248–254.
- Cui Y, Cai Q, Qu S, Chow WH, Wen W, et al. (2012) Association of leukocyte telomere length with colorectal cancer risk: nested case-control findings from the Shanghai Women's Health Study. *Cancer Epidemiol Biomarkers Prev* 21: 1807–1813.
- Segui N, Pineda M, Guino E, Borrás E, Navarro M, et al. (2013) Telomere length and genetic anticipation in lynch syndrome. *PLoS One* 8: e61286.
- Cawthon RM (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 37: e21.
- Martinez-Delgado B, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M, et al. (2011) Genetic anticipation is associated with telomere shortening in hereditary breast cancer. *PLoS Genet* 7: e1002182.
- Savage SA, Gadalla SM, Chanock SJ (2013) The long and short of telomeres and cancer association studies. *J Natl Cancer Inst* 105: 448–449.
- Lan Q, Cawthon R, Shen M, Weinstein SJ, Virtamo J, et al. (2009) A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of non-Hodgkin lymphoma. *Clin Cancer Res* 15: 7429–7433.
- Bisoffi M, Heaphy CM, Griffith JK (2006) Telomeres: prognostic markers for solid tumors. *Int J Cancer* 119: 2255–2260.
- Hu J, Hwang SS, Liesa M, Gan B, Sahin E, et al. (2012) Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. *Cell* 148: 651–663.

ARTÍCULO 3***GALNT12* is not a major contributor of familial colorectal cancer type X**

Nuria Seguí, Marta Pineda, Matilde Navarro, Conxi Lázaro, Joan Brunet, Mar Infante, Mercedes Durán, José Luis Soto, Ignacio Blanco, Gabriel Capellá, Laura Valle.

Human Mutation 35(1):50-2; 2014. doi: 10.1002/humu.22454.

Resumen:

La hipótesis de este estudio se planteó en base a evidencias previas de que mutaciones en *GALNT12* podían explicar el riesgo a desarrollar CCR en una fracción de casos hereditarios sin causa genética identificada: Por un lado, el gen está situado cerca de un pico de ligamiento en el cromosoma 9q asociado con el CCR; y por otro lado, habían identificado mutaciones germinales que inactivan o reducen la actividad enzimática de la proteína en pacientes con CCR, algunos con historia familiar. Teniendo en cuenta estos indicios nosotros quisimos estudiar el gen en familias de alto riesgo agrupadas como CCRf-X. Secuenciamos las regiones codificantes de *GALNT12* y secuencias flanqueantes en 103 individuos índice de familias CCRf-X de distintas zonas de la geografía española: Cataluña, Castilla-León y Valencia. Se identificaron un total de 20 variantes en 90 pacientes, ocho de ellas localizadas en las regiones codificantes. Mediante estudios *in silico* de predicción de función, de cosegregación con la enfermedad, de frecuencia en población sana, y en base a evidencias previas de afectación de la actividad enzimática de la proteína, concluimos que ninguno de los cambios genéticos identificados era causante de la agregación familiar de cáncer de estos casos. Por ello, nuestros resultados desestiman a *GALNT12* como un gen de alta penetrancia para CCR y relevante para su inclusión en el diagnóstico genético del CCR hereditario.

***GALNT12* is Not a Major Contributor of Familial Colorectal Cancer Type X**

Nuria Seguí,¹ Marta Pineda,¹ Matilde Navarro,¹ Conxi Lázaro,¹ Joan Brunet,² Mar Infante,³ Mercedes Durán,³ José Luis Soto,⁴ Ignacio Blanco,¹ Gabriel Capellá,¹ and Laura Valle^{1*}

¹Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Spain; ²Hereditary Cancer Program, Catalan Institute of Oncology, IDIBGI, Girona, Spain; ³Instituto de Biología y Genética Molecular, IBGM-UVA-CSIC, Valladolid, Spain; ⁴Elche University Hospital, Elche, Spain

Communicated by Stephen J. Chanock

Received 3 July 2013; accepted revised manuscript 23 September 2013.

Published online 1 October 2013 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22454

ABSTRACT: Previous evidence indicates that mutations in the *GALNT12* gene might cause a fraction of the unexplained familial colorectal cancer (CRC) cases: *GALNT12* is located in 9q22-33, in close proximity to a CRC linkage peak; and germline missense variants that reduce the enzymatic activity of the protein have been identified in CRC patients, some of them with familial CRC history. We hypothesized that mutations in *GALNT12* might explain part of the high-risk families grouped as familial CRC type X (fCRC-X), that is, Amsterdam-positive families with mismatch repair proficient tumors. We sequenced the coding regions of the gene in 103 probands of fCRC-X families, finding no functionally relevant mutations. Our results rule out *GALNT12* as a major high CRC susceptibility gene. Additional studies are required to provide further evidence about its role as a moderate/low susceptibility gene in familial aggregation of cancer.

Hum Mutat 00:1–3, 2013. © 2013 Wiley Periodicals, Inc.

KEYWORDS: *GALNT12*; hereditary colorectal cancer; colorectal cancer susceptibility; penetrance

Family history is well established to be one of the strongest risk factors for the development of colorectal cancer (CRC) and it is thought to involve approximately 20% of all CRC cases. However, only a minority of all CRC cases (2%–6%) are explained by germline genetic mutations in well-known high penetrance genes [reviewed by Lynch et al., 2009]. The proportion of CRC families that fulfill the most stringent criteria for hereditary CRC, that is, the Amsterdam criteria, and do not show a mismatch repair (MMR) defect is high (~40%). These individuals are at increased risk of developing CRC and therefore require strict cancer surveillance strategy. The genetic cause of the CRC familial aggregation in these families is unknown and they have been grouped as familial CRC type X (fCRC-X) [reviewed by Ku et al., 2012]. Several dominantly act-

ing predisposition loci mapping to different chromosomal regions, such as 9q22.2-31.2, 5q14-q22, or 3q13.31-q27.1, have been identified through genome-wide linkage studies in CRC families, but so far no causal gene has been reported [Gray-McGuire et al., 2010; Picelli et al., 2008; Tomlinson et al., 1999].

Previous evidence indicates that mutations in the *GALNT12* gene (MIM #610290), which codes for the enzyme *N-acetylgalactosaminyltransferase-type 12*, might explain familial CRC cases of unknown etiology [Clarke et al., 2012; Guda et al., 2009]. This gene, whose protein product is involved in the O-glycosylation of mucin-type glycans, shows high expression levels in the normal colon and is downregulated in a significant proportion of colorectal tumors [Guo et al., 2002; 2004]. *GALNT12* is located in 9q21-33, in close proximity to the linkage peak in 9q22-31, recurrently found when studying familial CRC cases [Gray-McGuire et al., 2010; Kemp et al., 2006; Skoglund et al., 2006; Wiesner et al., 2003]. In 2009, Guda et al. identified functionally relevant germline *GALNT12* mutations in seven out of 272 (2.6%) colon cancer patients. Of note, no mutations were present in 192 cancer-free controls. Recently, Clarke et al. (2012) studied the occurrence of *GALNT12* mutations in a cohort of 118 CRC families of unknown genetic cause. They identified two missense mutations present in four different Bethesda-positive families. However, no mutations were detected in 26 probands that met the Amsterdam I criteria. Based on the previous evidence implicating *GALNT12* in CRC predisposition, we wanted to expand the study of the relevance of *GALNT12* mutations in the etiology of fCRC-X.

A total of 103 fCRC-X families (all Caucasian) were included in the study: forty-three had been referred to the Genetic Counseling Units of the Catalan Institute of Oncology in the Spanish region of Catalonia between 1998 and 2011; 27 families were recruited by the Hereditary Cancer Unit of the Instituto de Biología y Genética Molecular (IBGM) of the University of Valladolid and the Spanish Research Council (CSIC), between 2006 and 2011; and 33 came from the Hereditary Cancer Program of the Spanish region of Valencia, which belongs to the Valencian Biobank Network, and were collected between 2005 and 2012. All families fulfilled Amsterdam criteria and did not have MMR deficiency (either tumor microsatellite instability and/or lack of expression of the MMR proteins MLH1, MSH2, MSH6, or PMS2). Of all 103 families, 45.6% fulfilled Amsterdam I and 54.4% Amsterdam II. The mean age at cancer diagnosis was 48.4 for the sequenced probands. Informed consent was obtained from all subjects and the study received the approval of the Ethics Committee of IDIBELL (ref. PR073/12). Supp. Table S1 shows the characteristics of the families included in the study.

All exons and exon-intron boundaries were sequenced using a standard protocol for automated direct Sanger sequencing. Primer sequences and PCR conditions are available upon request.

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

*Correspondence to: Laura Valle, Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Av. Gran Via 199-203, 08908 Hospitalet de Llobregat, Barcelona, Spain. E-mail: lvalle@iconcologia.net

Contract grant sponsors: Spanish Ministry of Economy (State Secretariat for Research, Development and Innovation) (SAF2012-38885, SAF2012-33636); L'Oréal-UNESCO "For Women in Science"; Scientific Foundation Asociación Española Contra el Cáncer; Catalan Government (2009SGR290); Carlos III Health Institute.

Table 1. Germline *GALNT12* Variants Identified in 103 fCRC-X Index Patients

Nucleotide change ^a	Amino-acid change	refSNP	Location	<i>n</i>	Het/Hom	MAF	Population MAF ^b (dbSNP)	Population MAF ^c (ESP)	Reported in previous studies (MAF)
c.136G>A	p.G46R	rs10987768	Exon 1	1	1/0	0.005	0.09	–	Guda et al. (2009) (NA)
c.237C>T	p. =	–	Exon 1	1	1/0	0.005	–	–	
c.356A>T	p.E119V	rs1137654	Exon 1	18	18/0	0.087	–	0.058	Guda et al. (2009) (NA); Clarke et al. (2012) (0.08)
c.399T>C	p. =	–	Exon 2	1	1/0	0.011	–	–	
c.541+74G>T	–	rs73496150	Intron 2	2	2/0	0.010	0.022	–	
c.781G>A	p.D261N	rs41306504	Exon 4	2	2/0	0.010	0.008	0.008	Guda et al. (2009) (NA); Clarke et al. (2012) (0.015)
c.917+24C>T	–	rs41297187	Intron 4	4	3/1	0.024	0.043	0.017	
c.1036–42delT	–	rs3216734	Intron 5	31	25/6	0.180	0.151	0.163	
c.1036–4G>A	–	–	Intron 5	1	1/0	0.005	–	–	
c.1344+61G>T	–	rs3824516	Intron 7	12	11/1	0.063	0.058	–	
c.1392C>T	p. =	rs35616709	Exon 8	1	1/0	0.005	0.001	–	
c.1458+58G>T	–	rs1885608	Intron 8	81	44/37	0.573	0.230	–	
c.1497C>T	p. =	rs35632007	Exon 9	1	1/0	0.005	0.001	0.002	Clarke et al. (2012) (0.008)
c.1605+4G>A	–	rs79574929	Intron 9	1	1/0	0.005	0.006	0.012	
c.1707G>C	p. =	rs2273846	Exon 10	5	5/0	0.024	0.147	0.083	Clarke et al. (2012) (0.043)
c.*67A>T	–	rs78514784	3'-UTR	2	2/0	0.010	0.017	–	
c.*171A>G	–	rs2273847	3'-UTR	5	5/0	0.024	0.147	–	Clarke et al. (2012) (0.031)
c.*421G>A	–	rs2273848	3'-UTR	11	11/0	0.053	0.059	–	Clarke et al. (2012) (0.088)
c.*499T>A	–	–	3'-UTR	1	1/0	0.011	–	–	
c.*547dup	–	–	3'-UTR	1	1/0	0.011	–	–	

^aRefSeq NM_024642.4.^bMAF reported at the dbSNP and 1000 Genomes databases (<http://ncbi.nlm.nih.gov/projects/SNP/>).^cMAF reported at the NHLBI Exome Sequencing Project (ESP) (<http://evs.gs.washington.edu/EVS/>). Most intronic changes are not covered.

Het, heterozygous; Hom, homozygous; MAF, minor allele frequency; NA, not available.

Table 2. In Silico Prediction of Functional Effects of the Identified Variants

Variant	refSNP	Location	Protein prediction (score)			Splice site (SS) prediction ^a			
			PolyPhen-2 (HumDiv / HumVar)	Condel	SIFT	Distance to the nearest SS (bp)	SS	WT score	Variant score
c.136G>A (p.G46R)	rs10987768	Exon 1	Benign (0.002) / benign (0.001)	Neutral (0.011)	Tolerated (0.18)	236	D	0.8	0.8
c.237C>T (p. =)	–	Exon 1	–	–	–	135	D	0.8	0.8
c.356A>T (p.E119V)	rs1137654	Exon 1	Possibly damaging (0.791) / benign (0.198)	Neutral (0.011)	Tolerated (0.18)	16	D	0.8	0.8
c.399T>C (p. =)	–	Exon 2	–	–	–	28	A	0.89	0.89
c.541+74G>T	rs73496150	Intron 2	–	–	–	143	D	1	1
c.781G>A (p.D261N)	rs41306504	Exon 4	Probably damaging (1)/probably damaging (0.968)	Neutral (0.449)	Tolerated (0.22)	74	D	1	1
c.917+24C>T	rs41297187	Intron 4	–	–	–	50	A	0.56	0.56
c.1036–42delT	rs3216734	Intron 5	–	–	–	137	D	0.87	0.87
c.1036–4G>A	–	Intron 5	–	–	–	24	D	0.87	0.87
c.1344+61G>T	rs3824516	Intron 7	–	–	–	42	A	0.95	0.95
c.1392C>T (p. =)	rs35616709	Exon 8	–	–	–	4	A	0.95	0.98
c.1458+58G>T	rs1885608	Intron 8	–	–	–	61	D	0.98	0.98
c.1497C>T (p. =)	rs35632007	Exon 9	–	–	–	48	A	0.88	0.88
c.1605+4G>A	rs79574929	Intron 9	–	–	–	67	D	0.99	0.99
c.1707G>C (p. =)	rs2273846	Exon 10	–	–	–	58	D	0.99	0.99
c.*67A>T	rs78514784	3'-UTR	–	–	–	42	A	0.8	0.8
c.*171A>G	rs2273847	3'-UTR	–	–	–	106	D	0.96	0.96
c.*421G>A	rs2273848	3'-UTR	–	–	–	4	D	0.96	1
c.*499T>A	–	3'-UTR	–	–	–	102	A	0.98	0.98
c.*547dup	–	3'-UTR	–	–	–	208	A	0.98	0.98
						312	A	0.98	0.98
						562	A	0.98	0.98
						640	A	0.98	0.98
						684	A	0.98	0.98

^aPrediction calculated by NNSplice 0.9.For splice site prediction, major alterations represent creation or destruction of a splice site, or score modifications $\geq 45\%$.

SS, splice site; bp, base pairs; WT, wild-type; A, acceptor consensus splice site; D, donor consensus splice site; D2, new donor splice site; NR, splice site not recognized.

Sequencing was performed on an ABI Sequencer 3730 and data analyzed using Mutation Surveyor v. 3.10. Identified variants were submitted to the LOVD database (<http://www.lovd.nl/GALNT12>).

A total of 20 variants were identified in 90 patients. Of the 20 variants, eight were located in protein-coding regions, seven in introns, and five in the 3'-UTR. Three of the eight changes in protein-coding regions, c.136G>A (p.G46R), c.356A>T (p.E119V), and c.781G>A

(p.D261N), were nonsynonymous. The change p.E119V was identified in 18 fCRC-X patients, p.D261N in 2, and p.G46R in 1 patient. The three variants had been previously described in public databases with population minor allele frequencies of 0.06, 0.008, and 0.09, respectively (Table 1). Also, previous studies identified these three variants both in cancer cases and controls [Clarke et al., 2012; Guda et al., 2009]. No relevant effects on the protein were predicted by the

in silico algorithms SIFT and Condel, while PolyPhen-2 predicted functional effects for p.D261N (score ~1) and p.E119V (score 0.8, when applying HumanDiv) (Table 2). Nevertheless, Guda et al. (2009) reported that none of these three variants altered the enzymatic activity of the protein.

The remaining five variants in protein-coding regions predicted to translate into synonymous amino-acid changes, affecting codons 79, 133, 464, 499, and 569. All but two, c.237C>T and c.399C>T, had been previously described in public databases (dbSNP or ESP). The population allelic frequencies reported for other two, c.1392C>T and c.1497C>T, were below 2% (Table 1). The in silico algorithm NNSplice [Reese et al., 1997] only predicted a possible effect on the splicing process for c.237C>T, where a new donor splice would be created (score 0.52) (Table 2). However, the generation of this new donor splice site was not predicted by other algorithms, such as NetGene2 [Hebsgaard et al., 1996] or SoftBerry [Burset et al., 2001]. Moreover, this variant did not segregate with the disease in the family: the father of the proband, diagnosed with rectal cancer at age 64, did not carry the variant, being therefore absent in the CRC-affected family branch.

Of the intronic variants identified, only one, c.1036-4G>A, had not been reported as a polymorphism. Likewise, the change c.1605+4G>A (rs79574929) had a population MAF of 0.006–0.012 (sources: dbSNP and ESP, respectively). Despite their proximity to the corresponding consensus splice sites, no relevant effect on the splicing process was predicted (Table 2). Regarding c.1036-4G>A, it was not present in the mother of the proband, who was diagnosed with two metachronous colon tumors at the age of 57 and 59 years, discarding it as the genetic cause of the CRC familial aggregation. Unfortunately, no cosegregation analysis could be performed for c.1605+4G>A. In summary, no functionally relevant variants were identified in any of the 103 fCRC-X families evaluated.

The results obtained in our series ($n = 103$ fCRC-X families), in line with those of Clarke et al., who found no mutations in 26 Amsterdam-positive families, support the notion that *GALNT12* is not a major high-penetrance gene for CRC predisposition. The lack of information on the familial cancer history of the carriers of *GALNT12* germline mutations in Guda et al. (2009) does not allow further investigation in this regard.

Considering the presence of *GALNT12* mutations in non-Amsterdam familial CRC cases, Clarke et al. identified two missense variants, c.907G>A (p.D303N) and c.1187A>G (p.Y396C) in 4 families that met the revised Bethesda criteria [Clarke et al., 2012]. The c.907G>A change, rs145236923, identified in three different families, partially inactivates the enzymatic activity of the protein, in contrast to the other mutations, germline and somatic, identified by Guda et al. (2009) which cause almost total inactivation of the enzyme. Moreover, the evidence of cosegregation of c.907G>A with CRC in one of the two families studied was weak. The second missense variant identified, c.1187A>G (p.Y396C), is not reported in public databases, is located in the catalytic domain of the protein, and in silico algorithms predict it to be functionally relevant [Clarke et al., 2012]. All reported patients carrying inactivating germline mutations developed CRC later in life (median age: 71), and half of them were diagnosed with multiple primary epithelial tumors, including breast and colon cancers [Guda et al., 2009]. Further studies are required to provide a definitive answer about the role of *GALNT12* mutations in this subset of familial cancer cases. Either germline mutations in *GALNT12*, as a moderate-risk gene, or in other genes located within the 9q22-31 linkage peak might still explain some familial CRC cases. Indeed, the strength of the studies that replicate the linkage signal in 9q22-31, which suggest that the disease

locus housed on 9q is specific to a familial syndrome with a phenotype of young age of onset and/or severity of colon neoplasia [Gray-McGuire et al., 2010], supports the idea that probably the other genes under the linkage peak, such as *ZNF367*, *HABP4*, and *GABBR2*, might be relevant in CRC susceptibility.

In conclusion, our findings indicate that *GALNT12* is not a major high penetrance gene for CRC. Further comprehensive studies in Amsterdam-negative CRC families are required to clarify the gene's role as moderate/low susceptibility gene in CRC familial aggregation.

Acknowledgment

Disclosure statement: The authors declare no conflict of interest.

References

- Burset M, Seledtsov IA, Solovyev VV. 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res* 29(1):255–259.
- Clarke E, Green RC, Mahoney K, Parfrey PS, Youngusband HB, Woods MO. 2012. Inherited deleterious variants in *GALNT12* are associated with CRC susceptibility. *Hum Mutat* 33(7):1056–1058.
- Gray-McGuire C, Guda K, Adrianto I, Lin CP, Natale L, Potter JD, Newcomb P, Poole EM, Ulrich CM, Lindor N, Goode EL, Fridley BL, et al. 2010. Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. *Cancer Res* 70(13):5409–5418.
- Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, et al. 2009. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci USA* 106(31):12921–12925.
- Guo JM, Chen HL, Wang GM, Zhang YK, Narimatsu H. 2004. Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-12 in gastric and colonic cancer cell lines and in human colorectal cancer. *Oncology* 67(3–4): 271–276.
- Guo JM, Zhang Y, Cheng L, Iwasaki H, Wang H, Kubota T, Tachibana K, Narimatsu H. 2002. Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12. *FEBS Lett* 524(1–3):211–218.
- Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* 24(17):3439–3452.
- Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, et al. 2006. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. *Cancer Res* 66(10):5003–5006.
- Ku CS, Cooper DN, Wu M, Roukos DH, Pawitan Y, Soong R, Iacopetta B. 2012. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. *Mod Pathol* 25(8):1055–1068.
- Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 76(1):1–18.
- Picelli S, Vandrovcova J, Jones S, Djureinovic T, Skoglund J, Zhou XL, Velculescu VE, Vogelstein B, Lindblom A. 2008. Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. *BMC Cancer* 8:87.
- Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* 4(3):311–323.
- Skoglund J, Djureinovic T, Zhou XL, Vandrovcova J, Renkonen E, Iselius L, Bisgaard ML, Peltomaki P, Lindblom A. 2006. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32–31.1. *J Med Genet* 43(2):e7.
- Tomlinson I, Rahman N, Frayling I, Mangion J, Barfoot R, Hamoudi R, Seal S, Northover J, Thomas HJ, Neale K, Hodgson S, Talbot I, et al. 1999. Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology* 116(4):789–795.
- Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, Markowitz SD. 2003. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2–31.2. *Proc Natl Acad Sci USA* 100(22):12961–12965.

ARTÍCULO 4**Exome sequencing identifies *MUTYH* mutations in a family with an atypical phenotype**

Nuria Seguí, Matilde Navarro, Marta Pineda, Nicole Köger, Fernando Bellido, Sara González, Olga Campos, Silvia Iglesias, Rafael Valdés-Mas, Adriana López-Doriga, Marta Gut, Ignacio Blanco, Conxi Lázaro, Gabriel Capellá, Xose S. Puente, Guido Plotz, Laura Valle.

Gut Publicado online: 1 Abril 2014. doi: 10.1136/gutjnl-2014-307084.

Resumen:

Se estima el efecto atribuible a factores hereditarios en más del 20% de casos de CCR, sin embargo gran parte de la predisposición al CCR permanece sin explicación. Presentamos los resultados obtenidos al estudiar una familia con 6 miembros afectados de CCR en dos generaciones, que cumple los criterios de Ámsterdam I pero cuya base genética no se había identificado en el diagnóstico de rutina de cáncer familiar. La secuenciación de exomas de 4 afectados y un individuo sano de edad avanzada con progenie sin cáncer reveló la presencia en todos los familiares afectados de la mutación patogénica en *MUTYH*, c.1147delC (p.Ala385Profs*23). Además, los miembros afectados de la segunda generación eran portadores de una variante adicional en el mismo gen, c.43A>G (p.Met15Val). Esta variante, no descrita anteriormente, inactiva el codón de inicio de las isoformas nucleares (β y γ) de *MUTYH*, especialmente relevantes en el colon. Esta familia presenta unas características fenotípicas y moleculares atípicas para el CCR asociado a *MUTYH*, que es una enfermedad recesiva asociada al desarrollo de pólipos en el colon. En concreto, esta familia presentaba un patrón de herencia aparentemente autosómico dominante, pero con una gran diferencia en la edad de desarrollo del cáncer entre las dos generaciones afectas; ausencia o escasez de pólipos en el momento del diagnóstico; y presencia de un tumor con IMS. En base a estos hallazgos, recomendamos que se tengan en cuenta estas características para el diagnóstico genético de rutina, donde el cribado universal de la mutación *KRAS* c.34G>T (p.G12C), muy frecuente en tumores asociados a *MUTYH* y muy infrecuente en otros CCRs, podría ayudar a identificar familias portadoras con fenotipos inusuales.

LETTER

Exome sequencing identifies *MUTYH* mutations in a family with colorectal cancer and an atypical phenotype

Ma *et al*¹ comprehensively assessed the association of previously reported genetic variants with colorectal cancer (CRC) risk. The meta-analyses revealed strong evidence for association with rare *MUTYH* variants, even when excluding cases with *MUTYH*-associated polyposis. An article by Nieuwenhuis *et al*² accurately defined the phenotypical features of *MUTYH*-associated polyposis. However, the study was performed on clinic-based series ascertained based on the inheritance model or the presence of polyps, which may miss additional phenotypes relevant to improve the disease characterisation and therefore, its genetic diagnosis. To illustrate this, we report a family with a clinical phenotype that resembled Lynch syndrome but was caused by *MUTYH* mutations.

To identify novel hereditary CRC genes, we studied an Amsterdam I family (hereditary non-polyposis CRC) with no mutations in the DNA mismatch repair (MMR) genes (figure 1, table 1). By exome sequencing performed on four cancer-affected (II.2, II.6, III.1 and III.6) and one cancer-free (III.5) family members, we identified a total of 11 unreported or rare heterozygous variants present in the cancer-affected individuals (see online supplementary table S1). One of them was *MUTYH* c.1147delC (p. Ala385Profs*23) (NM_001128425.1), an European recurrent mutation.³ Cancer-affected individuals of the third generation also carried a novel *MUTYH* variant: c.43A>G (p.Met15Val). The presence of biallelic *MUTYH* mutations in II.2 and II.6 that could explain the pseudodominant transmission was discarded.

The functional studies performed for c.43A>G are described in online supplementary material. The variant inactivates the start codon of the two transcripts encoding the nuclear *MUTYH* isoforms (β and γ), highly relevant in ascending colon,⁴ the location of \geq five tumours developed in the family.

KRAS c.34G>T (p.G12C), hallmark of *MUTYH*-associated carcinomas,⁵ was present in the tumours developed by two *MUTYH* biallelic mutation carriers but not in the tumour developed by II.6, carrier of only c.1147delC. The MMR-deficient tumour developed by a *MUTYH* biallelic mutation carrier had a transition in the same *KRAS* codon; c.35G>A (p.G12D) (table 1).

The features of this family suggest that the selection criteria proposed for *MUTYH* testing,⁶ might fail to detect a number of mutated families due to infrequent phenotypes. First, *MUTYH* heterozygous mutations may, probably in the presence of other cancer risk factors, provide an increased risk of developing cancer in heterozygous carriers,^{5 7} and thereby disguise the *MUTYH* recessive inheritance to look like autosomal-dominant. One should be suspicious when an extreme anticipation in the age of cancer onset is observed between two affected generations. Second, absence or scarcity of polyps, even at relatively advanced ages (early 50 s) and with a prior CRC diagnosis, can occur in biallelic mutation carriers. When this occurs in several cancer-affected

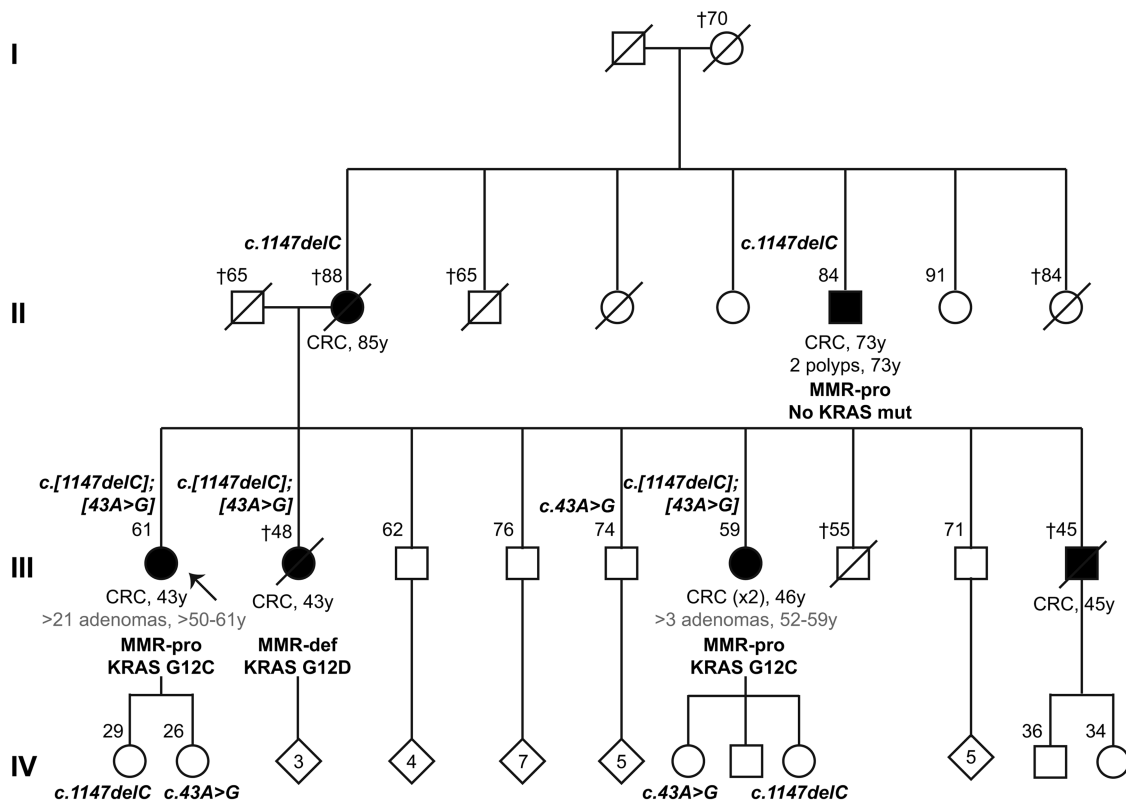


Figure 1 Family pedigree. Ages at information gathering or at death, when available, are indicated on the top left corner of each individual's symbol. Germline *MUTYH* mutations identified are indicated on the top left corner (above the age). Tumour *KRAS* mutations in codon 12 are indicated below the CRC and/or polyposis diagnosis information. In grey, number of adenomas identified at follow-up screenings. Filled symbol, CRC; numbers within the symbols, number of children; arrow, index case. CRC, colorectal cancer; y, years at cancer or polyposis diagnosis; MMR-pro, DNA mismatch repair proficiency in the tumour; MMR-def, DNA mismatch repair deficiency in the tumour.

Table 1 Characteristics of the tumours developed by family members and information on the number of polyps detected at diagnosis and at follow-up

Family member	<i>MUTYH</i> mutation	Age at CRC diagnosis	Tumour location	Stage	MMR status	<i>KRAS</i> mutation (codon 12)	Polyp no. at diagnosis and Ams I classification	Cumulative no. polyps at follow-up (age)
II.2	c.1147delC	85	Colon	n.a.	n.a.	n.a.	n.a.	Deceased
II.6	c.1147delC	73	Cecum	pT3pN0pM0	MMR-pro	No	2	No follow-up
III.1	c.[1147delC]; [43A>G]	43	Right colon	pT3pN0pM0	MMR-pro	c.34G>T (p.G12C)	0	n~34 (>50–61 years)
III.2	c.[1147delC]; [43A>G]	43	Right colon	pT3pN2pM0	MMR-def	c.35G>A (p.G12D)	0	Deceased
III.6 (a)	c.[1147delC]; [43A>G]	46	Right colon	pT3pN0pM0	MMR-pro	c.34G>T (p.G12C)	0	n=7 (52–59 years)
III.6 (b)	c.[1147delC]; [43A>G]	46	Right colon	pTis	n.a.	n.a.	0	n=7 (52–59 years)
III.9	n.a.	45	Colon	Disseminated	n.a.	n.a.	n.a.	Deceased

Ams I, Amsterdam I; CRC, colorectal cancer; MMR, DNA mismatch repair; MMR-def, MMR deficient; MMR-pro, MMR proficient; n.a., not available information.

mutation carriers within the same family it can lead to a misdiagnosis of hereditary non-polyposis CRC. Third, the presence of MMR-deficient tumours should not be an exclusion criterion for *MUTYH* genetic screening.⁵ Finally, as previously proposed, somatic *KRAS* codon 12 analysis might help select patients with CRC eligible for *MUTYH* testing,^{8,9} regardless of the presence of polyps or even in absence of a clear recessive pattern of inheritance.

The registry of colonoscopy follow-up results in family members of genetically uncharacterised families is key to, in cryptic polyposis cases, finally redirect the hereditary cancer diagnosis to right gene tests. In the studied family, no polyps had been detected at CRC diagnosis and at genetic counselling in any of the three *MUTYH* compound heterozygotes, however, several years later a total of 34 and 7 polyps (21 and 3 adenomas, respectively) were reported in two of them.

Targeted next generation sequencing using comprehensive panels of genes applied to routine genetic testing of familial cancers will overcome the problem of overlapping phenotypes and complex patterns of inheritance.

Nuria Seguí,¹ Matilde Navarro,¹ Marta Pineda,¹ Nicole Köger,² Fernando Bellido,¹ Sara González,¹ Olga Campos,¹ Sílvia Iglesias,¹ Rafael Valdés-Mas,³ Adriana López-Doriga,⁴ Marta Gut,⁵ Ignacio Blanco,¹ Conxi Lázaro,¹ Gabriel Capellá,¹ Xose S Puente,³ Guido Plotz,² Laura Valle¹

¹Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Spain

²Biomedical Research Laboratory, Medizinische Klinik 1, Universitätsklinikum Frankfurt, Frankfurt, Germany

³Department of Biochemistry and Molecular Biology,

Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo, Spain

⁴Unit of Biomarkers and Susceptibility, Catalan Institute of Oncology, IDIBELL and CIBERESP, Hospitalet de Llobregat, Spain

⁵Centro Nacional de Análisis Genómico, Fundació Parc Científic de Barcelona, Barcelona, Spain

Correspondence to Dr Laura Valle, Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Av. Gran Via 199-203, Hospitalet de Llobregat 08908, Spain; lvalle@iconcologia.net and to Dr Guido Plotz, Medizinische Klinik 1, Biomedizinisches Forschungslabor, Haus 11, Theodor-Stern-Kai 7, Universitätsklinikum Frankfurt, Frankfurt 60590, Germany; plotz@med.uni-frankfurt.de

LV and GP contributed equally to this work.

Contributors LV, GC, CL and GP conceived and designed the study and experiments. MN, MP, IB, OC, SG, CL, GC and LV collected data, performed genetic analyses, and provided samples and clinical data. NS, MG, FB, NK, OC, GP and LV performed the experiments and analysed the results. XSP, RV-M and AL-D analysed the next-gen sequencing data. LV and GP wrote the manuscript. All authors critically reviewed the manuscript and provided final approval.

Funding This work was funded by the Spanish Ministry of Economy and Competitiveness (State Secretariat for Research, Development and Innovation) (grants SAF2012-38885 and SAF2010-21165 to LV and XSP, Ramón y Cajal contract to LV, fellowship to FB, and RTICC networks RD12/0036/0031 and RD12/0036/0008); the Deutsche Krebshilfe (grant 110799 to GP); the Carlos III Health Institute (fellowship to NS); L'Oréal-UNESCO "For Women in Science"; the Scientific Foundation Asociación Española Contra el Cáncer; and the Government of Catalonia (grant 2009SGR290).

Ethics approval IDIBELL Ethics Committee.

Provenance and peer review Not commissioned; internally peer reviewed.

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2014-307084>).

To cite Seguí N, Navarro M, Pineda M, *et al.* Gut Published Online First: [please include Day Month Year]

doi:10.1136/gutjnl-2014-307084

Received 20 February 2014

Accepted 10 March 2014

Gut 2014;0:1–2. doi:10.1136/gutjnl-2014-307084

REFERENCES

- Ma X, Zhang B, Zheng W. Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Gut* 2014;63:326–36.
- Nieuwenhuis MH, Vogt S, Jones N, *et al.* Evidence for accelerated colorectal adenoma—carcinoma progression in *MUTYH*-associated polyposis? *Gut* 2012;61:734–8.
- Quesada V, Conde L, Villamor N, *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet* 2011;44:47–52.
- Plotz G, Casper M, Raedle J, *et al.* *MUTYH* gene expression and alternative splicing in controls and polyposis patients. *Hum Mutat* 2012;33:1067–74.
- Nielsen M, Morreau H, Vasen HF, *et al.* *MUTYH*-associated polyposis (MAP). *Crit Rev Oncol Hematol* 2011;79:1–16.
- Brand R, Nielsen M, Lynch H, *et al.* *MUTYH*-associated polyposis. In: Pagon RA, Adam MP, Bird TD, *et al.*, eds. *GeneReviews™ [Internet]*. Seattle, WA: University of Washington, Seattle: 1993–2014.
- Farrington SM, Tenesa A, Barnetson R, *et al.* Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* 2005;77:112–19.
- van Puijnenbroek M, Nielsen M, Tops CM, *et al.* Identification of patients with (atypical) *MUTYH*-associated polyposis by *KRAS2* c.34G>T prescreening followed by *MUTYH* hotspot analysis in formalin-fixed paraffin-embedded tissue. *Clin Cancer Res* 2008;14:139–42.
- Guarinos C, Juarez M, Egoavil CM, *et al.* Prevalence and characteristics of *MUTYH*-associated polyposis in patients with multiple adenomatous and serrated polyps. *Clin Cancer Res* 2014;20:1158–68.

SUPPLEMENTARY MATERIAL

RESULTS

***MUTYH* c.43A>G (p.Met15Val) functional analyses**

The *in silico* algorithm NNSplice0.9 predicted no splicing alterations. The potential impact of this novel variant on the function of the protein was assessed using different prediction servers PolyPhen-2, SIFT and Condel; all of them predicting damaging functional effects for p.Met15Val. This variant not only affects an evolutionary conserved methionine of α isoforms (mitochondrial), it also inactivates the start codon of the two transcripts encoding the nuclear *MUTYH* isoforms (β and γ).

To assess the functional effect of c.43A>G, we first tested if it affects the mRNA expression or stability of *MUTYH* transcripts, using qPCRs with RNA extracted from lymphocytes of two controls, three c.1147delC heterozygous carriers (IV.1, IV.24, and an unrelated carrier), one c.43A>G carrier (IV.2), and two c.[1147delC];[43A>G] compound heterozygotes (III.1 and III.6). This analysis revealed large inter-individual variation in *MUTYH* expression, overall and isoform-specific. When compared to controls and c.1147delC carriers, where a specific downexpression of the nuclear transcripts was not expected, the β and γ mRNA expression variability (reduction) observed in c.43A>G carriers might fall within the range of inter-individual variation (Suppl. Figure 1).

It has been previously observed that the expression of *MUTYH* nuclear transcripts is much lower in lymphocytes than in intestinal cells.[1] Therefore, In order to approximate the conditions that occur in colon, lymphocytes from three controls, one c.1147delC carrier (IV.24), one c.43A>G carrier (IV.2) and one c.[1147delC];[43A>G] compound heterozygous carrier (III.1) were cultured for 7 days to induce the expression of *MUTYH* nuclear isoforms. The qPCR results confirmed the increase of *MUTYH* β and γ transcripts (Suppl. Figure 2A), probably triggered by growth signals present under culture conditions.[1] A statistically significant reduction in the expression of β and γ transcripts in c.43A>G carriers compared to the wildtype controls and the c.1147delC carrier was detected (Suppl. Figure 2B). Nevertheless, this expression reduction does not represent the complete loss of mRNA expression of the c.43A>G allele. Transcript-specific reverse

transcription (RT)-PCR and sequencing confirmed these results, showing subtle or no mRNA expression reduction of nuclear isoforms in the c.43G allele (Suppl. Figure 3). Moreover, isoform-specific qPCRs performed in cultured lymphocytes treated with puromycin, which blocks the nonsense-mediated mRNA decay (NMD) mechanism, did not show any difference compared to untreated lymphocytes of c.43A>G carriers (data not shown).

Regarding protein effects, the genetic change c.43A>G leads to β and γ transcripts where the first in-frame translational start codon is located in exon 4, i.e. 92 aminoacids (276 bp) downstream of the original start codon. It is identical to the postulated protein translated from mRNA BAA89339. This protein, however, has no relevant functional elements since a highly conserved loop that binds the misincorporated adenine during the repair process is missing, and it is extremely unstable.[1] In summary, the evidence herein gathered suggests that the main functional effect of the c.43A>G variant probably depends on the disruption of the translation start of functional nuclear (β and γ) *MUTYH* isoforms.

METHODS

Whole exome capture and massive parallel sequencing

Library preparation for capturing selected DNA regions was performed according to Agilent's SureSelect protocol for Illumina paired-end sequencing (SureSelect XT Human All Exon V3, 50Mb, Agilent Technologies, Santa Clara, CA). In brief, 3.0 μ g of genomic DNA were sheared on a Covaris™ E220 instrument. Fragment size (150-300bp) and quantity were checked with the Agilent 2100 Bioanalyzer 7500 chip. Fragmented DNA was end-repaired, adenylated and ligated to Agilent indexing-specific paired-end adaptors. The DNA with adaptor-modified ends was PCR amplified (6 cycles, Herculase II fusion DNA polymerase) with SureSelect Primer and SureSelect Pre-capture Reverse PCR primers, quality controlled on the DNA 7500 assay for the library size range of 250 to 450 bp, and hybridized for 24hrs at 65°C (Applied Biosystems 2720 Thermal Cycler, Life Technologies, Carlsbad, CA) . The hybridization mix was washed in the presence of magnetic beads (Dynabeads MyOne Streptavidin T1, Life Technologies). The eluted fraction was PCR amplified (16 cycles) in order to add the index tags using SureSelectXT

Indexes for Illumina. The final library size and concentration was determined on Agilent 2100 Bioanalyzer 7500 chip and sequenced on an Illumina HiSeq 2000 platform with paired end run of 2x76bp following the manufacturer's protocol. Images generated by Illumina HiSeq 2000 were processed using the manufacturer's software to generate FASTQ sequence files.

Sequencing data analysis

Reads were aligned to the reference genome (GRCh37) using Burrows-Wheeler analysis (BWA) with the sample option,[2] and a BAM file was generated using SAMtools.[3] PCR-duplicates were removed using SAMtools and custom scripts, and initial SNP calling was performed using a combination of SAMtools and Sidrón as previously described.[4] Common variants, defined as those present in dbSNP135 with a minor allele frequency $\geq 1\%$, were filtered out.

In silico functional prediction of genetic variants

Protein damage prediction of missense genetic variants was performed by using the *in silico* algorithms PolyPhen-2,[5] SIFT,[6] and Condel.[7] Possible alterations of the splice sites were evaluated using NNSplice0.9.[8]

qPCR quantification of MUTYH

Reverse transcription and qPCR were used to quantify the three different transcript forms of *MUTYH* according to the MIQE guidelines, as previously described.[1] *MUTYH* c.43A>G sequence signals were assessed by specifically amplifying the three different *MUTYH* transcripts with the following primers: α -transcript: 5'-ctcctcgtggctagttcagg (F); β -transcript: 5'-ctcgcggcggaacgcgg (F); γ -transcript: 5'-taattgcctattggcctgtg (F); and 5'-catcacaggcagaaggc (R) for all transcripts. PCRs were performed according to standard conditions. Purified PCR products were sequenced with the reverse primer using BigDye v3.1, and analyzed on an ABI 3130XL sequencer.

Direct automated sequencing

For segregation analysis, exons 2 and 12 of *MUTYH* and their corresponding flanking regions were sequenced using a standard protocol for automated direct Sanger sequencing.

Codons 12 and 13 of the *KRAS* gene were amplified and sequenced in DNA extracted from FFPE tumor samples. Primer sequences and PCR conditions are available upon request. Sequencing was performed on an ABI Sequencer 3730 and data analyzed using Mutation Surveyor v.3.10.

REFERENCES

1. Plotz G, Casper M, Raedle J, *et al.* MUTYH gene expression and alternative splicing in controls and polyposis patients. *Hum Mutat* 2012;**33**:1067-74.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754-60.
3. Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078-9.
4. Puente XS, Pinyol M, Quesada V, *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2011;**475**:101-5.
5. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248-9.
6. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 2002;**12**:436-46.
7. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 2011;**88**:440-9.
8. Reese MG, Eeckman FH, Kulp D, *et al.* Improved splice site detection in Genie. *J Comput Biol* 1997;**4**:311-23.

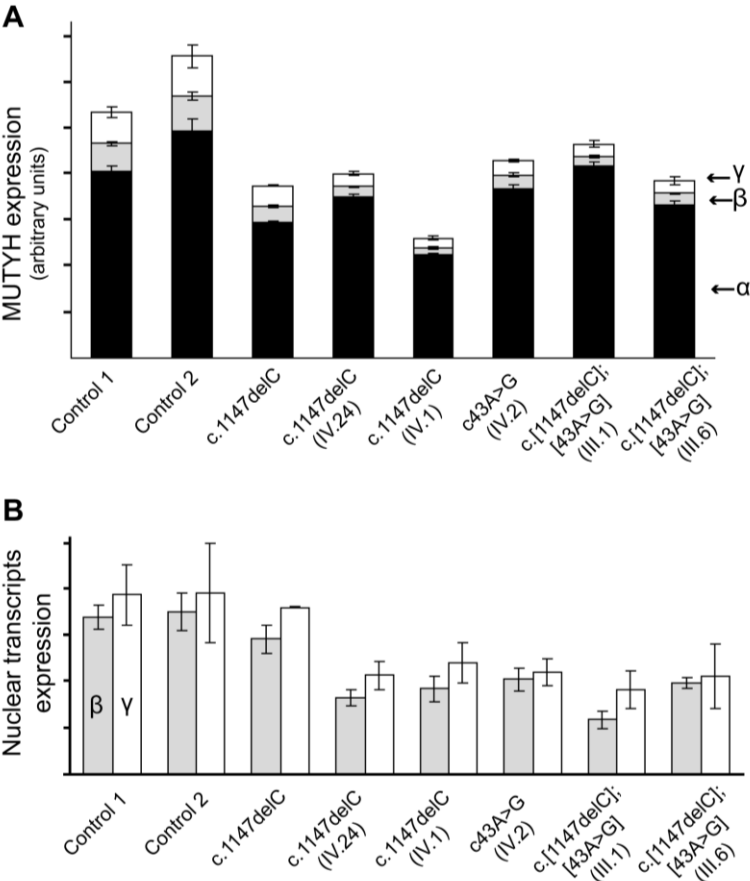
TABLES

Supplementary Table 1. Rare or unreported genetic variants identified by exome sequencing shared by the studied cancer-affected family members.

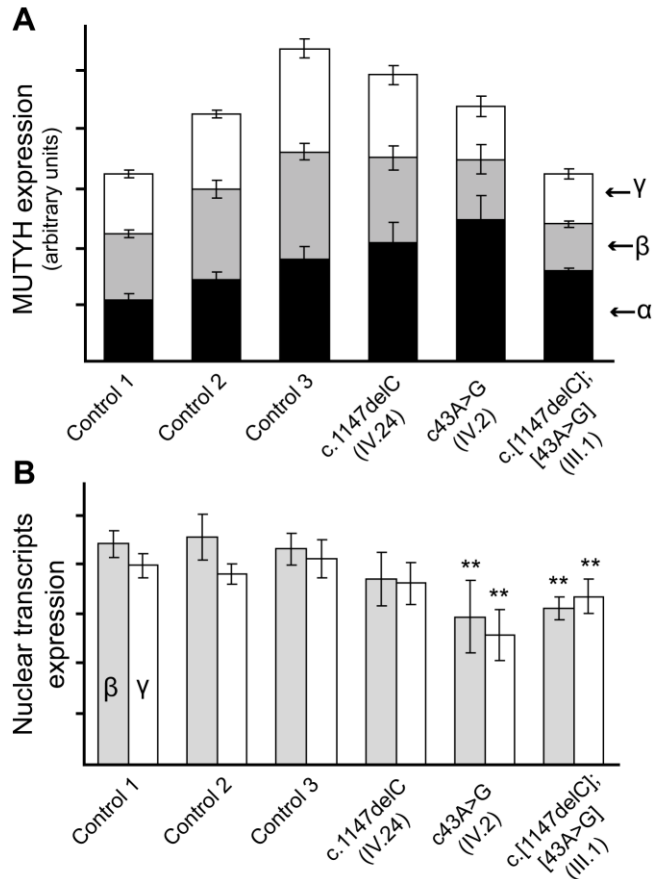
Gene	Nucleotide change	Aminoacid change	Location	Protein prediction (score)			Splice site prediction
				PolyPhen-2 (HumDiv / HumVar)	SIFT	Condel	
ANXA6	c.934G>A	p.Glu312Lys	Exon 13	Probably damaging (0,964) / (0,790)	Damaging (0)	Deleterious (0,842)	No change
*COL6A3	c.8846C>T	p.Pro2949Leu	Exon 40	Probably damaging (1) / (0,996)	NA	Neutral (0)	No change
C22orf33	c.356C>T	p.Thr119Met	Exon 4	Possibly damaging (0,648) / Benign (0,075)	Tolerated (0,1)	Neutral (0,314)	No change
DDIT4L	c.370A>T	p.Ile124Phe	Exon 3	Probably damaging (0,993) / (0,929)	Damaging (0,01)	Deleterious (0,746)	No change
MUTYH	c.1147delC	p.Ala385Profs*23	Exon12	-	-	-	-
*NRP2	c.1198G>A	p.Ala400Thr	Exon 8	Benign (0) / (0)	Tolerated (0,26)	Neutral (0,007)	No change
PRSS33	c.89A>G	p.Gln30Arg	Exon 3	Benign (0,002) / (0,005)	Tolerated (0,62)	Neutral (0,350)	No change
RPL19	c.179G>A	p.Arg60Gln	Exon 3	Benign (0,118) / (0,062)	Damaging (0)	Neutral (0,007)	No change
SLC8A3	c.876G>A	p.Met292Ile	Exon 2	Benign (0,021) / (0,016)	Damaging (0,015)	Neutral (0,013)	No change
SPPL2C	c.418delC	p.Arg140Alafs*68	Exon 1	-	-	-	-
UGT1A6	c.242C>T	p.Pro81Leu	Exon 1	Benign (0,063) / (0,052)	Tolerated (0,06)	Neutral (0,378)	No change

* variants reported at the NHLBI Exome Sequencing Project (Exome Variant Server) with minor allele frequencies (MAF) of 0.012 in individuals of European origin.

FIGURES

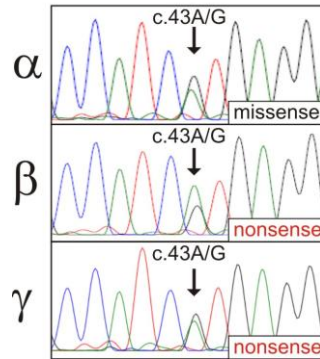


Suppl. Figure 1. qPCR quantification of *MUTYH*. Overall (A) and transcript-specific (A and B) mRNA expression quantification was performed in lymphocyte cDNA from two wildtype controls, three c.1147delC, one c.43A>G, and two c.[1147delC];[43A>G] compound heterozygous carriers.



Suppl. Figure 2. qPCR amplification of *MUTYH* (overall and isoform-specific expression).

A) Overall and relative mRNA expression of the α , β , and γ transcripts quantified with transcript-specific qPCR in cDNA obtained from cultured (non-immortalized) lymphocytes of three wildtype controls, one c.1147delC, one c.43A>G and one c.[1147delC];[43A>G] mutation carriers. **B)** Expression of the nuclear transcripts β and γ . The expression of β and γ transcripts in the c.43A>G carrier and the c.[1147delC];[43A>G] compound heterozygote is significantly lower ($p < 0.01$) than in controls and c.1147delC carrier. The comparisons were performed using t-test. Bars indicate the standard errors of the means.



Suppl. Figure 3. Sequences of the different *MUTYH* transcripts in a c.43A>G mutation carrier (RNA extracted from cultured lymphocytes) show the relatively weak reduction of RNA expression of the mutated allele in the nuclear transcripts (β and γ) compared to its expression in the α transcript.

ARTÍCULO 5**Germline mutations in *GEN1X* cause hereditary colorectal cancer through impaired DNA repair****Nuria Seguí et al.**

Artículo en preparación.

Resumen:

Parte de los factores heredables de susceptibilidad al CCR permanecen desconocidos. Mediante secuenciación de exomas completos de una familia clasificada como CCRf-X, hemos identificado un gen, *GEN1X*, como posible causa de la predisposición al cáncer. Tras el estudio del gen candidato en numerosos casos de CCR familiar, 176 familias Amsterdam y 71 familias Bethesda, sin defectos en la maquinaria de reparación de bases desapareadas, hemos detectado un total de dos mutaciones truncantes y tres mutaciones de cambio de sentido, todas ellas en familias que cumplen los criterios de Ámsterdam. Este gen participa en la reparación de uniones cruzadas entre las dos hebras de ADN. La haploinsuficiencia de *GEN1X* provoca sensibilidad a la mitomicina C (MMC) a concentraciones elevadas, y también parece causar la acumulación de un determinado tipo de errores en los tumores asociados debido a defectos en la reparación del ADN. Si se demuestra la patogenicidad de las mutaciones descritas, estaremos ante un nuevo gen de predisposición al CCR que podría explicar cerca del 3% de los casos Ámsterdam positivos con etiología genética desconocida. Estos hallazgos vinculan la predisposición al CCR con la vía de reparación del ADN de la anemia de Fanconi, fortaleciendo la relación entre el mantenimiento de la integridad genómica y el riesgo de cáncer.

Observación: Dado que los resultados de este estudio aún no han sido publicados se ha realizado la codificación de la nomenclatura del gen. Así, en la presente tesis, nos referimos a él como *GEN1X*.

Germline mutations in *GENIX* cause hereditary colorectal cancer through impaired DNA repair.

The identification of new genes associated with hereditary cancer will facilitate the management of patients whose predisposition is yet unexplained. Initially identified by exome sequencing, we implicate *GENIX* mutations in the etiology of hereditary colorectal cancer, accounting for ~3% of the genetically uncharacterized Amsterdam-positive mismatch repair-proficient families. These findings link colorectal cancer predisposition to the Fanconi Anemia DNA repair pathway, further strengthening the connection between maintenance of genome integrity and cancer risk.

Brief communication

Familial aggregation of colorectal cancer (CRC) is one of the strongest risk factors for CRC. Germline mutations in the DNA mismatch repair (MMR) genes, *EPCAM*, *APC*, *MUTYH*, *POLE*, *POLD1*, *GREM1*, *SMAD4*, *BMPRIA*, *STK11* and *PTEN* cause hereditary forms of CRC¹⁻³. However, much of the observed heritability and familial aggregation of the disease has not been explained.

With the aim of identifying novel hereditary CRC genes, we sequenced the exomes of three cancer-affected members of a high-risk, Amsterdam I MMR-proficient, CRC family (Fig. 1, Family 1). We identified a total of 32 unreported or rare (MAF<1%) germline heterozygous variants shared by all affected relatives (Supplementary Table 1), including a nonsense mutation in a gene involved in DNA interstrand cross-links repair, *GENIX* c.141C>A (p.C47*). The identified *GENIX* mutation is not currently described by the NHLBI Exome Sequencing Project (ESP)⁵ or 1000 Genomes, and has not been found in 1648 alleles of Spanish origin, including 286 sporadic CRC patients.

To determine whether mutations in the candidate gene might be involved in the pathogenesis of hereditary CRC, *GENIX* was sequenced in 176 MMR-proficient Amsterdam I and II families, and in 71 MMR-proficient Bethesda CRC families (Supplementary Table 2). Four additional unreported or rare genetic variants were identified in Amsterdam-positive families: a truncating mutation, c.2854C>T (p.R952*),

and three missense variants, c.418G>T (p.D140Y), c.1018C>T (p.P340S) and c.1771C>T (p.R591W) (Fig. 1). No *GENIX* mutations, including nonsense, frameshift, splice-site or missense variants, were identified in the normal colonic mucosae of 42 sporadic CRC patients from Spain and of 100 TCGA CRC patients⁶, and in 250 Spanish individuals without CRC⁷. This was not the case for the 6503 individuals included in ESP⁴ where a total of 10 nonsense, frameshift or splice-site mutations with MAF<5% were identified in 16 individuals (0.24%). Unfortunately, no information about personal or family history of cancer is available. Mutation carrier status could be assessed in 14 family members: all 11 cancer-affected (10 CRC and 1 breast cancer) were carriers, two unaffected 21 and 43 year-old individuals were carriers and one unaffected 53 year-old was non-carrier (Fig. 1).

The identification of two truncating mutations in *GENIX* prompted us to investigate whether the other three variants might also affect the protein function (Table 1). *In silico* algorithms predicted damaging effects for c.1771C>T (p.R591W), being the R591 residue evolutionary conserved during evolution (Supplementary Fig. 1a). Three-dimensional (3D) structure predictions of the region indicated that R591 is located in an exposed loop that connects two α -helices, in the vicinity of a DNA-binding domain (Supplementary Fig. 1b), predicting destabilization of the structure for R591W. While *in silico* predictors are not conclusive for c.418G>T and c.1018C>T (Table 1), a splicing predictor based on ESRseq scores⁸, indicated the creation of an exonic splicing silencer for c.418G>T (Supplementary Table 3). Therefore, most germline *GENIX* mutations identified are predicted to have a functional effect.

Five colorectal tumors developed by *GENIX* mutation carriers (three c.141C>A and two c.418G>T) were available for somatic testing. Whole-exome sequencing of the Family 1 proband's tumor identified a total of 236 somatic mutations in transcribed sequences, including mutations in classic driver genes for CRC, such as *APC*, *KRAS* or *PIK3CA* (Supplementary Table 4), with an average mutation rate of 5/Mb, or 1.3/Mb for non-synonymous changes. This mutation burden corresponds to that of non-hypermutant CRCs⁶. However, the mutation spectrum is characterized by an excess of T:A>G:C (10.5%) and C:G>G:C transversions (12.5%), both exceeding the 95th percentiles observed in non-hypermutant TCGA CRCs (Supplementary Fig. 2). On the other hand, no clear evidence of somatic *GENIX* second hits was obtained. No loss of heterozygosity (LOH)

(0/5), or somatic mutation (0/3) was evidenced (Supplementary Fig. 3). Furthermore, neither loss of RNA expression of the wildtype allele nor reduction of expression of the *GENIX* protein were observed in a tumor developed by a c.141C>A carrier (Family 1, proband). However, *GENIX* protein levels of normal colon mucosa from the *GENIX* c.141C>A carrier were lower than those of a wildtype individual, supporting the predicted functional effect of c.141C>A (p.C47*) (Supplementary Fig. 4). Altogether, our results point to a *GENIX* haploinsufficiency that might cause a bias towards a specific type of errors due to defective DNA maintenance. Study of additional *GENIX*-mutated tumors is required to validate this hypothesis.

We tested the survival in response to MMC of the lymphoblastoid cell lines derived from: a c.141C>A (p.C47*) heterozygous carrier (*GENIX*^{+/-}); two *GENIX*-deficient patients (*GENIX*^{-/-}); a *FANCA*-deficient patient; and a wildtype control. The results obtained indicate that at relatively high concentrations of MMC (10-70 nM), *GENIX*^{+/-} cells are more sensitive to MMC than wildtype cells but less sensitive than *GENIX*^{-/-} (Fig. 2). The observation that *GENIX*^{+/-} cells are hypersensitive to high doses of MMC is consistent with the haploinsufficiency model suggested above.

Our data suggests that *GENIX* mutations increase the risk to develop CRC. To test whether *GENIX* silencing also contributes to tumor growth, tumors derived from the subcutaneous implantation in athymic mice of Caco-2 cell lines with and without *GENIX* (sh-*GENIX* and sh-empty) were grafted in the cecum of athymic mice. Tumors derived from sh-*GENIX* Caco-2 were slightly larger than the control group (statistically non-significant) (Supplementary Fig. 5). In addition, two mice from the sh-*GENIX* group showed macroscopic tumor masses in the diaphragm (2/10), while no metastases were observed in the control group (0/9).

Our findings implicate *GENIX* mutations in the inherited susceptibility to CRC. The analysis of a larger series of cases will provide further information about the prevalence of *GENIX* mutations in hereditary CRC (2.8% of Amsterdam-positive MMR-proficient families in our series). Likewise, a thorough analysis of genetic and genomic alterations found in *GENIX*-associated tumors will clarify the underlying repair defects accumulated and therefore the mechanism of action of *GENIX* haploinsufficiency in colorectal

carcinogenesis. Our findings further support the relationship between defective DNA repair and cancer predisposition, being the first unequivocal evidence linking the FA pathway and CRC through *GENIX*.

REFERENCES

1. Kastrinos, F. & Syngal, S. *Cancer J* **17**, 405-15 (2011).
2. Palles, C. et al. *Nat Genet* **45**, 136-44 (2013).
3. Jaeger, E. et al. *Nat Genet* **44**, 699-703 (2012).
4. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. & Jiricny, J. *J Biol Chem* **282**, 2976-86 (2007).
5. NHLBI GO Exome Sequencing Project, <http://evs.gs.washington.edu/EVS/>. (Seattle, WA).
6. Cancer Genome Atlas Network. *Nature* **487**, 330-7 (2012).
7. Quesada, V. et al. *Nat Genet* **44**, 47-52 (2011).
8. Ke, S. et al. *Genome Res* **21**, 1360-74 (2011).

FIGURES

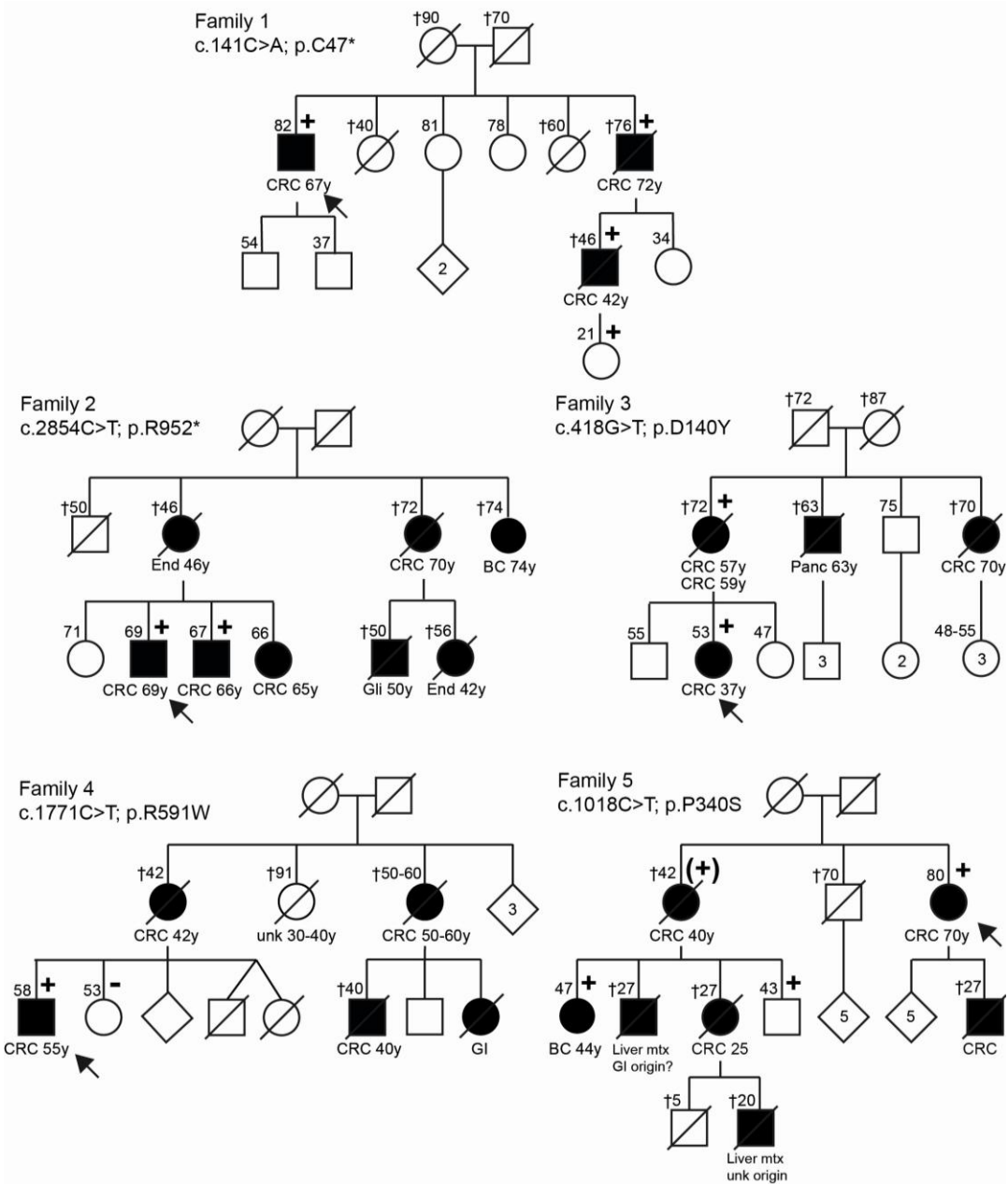


Figure 1. Pedigrees of the families with germline *GENIX* mutations. Filled symbol, cancer; +, mutation carrier; (+), obliged mutation carrier; -, wild-type; arrow, index case. Ages at information gathering or at death, when available, are indicated on the top-right corner of each individual's symbol. CRC, colorectal cancer; End, endometrial cancer; BC, breast cancer; Gli, glioma; Panc, pancreatic cancer; unk, cancer of unknown location; GI, gastrointestinal tumor; mtz, metastasis; y, years.

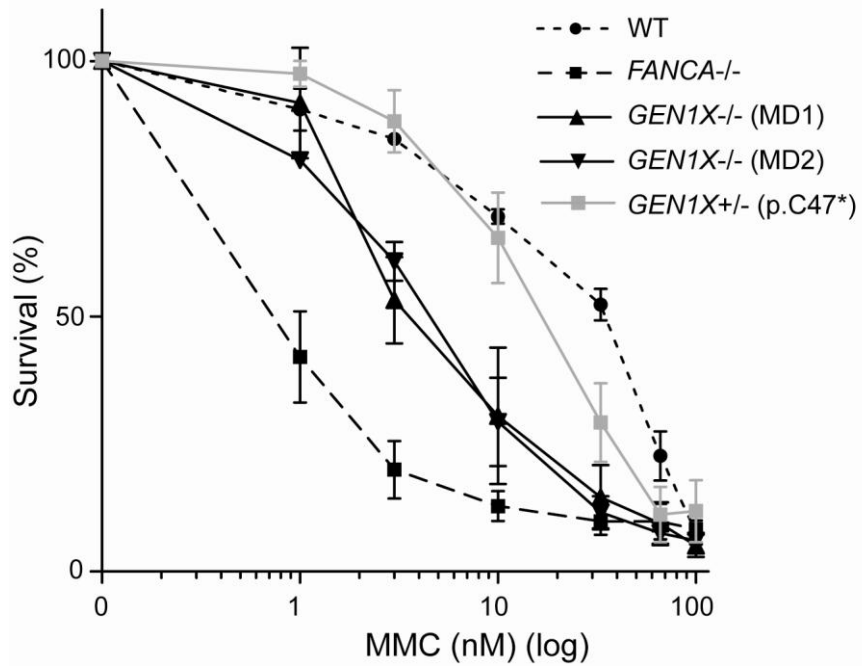


Figure 2. Sensitivity to MMC in lymphoblastoid cell lines derived from a FA patient with biallelic *FANCA* mutations, two *GEN1X*^{-/-} patients, a carrier of *GEN1X* c.141C>A (p.C47*) (*GEN1X*^{+/-}) and a wildtype individual. Error bars indicate SEM. At 30 nM of MMC, the cell survival observed in *GEN1X*^{+/-} cells is significantly lower than in wildtype cells (p=0.010, two-sided Wilcoxon rank sum test).

TABLES

Table 1. Germline *GENIX* mutations identified in 176 MMR-proficient Amsterdam-positive CRC families. In bold, the evidence that supports the damaging nature of the variants.

Family	<i>GENIX</i> genetic variant	Protein prediction (score)			Structure prediction	ESRseq score change	Population MAF (%) (dbSNP/ESP)
		PolyPhen-2 (HumDiv / HumVar)	SIFT	Condel			
1	c.141C>A (p.C47*)	-	-	-	Protein truncation	-	0/0
2	c.2854C>T (p.R952*)	-	-	-	Protein truncation	-	0.05/0
3	c.418G>T (p.D140Y)	Benign (0.03/0.019)	Damaging (0.04)	Deleterious (0,708)	N.I. ¹	-2.374	0/0
4	c.1771C>T (p.R591W)	Probably damaging (1/0.998)	Damaging (0)	Deleterious (1)	Protein destabilization (3D models)	-0.466	0/0.0154
5	c.1018C>T (p.P340S)	Benign (0.221/0.024)	Tolerated (0.09)	Neutral (0,056)	N.I. ¹	-0.030	0/0

1. D140 and P340 are located in a region predicted to have a disordered structure.

Abbreviations: MAF, minor allele frequency; ESP, NHLBI Exome Sequencing Project; N.I., not informative.

ONLINE METHODS

Whole exome capture and massive parallel sequencing

Library preparation for capturing selected DNA regions was performed according to Agilent's SureSelect protocol for Illumina paired-end sequencing (SureSelect XT Human All Exon V3, 50Mb, Agilent Technologies, Santa Clara, CA). In brief, 3.0 µg of genomic DNA were sheared on a Covaris™ E220 instrument. Fragment size (150-300bp) and quantity were checked with the Agilent 2100 Bioanalyzer 7500 chip. Fragmented DNA was end-repaired, adenylated and ligated to Agilent indexing-specific paired-end adaptors. The DNA with adaptor-modified ends was PCR amplified (6 cycles, Herculase II fusion DNA polymerase) with SureSelect Primer and SureSelect Pre-capture Reverse PCR primers, quality controlled on the DNA 7500 assay for the library size range of 250 to 450 bp, and hybridized for 24hrs at 65°C (Applied Biosystems 2720 Thermal Cycler, Life Technologies, Carlsbad, CA). The hybridization mix was washed in the presence of magnetic beads (Dynabeads MyOne Streptavidin T1, Life Technologies). The eluted fraction was PCR amplified (16 cycles) in order to add the index tags using SureSelectXT Indexes for Illumina. The final library size and concentration was determined on Agilent 2100 Bioanalyzer 7500 chip and sequenced on an Illumina HiSeq 2000 platform with paired end run of 2x76bp following the manufacturer's protocol. Images generated by Illumina HiSeq 2000 were processed using the manufacturer's software to generate FASTQ sequence files. Both exome capture and sequencing were performed at the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain).

Sequencing data analysis I

Reads were aligned to the reference genome (GRCh37) using Burrows-Wheeler analysis (BWA) with the sample option⁹ and a BAM file was generated using SAMtools¹⁰. PCR-duplicates were removed using SAMtools and custom scripts, and initial SNP calling was performed using a combination of SAMtools and Sidrón as previously described¹¹. Common variants, defined as those present in dbSNP135 with a minor allele frequency $\geq 1\%$, were filtered out. This data processing was carried out to search for germline mutations in Family 1 and to characterize the somatic mutations located in transcribed sequences present in the tumor of a *GENIX* mutation carrier (Family 1, proband).

Sequencing data analysis II

BAM files generated from the exome sequencing of 239 colorectal tumors and 100 normal mucosae were downloaded from TCGA website⁶. Variant calling was processed using GATK software (quality and mapping quality >30, and read depth >10). SNPs were removed using the pool of normal samples and the 1000Genomes database. All identified variants were analyzed, also including intronic and intergenic changes. The same pipeline was applied to tumor and normal (blood) tissues of a *GENIX* mutation carrier (Family 1, proband) to extract somatic genetic variants/mutations.

Mutation screening of *GENIX* c.141C>A

Screening for the identified *GENIX* mutation in controls and sporadic CRC population was performed by Conformation-Sensitive Capillary Electrophoresis (CSCE) (conditions available upon request). DNA samples from 288 controls and 286 consecutively recruited CRC patients from the same geographical area were analyzed¹². Patterns differing from the wildtype were analyzed by direct automated sequencing of the corresponding exonic fragment. Also, *GENIX* c.141C>A was searched in the data obtained from the exome sequencing of 250 Spanish individuals without CRC (subjects described in Quesada et al.⁷)

CRC families: validation series

GENIX was analyzed for new mutations in a total of 176 Caucasian CRC patients from unrelated families, recruited at different hospitals in Spain. All families fulfilled the Amsterdam I (56.8%) or II (43.2%) criteria and did not have MMR deficiency (either tumor microsatellite instability and/or lack of expression of the MMR proteins MLH1, MSH2, MSH6 or PMS2). The mean age at cancer diagnosis was 48 years for the sequenced patients (Supplementary Table 2).

Likewise, we studied the index CRC patients of 71 MMR-proficient CRC families characterized by the presence of two first-degree relatives affected with CRC or other related tumors, one of which diagnosed before the age of 50, or by the presence of three relatives with CRC or other related tumors independent of the age at diagnosis (Supplementary Table 2). All of them were recruited by the Genetic Counseling Unit of the Catalan Institute of Oncology. Informed consent was obtained from all subjects.

Direct automated sequencing

Exons and intron-exon boundaries of *GENIX* were sequenced using a standard protocol for automated direct (Sanger) sequencing. Sequencing was performed on an ABI Sequencer 3730 and data analyzed using Mutation Surveyor v.3.10. Sanger sequencing was used to validate the results obtained by massive parallel sequencing (exome and gene targeted sequencing), and to sequence *GENIX* in FFPE samples and in blood DNA from 85 Amsterdam-positive and 71 Bethesda-positive index cases.

Mutation identification in pooled samples

The remaining of 91 Amsterdam-positive MMR-proficient index patients were screened for new mutations in *GENIX* by using a combination of pooled samples, PCR amplification and high-throughput sequencing, as previously described¹¹. Amplification was performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) and custom-designed primers.

In silico functional prediction of genetic variants

Protein damage prediction of missense genetic variants was performed by using the *in silico* algorithms PolyPhen-2, SIFT and Condel¹³⁻¹⁵. Possible alterations of the splice sites were evaluated using NNSplice0.9 and SplicePort^{16,17}.

The functional impact of c.418G>T, c.1018C>T and c.1771C>T was also assessed by using a recently developed bioinformatics approach based on the evaluation of all RNA hexamers as potential exonic splicing elements⁸. This approach has been proven to be very sensitive in predicting the effect of variants in exon skipping. A total ESRseq score change lower than -0.663 is highly suggestive of a variant that causes exon skipping, whereas an increase or a lower decrease in the ESRseq score change usually occur in variants that do not cause exon skipping¹⁸.

Structural modeling of GENIX

Human *GENIX* is a multi-domain protein whose three-dimensional (3D) structure has not been determined yet. Protein domain annotations of *GENIX* were retrieved from UniProt.

ModBase (<http://modbase.compbio.ucsf.edu/>), a database of comparative protein structure models, currently contains six 3D models for *GENIX* covering the amino acid region from

position 14 to 1002. The sequence identity between GEN1X and the proteins whose crystal structures were used as template, ranges from 7% to 20%, well below the threshold of 30% for reliable fold assignment. Interestingly, amino acid regions present in all these 3D models show different topologies, which supports their low reliabilities. To calculate a more reliable 3D model of GEN1X, specifically for the protein domains of interest in this study, we first predicted the regions with disordered structure by using DisMeta (<http://www.wenmr.eu/wenmr/dismeta-disorder-prediction-metaserver>), DRIP-PRED (<http://www.sbc.su.se/~maccallr/disorder/cgi-bin/submit.cgi>) and MetaDisorder (<http://iimcb.genesilico.pl/metadisorder/>). We identified two regions with disordered structure: residues 1-30 and 95-370. The missense mutations D140Y and P340S are located in the second structural-disordered region.

We next focused on the 3D modeling of the GEN1X region that accommodates the missense mutation R591W. We threaded the GEN1X sequence to the PDB database by using HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>) and I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>). The resulting 3D models were refined using ModRefiner (<http://zhanglab.ccmb.med.umich.edu/ModRefiner/>).

Loss of heterozygosity (LOH)

Two microsatellite markers mapping close to *GEN1X* and an intragenic microsatellite, were analyzed. Also, SNaPshot® (Life Technologies) targeting the mutation c.141C>A was used to assess LOH and to discriminate the wildtype and the mutated allele. LOH was scored if the intensity of any allele was reduced by $\geq 50\%$ relative to the other allele after taking account of the relative allelic intensities in paired constitutional DNA.

Allele specific expression assessment

Allele specific expression assessment in RNA samples was performed by SNaPshot® using the mutation as allelic marker.

Western blotting

Tissue samples were lysed in ice-cold buffer: 40 mM HEPES (pH 7.4), 120 mM NaCl, 1% (v/v) Triton X-100, 1 mM EDTA, 0.5 M NaF, 10 mM PMSF and with protease inhibitors (Roche, Roche Diagnostics GmbH, Germany). Lysates were incubated on ice 15 minutes,

and the supernatants were cleared at 13,000 rpm at 4°C. GEN1X protein expression level was examined by western blotting with anti-GEN1X antibody. Actin was used as loading control (Millipore MAB1501R).

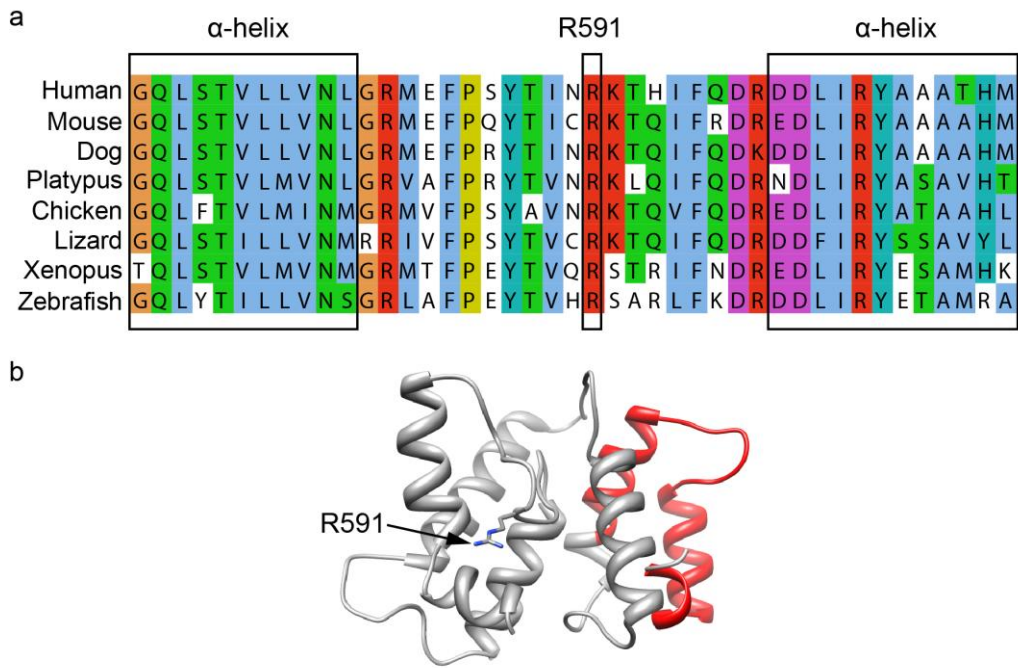
Survival assay in response to mitomycin C

Lymphocytes were cultured in RPMI 1640 medium (Invitrogen, Carlsbad, CA) supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin-streptomycin (Sigma, Saint Quentin Fallavier, France). They were exposed to increasing concentrations of MMC (0-100 nM) until they had reach at least three population doublings. Subsequently, cells for each MMC concentration were counted using a cell counter and size analyzer Beckman-Coulter Z2. Three experimental replicates in which each condition was performed in duplicate were carried out. *GEN1X*^{-/-} lymphoblastoid cell lines correspond to two patients with a homozygous deletion of GEN1X.

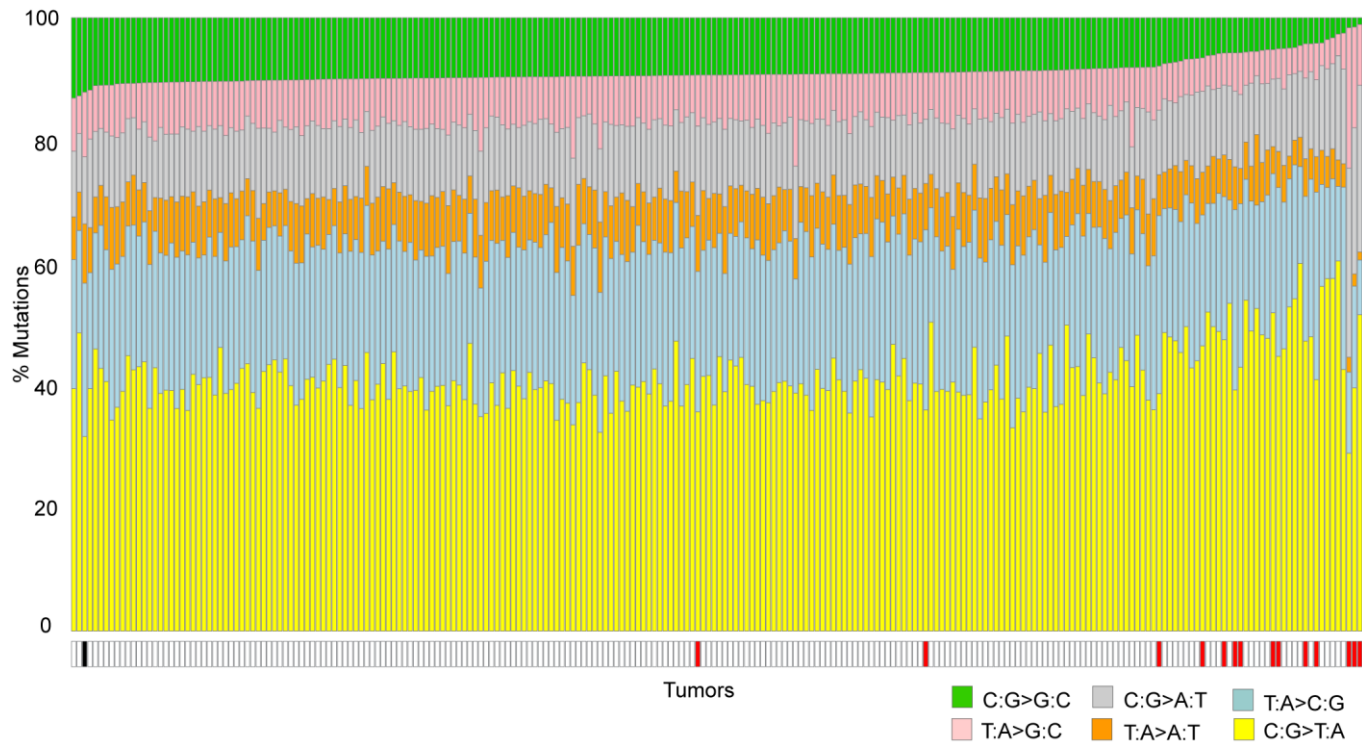
REFERENCES (Methods)

9. Li, H. & Durbin, R. *Bioinformatics* **25**, 1754-60 (2009).
10. Li, H. et al. *Bioinformatics* **25**, 2078-9 (2009).
11. Puente, X.S. et al. *Nature* **475**, 101-5 (2011).
12. Moreno, V. et al. *Clin Cancer Res* **12**, 2101-8 (2006).
13. Adzhubei, I.A. et al. *Nat Methods* **7**, 248-9 (2010).
14. Kumar, P., Henikoff, S. & Ng, P.C. *Nat Protoc* **4**, 1073-81 (2009).
15. Gonzalez-Perez, A. & Lopez-Bigas, N. *Am J Hum Genet* **88**, 440-9 (2011).
16. Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. *J Comput Biol* **4**, 311-23 (1997).
17. Dogan, R.I., Getoor, L., Wilbur, W.J. & Mount, S.M. *BMC Bioinformatics* **8**, 410 (2007).
18. Di Giacomo, D. et al. *Hum Mutat* **34**, 1547-57 (2013).

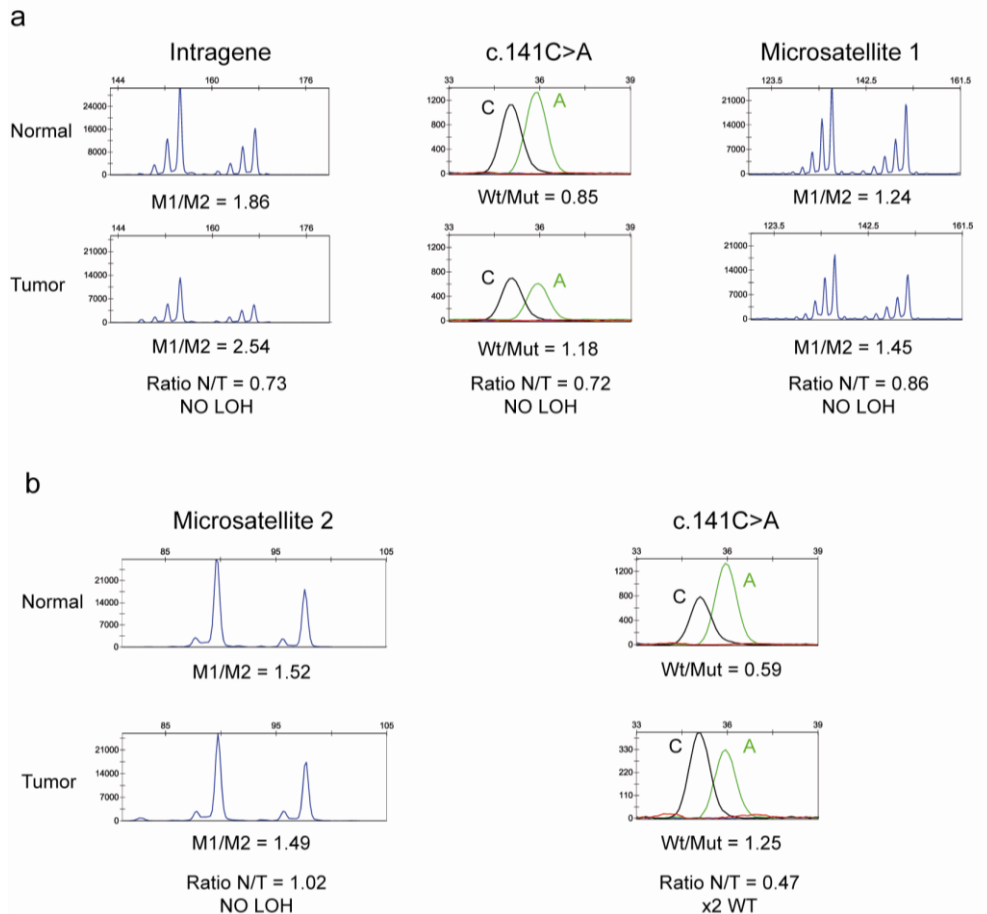
SUPPLEMENTARY FIGURES



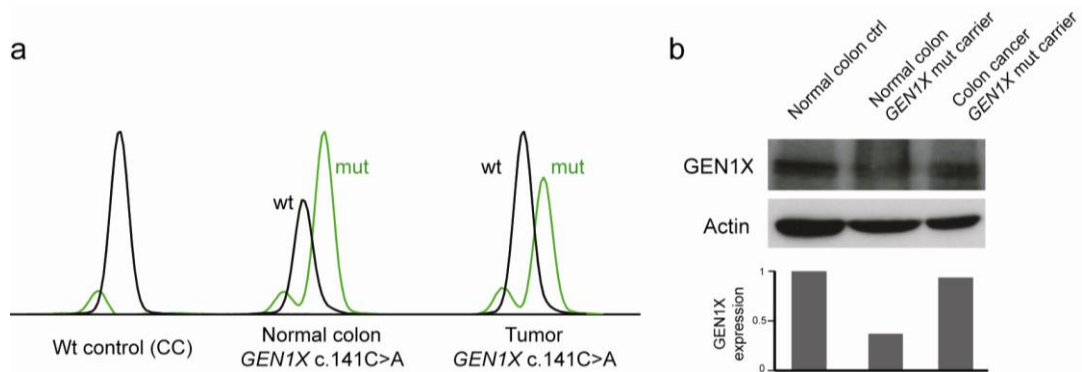
Supplementary Figure 1. Protein sequence and 3D structure of the region containing R591. a) Multiple sequence alignments for the region comprised between the α -helix motives flanking R591 in different species. b) 3D model for the GEN1X region containing R591.



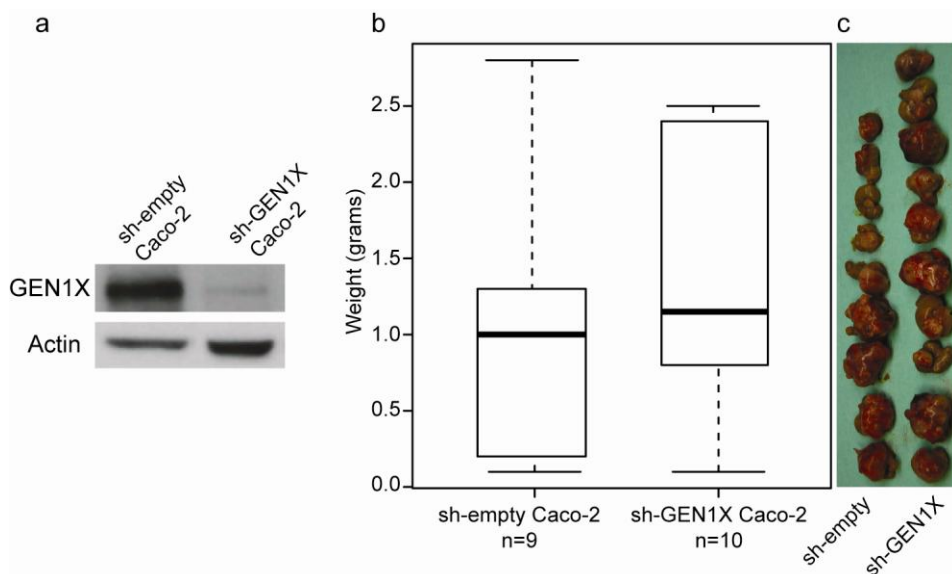
Supplementary Figure 2. Mutation spectrum of the colon tumor of a *GENIX* c.141C>A (p.C47*) mutation carrier (Family 1, proband) in the context of the mutation spectrum of 239 TCGA colorectal tumors. The lower panel depicts the tumor developed by the p.C47* mutation carrier in black, the 15 most hypermutated TCGA colorectal tumors in red, and the remaining 224 TCGA CRCs uncolored. Exome-sequencing data from the proband's tumor and the 239 TCGA CRCs were processed applying the same analytical method (Online Methods, Sequence data analysis II).



Supplementary Figure 3. Absence of LOH of the wildtype allele at the *GENIX* locus in the colon tumors developed by *GENIX* c.141C>A (p.C47*) mutation carriers: a) Family 1, individual IL7; b) Family 1, individual III.6. The LOH was studied by using three microsatellites, microsatellite 1, microsatellite 2 and an intragenic microsatellite, as well as the mutation assessed by SNaPshot. Also, no LOH, assessed by SNaPshot and microsatellite marker analysis, was detected in two colorectal tumors developed by *GENIX* c.418G>T (p.D140Y) mutation carriers (Family 3, proband and individual II.1).



Supplementary Figure 4. *GEN1X* expression in the tumor developed by a *GEN1X* c.141C>A (p.C47*) mutation carrier (Family 1, proband). A) Allele specific expression of *GEN1X* in normal and tumor assessed by SNaPshot and using the mutation as allelic marker. B) Expression of the *GEN1X* protein in normal and tumor tissue of the patient assessed by Western blot and relative quantification (*GEN1X*/actin).



Supplementary. Figure 5. Tumors developed after the orthotopic implantation in the cecum of athymic mice of tumors previously originated after subcutaneous implantation in athymic mice of Caco-2 cell lines with (sh-*GEN1X*) and without (sh-empty) silencing of *GEN1X*. a) Western blot showing the silencing of *GEN1X* in the Caco-2 cell line. b) Box blot indicating the distribution of tumor weights at sacrifice. c) Size of the tumors at sacrifice.

SUPPLEMENTARY TABLES

Suppl. Table 1. Rare or unreported genetic variants identified by exome sequencing in all the studied cancer-affected family members (Family 1).

Gene	Nucleotide change	Aminoacid change	Location	dbSNP (MAF)	Function prediction (score)		
					¹ PolyPhen-2 (HumDiv / HumVar)	² SIFT	Condel
AL359195.1	c.227G>A	p.Cys76Tyr	Exon 1		0	0,14	Neutral (0,014)
ATP13A3	c.1057A>G	p.Thr353Ala	Exon 12		0/0,005	0,09	Neutral (0,019)
BABAM1	c.686C>T	p.Thr229Met	Exon 7		0,394/0,055	n.a.	Deleterious (0,949)
BRCA1	c.4534C>G	p.Pro1512Ala	Exon 14	rs111034213 (n.a.)	0,404/0,141	n.a.	n.a.
C2orf16	c.2648G>A	p.Arg883Gln	Exon 1	rs369073751 (n.a.)	n.a.	n.a.	n.a.
CTTNBP2NL	c.832G>A	p.Glu278Lys	Exon 6		0,181/0,033	1	Neutral (0)
DACT2	c.791C>G	p.Ser264*	Exon 5		-	-	-
DARC	c.768G>A	p.Trp256*	Exon 1		-	-	-
DBH	c.1835T>C	p.Ile612Thr	Exon 12		0,002/0,003	0	Neutral (0)
EPHA8	c.1093G>A	p.Val365Met	Exon 5		0,045/0,010	0,21	Neutral (0,09)
FAM55B	c.1587G>A	p.Trp529*	Exon 6	rs201200145 (n.a.)	-	-	-
GEN1X	c.141C>A	p.Cys47*	Exon 2		-	-	-
FST	c.691A>G	p.Ile231Val	Exon 4		0,889/0,618	0	Deleterious (0,905)
HABP2	c.1598G>T	p.Cys533Phe	Exon 13		1/1	0	Deleterious (0,990)
HOXB2	c.1012G>A	p.Gly338Lys	Exon 2	rs369806240(n.a.)	n.a.	n.a.	n.a.
HRASLS2	c.377G>A	p.Arg126His	Exon 3		0,227/0,055	0,15	Neutral (0,437)
IFNA10	c.364T>C	p.Cys122Arg	Exon 1		1/1	0	Deleterious (0,990)
LDHA	c.92C>T	p.Ala31Val	Exon 2		0,194/0,169	0	Neutral (0,012)

LRRC14B	c.1249G>A	p.Glu417Lys	Exon 2		0,001/0,001	0,41	Neutral (0,003)
MCC	c.587G>A	p.Gly196Asp	Exon 3		n.a.	0,16	Neutral (0,012)
MYO1H	c.1886G>A	p.Arg629Gln	Exon 18		n.a.	0	Deleterious (1)
OR6T1	c.928C>T	p.Leu310Phe	Exon 1	rs201332040 (n.a.)	0.021/0.011	0.44	Neutral (0.003)
PCLO	c.14337G>A	p.Met4779Ile	Exon 17		0/0	n.a.	Neutral (0)
PKP3	c. 2017C>T	p.Arg673Cys	Exon 10	rs200807624 (0.001)	1/0.996	0	Deleterious (1)
PTPN20C	c.1004A>G	No protein product	Exon 6		-	-	-
RABL5	c.286A>G	p.Arg96Trp	Exon 4		0.002/0.001	n.a.	n.a.
RAPH1	c.2432C>T	p.Ala811Val	Exon 15		0/0	0.57	Neutral (0.025)
RPL6	c.668G>A	p.Arg223Gln	Exon 6		0.007/0.009	0	Neutral (0.001)
SNX6	c.238C>T	p.Arg80Trp	Exon 4		n.a.	n.a.	n.a.
SPECC1	c.1595T>A	p.Met532Lys	Exon 4		0.002/0.001	0.91	Neutral (0)
TAF4	c.1501G>A	p.Val501Ile	Exon 2		0.959/0.228	0.03	Deleterious (0.786)
WDR5B	c.167_170del TAGC	p.Leu56GInfs*8	Exon 1		-	-	-

1. PolyPhen-2: score > 0.85, probably damaging; score= 0.15-0.85, possibly damaging; score ≤ 0.15, benign.

2. SIFT: score ≤ 0.05, damaging; score > 0.05, tolerated.

Abbreviations: MAF, population minor allele frequency; n.a., not available information.

Suppl. Table 2. Characteristics of the CRC families and patients studied.

Geographic origin (Spain)	Criteria n(%)			¹ Age at cancer diagnosis Mean (\pm SD)	² Tumor MMR status n (%)		
	Ams I	Ams II	Beth		MSS	IHQ+	Total
Amsterdam-positive							
³ Catalonia (n=43)	32 (74.4%)	11 (25.6%)	-	45.3 (\pm 9.2)	39 (90.7%)	24 (55.8%)	43 (100%)
Madrid (n=66)	47 (71.2%)	19 (28.8%)	-	48.6 (\pm 11.5)	61 (92.4%)	48 (72.7%)	66 (100%)
Valencia (n=33)	5 (15.2%)	28 (84.8%)	-	50.8 (\pm 11.0)	33 (100%)	32 (97%)	33 (100%)
Castilla-Leon (n=27)	10 (37%)	17 (63%)	-	50.3 (\pm 13.7)	27 (100%)	27 (100%)	27 (100%)
Asturias (n=7)	6 (85.7%)	1 (14.3%)	-	46.7 (\pm 5.5)	4 (57.1%)	6 (85.7%)	7 (100%)
⁴Bethesda-positive							
Catalonia (n=71)	-	-	71 (100%)	57.8 (\pm 12.7)	71 (100%)	n.a.	71 (100%)
TOTAL (n=247)	100 (40.5%)	76 (30.8%)	71 (28.7%)	50.9 (\pm12.06)	235 (95.1%)	137 (55.5%)	247 (100%)

1. Age at cancer diagnosis of the individual where *GEN1X* was sequenced

2. Assessed by testing microsatellite instability and/or immunohistochemistry of the MMR proteins.

3. Family 1 belongs to the Amsterdam I-positive families of Catalan origin.

4. Characterized by the presence of two first-degree relatives affected with CRC or other related tumors, one of which diagnosed before the age of 50, or by the presence of three relatives with CRC or other related tumors independent of the age at diagnosis.

Abbreviations: Ams, Amsterdam criteria; Beth, Bethesda criteria; MMR, DNA mismatch repair; MSS, microsatellite stability; IHQ+, positive expression of the MMR proteins MLH1, MSH2, MSH6 and PMS2; SD, standard deviation.

Suppl. Table 3. Calculation of the ESRseq score change for the *GENIX* missense variants c.418G>T, c.1018C>T and c.1771C>T. The ESRseq scores of each hexamer (ESEseq or ESSseq) were retrieved from Ke et al. (2011). The total ESRseq scores and the ESRseq score changes were calculated as previously described (Ke et al. 2011; Di Giacomo et al. 2013).

Wildtype (WT)			Variant (VAR)			ESRseq score change (VAR-WT)
Overlapping hexamers	ESEseq or ESSseq score	Total ESRseq score	Overlapping hexamers	ESEseq or ESSseq score	Total ESRseq score	
c.418 G>T						-2.374 (potential ESS)
ATCAAG	0 (N)	1.567	ATCAAT	0 (N)	-0.807	
TCAAGA	0.464 (E)		TCAATA	0 (N)		
CAAGAT	0.638 (E)		CAATAT	-0.268 (S)		
AAGATG	0.465 (E)		AATATG	-0.245 (S)		
AGATGA	0 (N)		ATATGA	0 (N)		
GATGAG	0 (N)		IATGAG	-0.294 (S)		
c.1018C>T						-0.030
GGCTCC	0 (N)	0.469	GGCTCI	-0.222 (S)	0.439	
GCTCCT	0 (N)		GCTCIT	-0.257 (S)		
CTCCTC	0 (N)		CTCITC	0 (N)		
TCTCT	0 (N)		TCITCT	0.323 (E)		
CCTCTG	0.206 (E)		CITCTG	0.309 (E)		
CTCTGC	0.263 (E)		ITCTGC	0.286 (E)		
c.1771C>T						-0.466
TCAATC	0 (N)	1.352	TCAATI	0 (N)	0.886	
CAATCG	0.297 (E)		CAATIG	0 (N)		
AATCGG	0.266 (E)		AATIGG	0 (N)		
ATCGGA	0.342 (E)		ATIGGA	0.347 (E)		
TGGAA	0.447 (E)		TIGGAA	0.206 (E)		
CGGAAA	0 (N)		IGGAAA	0.333 (E)		

Abbreviations: ESE, exonic splicing enhancer; ESS, exonic splicing silencer; E, positive ESRseq score; S, negative ESRseq score; N, null ESRseq score.

Suppl. Table 4. Somatic mutations indentified in the tumor developed by a GEN1X c.141C>A (p.C47*) mutation carrier (Family 1, proband) (Online Methods, Sequencing data analysis I).

Gene	Chr.	Position	Wildtype (WT)	Variant	Exon	Type	Residue	WT residue	Mut residue	Conservation	dbSNP	Biotype
ABCA3	16	2339346	C	Y	-	intron	-	-	-	-0,642	-	protein-coding
ABHD1	2	27353355	G	R	-	intron	-	-	-	-0,16	-	protein-coding
AC016708.2	2	215712152	C	Y	1	3pUTR	-	-	-	0,736	-	processed pseudogene
AC097523.2	2	139046157	T	Y	1	3pUTR	-	-	-	0,489	-	processed pseudogene
AC104809.3	2	241906921	G	R	-	intron	-	-	-	-0,433	-	protein-coding
ADAD2	16	84225005	G	R	1	non_syn	'57/666	G	R/G	-0,048	-	protein-coding
ADAM19	5	156896069	C	Y	-	intron	-	-	-	-0,357	-	protein-coding
ADAMTS16	5	5242189	G	R	17	syn	'849/1225	P	P/P	-2,633	-	protein-coding
ADAMTS4	1	161168450	C	Y	1	5pUTR	-	-	-	2,122	-	protein-coding
ADCY5	3	123010055	G	R	18	non_syn	'1078/1262	A	T/A	0,959	-	protein-coding
ADH1A	4	100197996	A	M	-	intron	-	-	-	-0,34	-	protein-coding
AGRN	1	978964	C	Y	9	syn	'550/2046	T	T/T	-1,578	-	protein-coding
AHCYP2	9	123483491	G	R	1	3pUTR	-	-	-	1,584	-	processed pseudogene
ANKRD44	2	197860045	T	K	-	intron	-	-	-	-0,132	-	protein-coding
AP003900.6	21	11181417	G	S	-	intron	-	-	-	-0,179	rs915519	lincRNA
AP3B1	5	77511994	A	W	7	non_syn	'175/1046	K	K/I	1,112	-	protein-coding
APC	5	112173917	C	Y	16	non_syn	'876/2844	R	R/*	2,324	-	protein-coding
ASH1L	1	155340563	C	Y	-	intron	-	-	-	-0,127	-	protein-coding
ATL1	14	51096610	G	K	-	intron	-	-	-	-0,292	rs61985479	protein-coding
ATP12A	13	25281222	A	W	16	non_syn	'744/1040	D	D/V	4,706	-	protein-coding
ATP13A5	3	193090100	A	R	-	intron	-	-	-	0,603	rs7618476	protein-coding
BAI3	6	69758214	G	R	14	non_syn	'749/1523	V	M/V	1,231	-	protein-coding
BCAP31	X	152989105	G	R	1	syn	'5/314	A	A/A	-1,367	-	protein-coding
C10orf68	10	33134724	T	Y	-	intron	-	-	-	-0,783	rs181166524	protein-coding
C14orf164	14	23732311	T	Y	-	intron	-	-	-	0,881	-	protein-coding
			CTGCTGCTGC	CTGCTGC								
C16orf74	16	85743878	TG	TG	3	del_1res	'22/77	H	-	40811	-	protein-coding
C1orf222	1	1896465	C	Y	2	syn	'64/173	G	G/G	-0,717	-	protein-coding
C4orf40	4	71024675	A	R	-	intron	-	-	-	-0,05	-	protein-coding
C7orf10	7	40789122	T	W	-	intron	-	-	-	1,049	-	protein-coding
C8orf34	8	69243327	G	R	1	non_syn	'27/539	R	H/R	0,003	-	protein-coding
CABP4	11	67223848	G	R	3	non_syn	'159/276	R	Q/R	4,894	-	protein-coding
CACNA1H	16	1260096	G	R	18	non_syn	'1269/2348	R	H/R	1,189	-	protein-coding
CACNG6	19	54501442	C	Y	-	intron	-	-	-	0,394	-	protein-coding
CCDC147	10	106139938	G	R	9	non_syn	'442/873	R	H/R	5,708	-	protein-coding
CCDC15	11	124908956	A	M	15	non_syn	'920/963	N	N/H	3,059	-	protein-coding
CCDC178	18	30825115	A	W	-	intron	-	-	-	1,811	-	protein-coding
CCDC40	17	78058455	C	Y	-	intron	-	-	-	-1,296	-	protein-coding
CCT2	12	69992192	G	R	14	non_syn	'476/536	A	T/A	1,994	-	protein-coding
CDH16	16	66943281	G	R	17	non_syn	'761/830	V	M/V	3,93	-	protein-coding
CELF4	18	34901930	C	Y	-	intron	-	-	-	-0,707	-	protein-coding
CHD3	17	7812680	C	Y	-	intron	-	-	-	1,279	-	protein-coding
CLK2	1	155235612	C	Y	-	intron	-	-	-	0,932	-	protein-coding

Gene	Chr.	Position	Wildtype (WT)	Variant	Exon	Type	Residue	WT residue	Mut residue	Conservation	dbSNP	Biotype
COL12A1	6	75904564	C	Y	3	non_syn	'58/3063	T	T/M	2,71	rs140583215	protein-coding
CPA3	3	148603749	A	R	-	intron	-	-	-	-2,195	-	protein-coding
CPSF1P1	22	32666442	C	Y	1	3pUTR	-	-	-	-1,3	-	processed pseudogene
CRISPLD1	8	75929458	G	K	-	intron	-	-	-	-0,042	-	protein-coding
CSMD1	8	3046456	T	K	35	non_syn	'1826/3565	L	V/L	0,22	-	protein-coding
CST2	20	23807017	G	R	-	intron	-	-	-	0,301	-	protein-coding
CTD-2057J6.1	5	29882786	C	Y	1	3pUTR	-	-	-	0,234	-	processed pseudogene
CTD-3233P19.7	19	21416580	A	W	-	intron	-	-	-	-0,012	-	unprocessed pseudogene
DAB1	1	57480729	C	Y	13	non_syn	'457/589	P	P/L	4,103	-	protein-coding
DAOA	13	106119259	G	R	-	intron	-	-	-	-0,732	rs72549470	protein-coding
DAPP1	4	100761572	C	Y	3	syn	'117/281	S	S/S	-0,685	-	protein-coding
DCP1A	3	53376092	C	Y	-	intron	-	-	-	-0,836	-	protein-coding
DHX32	10	127576297	C	Y	-	intron	-	-	-	-2,306	-	protein-coding
DNAH5	5	13700780	G	K	78	non_syn	'4564/4625	M	M/I	3,287	-	protein-coding
DOCK11	X	117817223	T	W	-	intron	-	-	-	-0,045	-	protein-coding
DPP10	2	115919591	C	Y	1	non_syn	'20/801	Q	Q/*	1,518	-	protein-coding
DUSP27	1	167097292	C	S	6	non_syn	'975/1159	S	C/S	0,877	-	protein-coding
DYSF	2	71795104	C	Y	25	syn	'863/2120	G	G/G	0,266	-	protein-coding
EBF3	10	131641421	G	R	12	non_syn	'392/597	G	R/G	6,323	-	protein-coding
EIF2AK4	15	40282425	A	R	-	intron	-	-	-	0,042	-	protein-coding
EIF4E1B	5	176070798	C	Y	-	intron	-	-	-	-0,073	rs184522665	protein-coding
ELL2	5	95297271	C	M	-	intron	-	-	-	0,319	-	protein-coding
ELMOD3	2	85598100	G	K	-	intron	-	-	-	0,436	-	protein-coding
EMILIN1	2	27305403	C	M	4	syn	'322/1017	R	R/R	2,607	-	protein-coding
EPHA3	3	89528517	C	Y	-	intron	-	-	-	-1,321	-	protein-coding
ESPNP	1	17044695	A	R	-	intron	-	-	-	0,49	rs2773186	unprocessed pseudogene
EVC2	4	5576398	C	Y	-	intron	-	-	-	-0,902	-	protein-coding
EXOC3	5	457036	C	Y	5	non_syn	'360/746	P	P/L	5,611	-	protein-coding
EXOC3L2	19	45731456	C	Y	2	syn	'53/410	D	D/D	0,341	-	protein-coding
EYA3	1	28316267	A	M	14	non_syn	'407/537	E	E/D	1,28	-	protein-coding
EYS	6	66205262	T	Y	4	syn	'14/3166	F	F/F	0,126	-	protein-coding
FAM122B	X	133927883	C	Y	2	non_syn	'47/267	R	R/W	2,719	-	protein-coding
FAM127A	X	134166794	G	R	1	3pUTR	-	-	-	2,625	-	protein-coding
FAM216B	13	43358253	T	K	2	non_syn	'17/140	L	R/L	2,194	-	protein-coding
FBN2	5	127614381	G	K	63	non_syn	'2431/2913	A	A/S	0,965	-	protein-coding
FBXL6	8	145581945	CGGGG	CGGG	1	fs_del_1	'55/540	A	-	227	-	protein-coding
FBXW9	19	12799203	C	Y	-	intron	-	-	-	0,142	-	protein-coding
FES	15	91433735	G	S	-	intron	-	-	-	-0,606	-	protein-coding
FGF18	5	170883640	G	R	5	non_syn	'152/208	G	D/G	3,696	-	protein-coding
FKTN	9	108370021	C	Y	-	intron	-	-	-	-2,538	-	protein-coding
FRG1	4	190862307	G	S	-	intron	-	-	-	-0,647	-	protein-coding
FRG1	4	190862318	C	Y	-	intron	-	-	-	-0,478	rs111800301	protein-coding
FRG1	4	190862330	C	M	-	intron	-	-	-	-0,948	rs113358506	protein-coding
FRG1	4	190862350	C	S	-	intron	-	-	-	-0,008	-	protein-coding

Gene	Chr.	Position	Wildtype (WT)	Variant	Exon	Type	Residue	WT residue	Mut residue	Conservation	dbSNP	Biotype
GABRA5	15	27193051	G	R	-	intron	-	-	-	-0,832	-	protein-coding
GBP6	1	89849060	G	R	-	intron	-	-	-	0,315	-	protein-coding
GNAS	20	57485205	A	M	-	intron	-	-	-	-1,264	-	protein-coding
GPR143	X	9709582	C	S	-	intron	-	-	-	0,216	-	protein-coding
GPR64	X	19009101	G	R	29	non_syn	'979/1018	D	N/D	4,099	-	protein-coding
GRIA4	11	105789692	T	K	-	intron	-	-	-	1,545	-	protein-coding
GRIP2	3	14548229	A	R	-	intron	-	-	-	-1,076	-	processed-transcript
GSG2	17	3627704	G	R	1	3pUTR	-	-	-	0,212	-	protein-coding
HAAO	2	42994915	G	S	-	intron	-	-	-	-1,258	-	protein-coding
HBG2	11	5373458	T	Y	-	intron	-	-	-	1,337	-	protein-coding
HDC	15	50546973	C	S	-	intron	-	-	-	-0,31	rs79724974	protein-coding
HEATR2	7	821961	G	R	-	intron	-	-	-	-0,791	-	protein-coding
HIPK4	19	40889848	G	R	2	non_syn	'222/617	D	N/D	6,05	-	protein-coding
HRH3	20	60791478	G	R	3	non_syn	'308/454	G	S/G	0,544	-	protein-coding
ICOSLG	21	45649510	T	Y	6	non_syn	'442/474	L	P/L	-2,663	rs13048869	protein-coding
IFFO2	1	19243444	C	Y	5	syn	'352/518	V	V/V	-0,988	-	protein-coding
IQCB1	3	121544967	T	Y	5	syn	'108/466	L	L/L	1,351	-	protein-coding
IQCH	15	67601072	A	M	-	intron	-	-	-	-2,601	-	protein-coding
JPH3	16	87717757	C	Y	3	syn	'390/749	H	H/H	1,145	-	protein-coding
KCTD7	7	66105541	GTCTCTCTCT	GTCTCTCTCT	4	indel_3pU	-	-	-	11673	-	protein-coding
KIF2A	5	61681210	T	Y	-	intron	-	-	-	0,985	-	protein-coding
KIR2DS4	19	55358807	G	R	-	intron	-	-	-	-1,29	rs142705855	pseudogene
KLHL4	X	86887395	A	R	7	non_syn	'504/721	T	T/A	0,162	-	polymorphic
KRAS	12	25398284	G	R	2	non_syn	'12/190	G	D/G	5,862	-	protein-coding
KSR2	12	118198840	G	R	4	non_syn	'321/951	R	H/R	5,096	-	protein-coding
LATS2	13	21562611	G	R	4	syn	'436/1089	T	T/T	-2,498	-	protein-coding
LCE2A	1	152671619	G	R	2	non_syn	'81/107	R	Q/R	1,438	-	protein-coding
LILRB1	19	55112169	C	Y	-	intron	-	-	-	-0,353	-	protein-coding
LINC00587	9	105405500	A	M	-	intron	-	-	-	0,178	-	lincRNA
LPAR4	X	78011680	G	K	2	3pUTR	-	-	-	-0,098	-	protein-coding
LPHN1	19	14269142	C	S	-	intron	-	-	-	-0,903	-	protein-coding
LPHN2	1	82408925	G	K	6	non_syn	'224/1404	E	E/*	6,289	-	protein-coding
MAP2K7	19	7976202	TGGG	TGGGG	8	fs_ins_1	'308/420	L	-	7426	-	protein-coding
MFS12	19	3556883	A	M	-	intron	-	-	-	0,579	rs111812900	protein-coding
MROH2B	5	41007754	A	R	-	intron	-	-	-	0,594	-	protein-coding
MRPL46	15	89003080	C	M	3	3pUTR	-	-	-	4,225	-	nonsense mediated decay
MRPS9	2	105708958	G	S	8	non_syn	'251/397	E	E/Q	4,039	-	protein-coding
MSTO1	1	155582459	C	Y	-	intron	-	-	-	-0,106	-	protein-coding
MUC20	3	195456707	C	Y	-	intron	-	-	-	0,119	-	protein-coding
MUC3A	7	100547278	G	R	1	non_syn	'8/336	G	S/G	-3,219	rs73163737	protein-coding
MYBBP1A	17	4442641	A	W	-	intron	-	-	-	0,149	-	protein-coding
NCOR2	12	124887058	TGCTGCTGCT	TGCTGCTCTGCTGCTG	14	ins_1res	'511/2515	P	-	92740	-	protein-coding
NDUFB2	7	140396634	A	R	1	syn	'30/106	G	G/G	-0,283	-	protein-coding

Gene	Chr.	Position	Wildtype (WT)	Variant	Exon	Type	Residue	WT residue	Mut residue	Conservation	dbSNP	Biotype
NEBL	10	21097441	C	S	-	intron	-	-	-	1,963	-	protein-coding
NEDD9	6	11188717	T	Y	-	intron	-	-	-	-0,311	-	protein-coding
NEIL2	8	11640974	T	K	-	intron	-	-	-	-0,248	-	protein-coding
NELL1	11	20950075	T	K	-	intron	-	-	-	-0,141	-	protein-coding
NEMF	14	50292466	A	M	-	intron	-	-	-	0,032	-	protein-coding
NF2	22	30054300	A	M	-	intron	-	-	-	0,73	-	protein-coding
NIPAL4	5	156896069	C	Y	-	intron	-	-	-	-0,357	-	protein-coding
NKX3-1	8	23538651	C	Y	3	3pUTR	-	-	-	1,862	-	protein-coding
NOX4	11	89223726	A	W	-	intron	-	-	-	1,664	-	protein-coding
NPHP4	1	6008133	C	Y	8	non_syn	'330/1427	T	T/I	-0,137	-	protein-coding
NTM	11	132200071	C	M	8	non_syn	'320/356	P	H/P	1,556	-	protein-coding
OCRL	X	128718216	T	Y	-	intron	-	-	-	3,045	-	protein-coding
ODF2	9	131262535	G	R	21	3pUTR	-	-	-	1,079	-	protein-coding
OFD1	X	13773308	A	M	12	non_syn	'390/1013	N	N/H	3,069	-	protein-coding
OR11A1	6	29408571	G	R	-	intron	-	-	-	0,674	-	protein-coding
OR13C3	9	107299103	C	Y	1	3pUTR	-	-	-	-0,028	-	protein-coding
OR8H2	11	55872859	T	K	1	3pUTR	-	-	-	4,421	-	protein-coding
ORC5	7	103801812	T	W	-	intron	-	-	-	-0,007	-	protein-coding
P2RY4	X	69478805	C	M	1	3pUTR	-	-	-	5,143	-	protein-coding
PARP1P1	13	111591061	G	R	1	3pUTR	-	-	-	1,851	-	processed pseudogene
PCDHB17	5	140535663	C	Y	1	3pUTR	-	-	-	-0,514	-	protein-coding
PCSK2	20	17434761	C	Y	-	intron	-	-	-	0,874	-	protein-coding
												transcribed
												unprocessed
												pseudogene
PHF2P2	13	19526605	T	K	-	intron	-	-	-	-1,823	-	pseudogene
PHF8	X	54012317	C	Y	17	syn	'687/1025	G	G/G	-0,278	-	protein-coding
PI4K2A	10	99361771	C	Y	-	intron	-	-	-	-0,907	-	protein-coding
PIAS4	19	4028028	G	R	-	intron	-	-	-	0,096	-	processed transcript
PIK3CA	3	178936082	G	R	10	non_syn	'542/1069	E	K/E	5,802	-	protein-coding
PJA2	5	108704228	T	W	-	intron	-	-	-	0,145	-	protein-coding
PKD1L3	16	71981262	G	K	-	intron	-	-	-	0,149	-	processed transcript
PLA2G15	16	68289274	C	Y	4	non_syn	'165/413	R	R/*	1,939	-	protein-coding
PLEKHB2	2	132060632	T	Y	-	intron	-	-	-	-0,031	-	protein-coding
PLEKHB2	2	132110046	G	R	-	intron	-	-	-	-0,138	-	protein-coding
PLIN1	15	90216578	A	M	3	non_syn	'38/523	Q	Q/P	3,494	-	protein-coding
POLA2	11	65030072	G	R	-	intron	-	-	-	0,059	-	protein-coding
PRDM15	21	43226190	C	M	-	intron	-	-	-	-	-	protein-coding
PRKCB	16	24135382	G	R	-	intron	-	-	-	-0,192	-	protein-coding
PRKCQ	10	6533669	G	R	8	non_syn	'256/644	A	T/A	4,288	-	protein-coding
PRMT3	11	20473796	A	W	-	intron	-	-	-	0,358	-	protein-coding
PRMT5	14	23391327	C	Y	-	intron	-	-	-	-0,147	-	protein-coding
PRPF38A	1	52876975	T	Y	-	intron	-	-	-	-0,671	-	protein-coding
PRSS3	9	33797560	G	K	-	intron	-	-	-	-0,864	-	protein-coding
												nonsense mediated
												decay
PSTPIP1	15	77310392	G	R	-	intron	-	-	-	-1,117	-	decay
PTN	7	137028185	T	K	1	5pUTR	-	-	-	0,748	-	protein-coding
PXDN	2	1696179	G	R	-	intron	-	-	-	-1,68	-	protein-coding
RAX2	19	3771759	C	Y	-	intron	-	-	-	-3,352	-	protein-coding

Gene	Chr.	Position	Wildtype (WT)	Variant	Exon	Type	Residue	WT residue	Mut residue	Conservation	dbSNP	Biotype
RBM28	7	127983892	C	S	1	5pUTR	-	-	-	0,233	-	protein-coding
RFTN1	3	16475434	G	R	3	non_syn	'86/579	V	M/V	3,258	-	protein-coding
RNF17	13	25367255	C	Y	10	syn	'337/1582	Y	Y/Y	1,264	rs141654694	protein-coding
RP11-241F15.9	4	49563469	C	M	1	3pUTR	-	-	-	0,538	-	processed pseudogene
RP11-241F15.9	4	49563487	T	Y	1	3pUTR	-	-	-	2,573	-	processed pseudogene
488L18.3	1	247347806	G	R	1	3pUTR	-	-	-	0,725	-	processed pseudogene
RPUSD1	16	836504	C	M	-	intron	-	-	-	-0,005	-	protein-coding
SAP130	2	128707869	C	Y	17	non_syn	'872/1084	R	R/*	3,87	-	protein-coding
SCGB2A2	11	62038439	G	K	2	non_syn	'48/121	E	E/*	-1,237	-	protein-coding
SCUBE2	11	9077299	G	R	-	intron	-	-	-	-0,181	-	protein-coding
SCYL2	12	100732812	A	R	18	syn	'884/930	T	T/T	-0,023	-	protein-coding
SH3BGRL	X	80532663	C	Y	3	3pUTR	-	-	-	2,287	-	processed transcript
SHANK3	22	51160917	G	R	-	intron	-	-	-	-0,08	-	protein-coding
SIGLEC8	19	51960326	C	Y	-	intron	-	-	-	0,095	-	protein-coding
SLC22A1	6	160560740	G	R	7	non_syn	'373/555	G	R/G	1,504	-	protein-coding
SLC22A12	11	64360260	G	R	2	non_syn	'138/554	V	M/V	2,18	rs149722479	protein-coding
SLC22A14	3	38350410	G	R	-	intron	-	-	-	0,247	-	protein-coding
SLC2A4	17	7187286	C	Y	-	intron	-	-	-	0,734	-	protein-coding
SLC44A4	6	31836945	C	M	13	non_syn	'407/711	S	*S	1,942	-	protein-coding
SLC4A8	12	51845738	C	Y	-	intron	-	-	-	-0,104	-	protein-coding
SLC8A2	19	47960734	C	Y	3	non_syn	'265/922	R	R/C	3,049	-	protein-coding
SMARCA2	9	2110160	G	R	-	intron	-	-	-	0,094	-	protein-coding
SMCHD1	18	2769834	A	M	-	intron	-	-	-	-0,511	-	protein-coding
SMIM17	19	57166675	G	S	3	3pUTR	-	-	-	0,219	-	protein-coding
SNHG14	15	25444497	G	R	3	3pUTR	-	-	-	-1,489	-	antisense
SNRPN	15	25200221	C	Y	-	intron	-	-	-	-1,174	-	protein-coding
SNX18P24	4	49563496	A	M	1	3pUTR	-	-	-	2,581	-	processed pseudogene
SORD	15	45332772	T	Y	-	intron	-	-	-	-0,054	-	protein-coding
SOX17	8	55370662	C	Y	1	5pUTR	-	-	-	0,542	-	protein-coding
SOX2	3	181431130	G	K	1	3pUTR	-	-	-	1,657	-	protein-coding
SOX3	X	139586471	A	W	1	3pUTR	-	-	-	1,717	-	protein-coding
SP6	17	45925143	C	Y	2	non_syn	'218/377	S	S/L	4,935	-	protein-coding
SPAG17	1	118571012	T	K	4	3pUTR	-	-	-	1,754	-	processed transcript
SPAG17	1	118570984	C	S	4	3pUTR	-	-	-	3,694	-	processed transcript
SPAG9	17	49123976	C	Y	-	intron	-	-	-	5,78	-	protein-coding
SPTBN2	11	66461317	C	Y	22	non_syn	'1507/2391	R	R/W	1,247	-	protein-coding
SSTR4	20	23016556	G	R	1	3pUTR	-	-	-	0,026	-	protein-coding
SYCE1L	16	77241999	A	W	-	intron	-	-	-	-0,29	-	protein-coding
SYCP2	20	58471610	T	Y	-	intron	-	-	-	0,242	-	protein-coding
SYN2	3	12232194	G	R	10	3pUTR	-	-	-	1,086	-	retained intron
SYTL3	6	159103748	G	R	-	intron	-	-	-	-0,594	rs191533718	protein-coding
			ATTATTTATTTA	ATTATTTA		indel_3pU						
			TTTATTTATTTA	TTTATTTA								
TAGLN3	3	111732646	TTT	TTTATTT	5	TR	-	-	-	15135	-	protein-coding
TBC1D32	6	121481279	G	S	-	intron	-	-	-	-0,927	-	protein-coding

Gene	Chr.	Position	Wildtype (WT)	Variant	Exon	Type	Residue	WT residue	Mut residue	Conservation	dbSNP	Biotype
TMTC2	12	83424576	G	R	-	splice	-	-	-	4,886	-	protein-coding
TRHDE	12	72956792	A	W	3	3pUTR	-	-	-	1,147	-	processed transcript
TRPM8	2	234878540	A	W	-	intron	-	-	-	-1,399	-	protein-coding
TRPV4	12	110230485	C	Y	10	non_syn	'599/872	T	T/M	5,716	-	protein-coding
TUBG1	17	40761758	G	K	1	non_syn	'1/452	M	M/I	5,363	-	protein-coding
TUSC3	8	15600997	A	W	-	intron	-	-	-	0,341	-	protein-coding
TUSC5	17	1183553	G	R	1	syn	'86/178	A	A/A	-0,628	-	protein-coding
UBAP1	9	34241061	G	R	-	intron	-	-	-	-0,913	-	protein-coding
USP28	11	113675639	C	Y	20	non_syn	'844/1078	R	R/*	2,731	-	protein-coding
VSTM2A	7	54612108	C	Y	-	intron	-	-	-	0,559	-	protein-coding
WASF2	1	27739170	C	Y	7	syn	'240/499	N	N/N	-0,948	-	protein-coding
WDR49	3	167284786	C	M	-	intron	-	-	-	-1,2	-	protein-coding
WWC3	X	10062329	C	Y	-	intron	-	-	-	-0,425	-	protein-coding
XXbac-												
BPG308J9.3	6	29231797	C	Y	-	intron	-	-	-	0,053	-	lincRNA
ZFC3H1	12	72013714	A	R	26	non_syn	'1681/1990	I	I/V	1,272	-	protein-coding
ZFP92	X	152686619	C	Y	4	non_syn	'262/417	R	R/W	0,802	-	protein-coding
ZFR2	19	3820215	G	R	11	non_syn	'569/940	G	S/G	0,16	-	protein-coding
ZNF454	5	178392809	C	Y	5	syn	'468/523	Y	Y/Y	0,078	-	protein-coding
			GGCGGCGGC									
			GGCGGCGGC	GGCGGCG		indel_5pU						nonsense mediated
ZNF713	7	55955293	GGCG	GCG	1	TR	-	-	-	88	-	decay

DISCUSIÓN

El propósito de esta tesis es la caracterización genética del CCR hereditario no polipósico desde un punto de vista constitucional o de línea germinal. Para ello nos hemos centrado en diversos aspectos que abarcan el estudio de factores modificadores de riesgo a desarrollar cáncer, como es la longitud telomérica, el análisis un gen causal candidato, y por último la búsqueda de nuevos genes que puedan explicar la agregación de CCR en familias de alto riesgo mediante secuenciación de exomas completos. A continuación discutiré los resultados obtenidos en cada uno de estos niveles y su posible aportación a la práctica clínica, es decir, al diagnóstico y consejo genéticos en cáncer hereditario, y a la comprensión general de la biología del cáncer.

1. ANÁLISIS DE LA LONGITUD TELOMÉRICA EN EL CÁNCER COLORRECTAL HEREDITARIO NO POLIPÓSICO

1.1. Aspectos a considerar sobre la cuantificación de la longitud de los telómeros

Hemos estudiado la longitud telomérica en familias con SL y CCRf-X, así como en controles sanos pertenecientes a la misma población caucásica. Clásicamente la cuantificación de la longitud de los telómeros se ha realizado mediante *Southern Blot* o hibridación fluorescente in situ cuantitativa (Q-FISH). Sin embargo, en los últimos años se ha generalizado el uso de la técnica de PCR cuantitativa monocroma multiplex (Cawthon et al., 2009). Gracias al diseño de unos cebadores específicos, este método permite la amplificación simultánea en cada muestra de la región telomérica y de un gen de copia única (*albúmina* o *β -globina*) respecto al que se normaliza, utilizando un único fluoróforo. Con ello se consigue reducir considerablemente la variabilidad y el coste, tanto de tiempo como económico.

La longitud telomérica es un rasgo que en los últimos años se ha estudiado ampliamente en el contexto del cáncer. Se han asociado tanto telómeros cortos como telómeros largos con el riesgo incrementado a padecer distintos tipos de cánceres, pero existe una gran variabilidad entre los diferentes estudios, incluso cuando se estudia el mismo tipo de neoplasia (Prescott et al., 2011). Se ha sugerido que estas discrepancias entre estudios son el resultado de diferencias entre los métodos de cuantificación de la longitud telomérica, de extracción y almacenamiento de las muestras, así como del tipo de diseño experimental, dependiendo de si es prospectivo o retrospectivo (Wentzensen et al., 2011). La variabilidad entre ensayos oscila entre el 2 y 6% según la técnica utilizada para determinar la longitud telomérica (Aviv et al., 2011), aunque puede ser mejorada con un

control de calidad riguroso (Olsen et al., 2012). Además, se ha demostrado que el método de extracción de ADN utilizado afecta a la medición de la longitud de los telómeros (Cunningham et al., 2013), por lo que es importante mantener unas condiciones de extracción y almacenamiento idénticas para todas las muestras a comparar. Con el fin de intentar minimizar al máximo todos estos factores de variabilidad, en nuestro estudio solo se incluyeron muestras que hubiesen seguido procesos idénticos de extracción y almacenamiento del ADN. Asimismo, se intentó repartir una proporción equitativa de muestras de todos los grupos estudiados en cada placa de amplificación para compensar la variabilidad inter-experimental, utilizando en todas ellas el mismo ADN de referencia para realizar la recta patrón.

Los telómeros se acortan progresivamente en cada división celular debido a la incapacidad de la ADN polimerasa de completar la síntesis en los extremos cromosómicos. Sin embargo, la mitosis no es el único factor que afecta las pérdidas teloméricas. Se ha observado que el tabaco, la obesidad y otros factores que contribuyen al estrés oxidativo se asocian a telómeros cortos, mientras que un estilo de vida saludable se relaciona con longitudes teloméricas más estables (von Zglinicki, 2002; Cherkas et al., 2008; Kim et al., 2009). Es por ello que se habla de la longitud de los telómeros como un indicador de la edad biológica más que de la edad cronológica (Aviv, 2006). Dejando a un lado los aspectos técnicos de la cuantificación de la longitud telomérica, el hecho de que ésta sea una característica ampliamente influenciada por el ambiente supone un factor más que complica su análisis. En un estudio realizado en cáncer de colon se demuestra cómo la asociación entre deficiencias teloméricas y el riesgo de cáncer se ve trastocada por factores del estilo de vida tales como el índice de masa corporal o el tabaco (Pellatt et al., 2012).

En cuanto al diseño experimental, hay que tener en cuenta y hacer una clara distinción entre los estudios retrospectivos respecto a los prospectivos. Se ha observado que el número de asociaciones significativas en estudios de carácter retrospectivo es mucho mayor que las descritas en estudios prospectivos. Un claro ejemplo es el trabajo de Pooley y colaboradores (2010), en el que se observó retrospectivamente un incremento con alta significancia del riesgo asociado a telómeros cortos en cáncer de mama y de colon, mientras que la asociación era nula en el estudio prospectivo realizado por el mismo grupo. Se ha de tener en cuenta que los estudios retrospectivos son muy susceptibles a múltiples sesgos: Selección de individuos, supervivencia, causalidad inversa, etc. Valorando estudios prospectivos y retrospectivos se ha llegado a la

conclusión de que el acortamiento telomérico observado en pacientes con cáncer podría ser el resultado de efectos fisiopatológicos de la enfermedad y no un factor de riesgo a desarrollar cáncer, es decir, que sea una consecuencia y no una causa del cáncer (Prescott et al., 2011). Se ha especulado que entre los factores que podrían acelerar el acortamiento telomérico en los pacientes con cáncer se podría encontrar la enfermedad en sí, el tratamiento que haya podido recibir el paciente o el impacto psicológico que tanto el diagnóstico como el tratamiento pueden causar, (Beeharry y Broccoli, 2005; Biegler et al., 2012). Por ello lo ideal es realizar estudios prospectivos para valorar la participación de la longitud de los telómeros en la predisposición al cáncer. Sin embargo, no siempre esto es posible, como ocurre en los casos hereditarios, cuya prevalencia es relativamente baja y por lo tanto reunir prospectivamente un tamaño muestral suficiente requeriría un periodo de tiempo muy amplio. Por ello, se han establecido unos criterios para disminuir en lo posible los factores confusores externos que pueden sesgar los estudios retrospectivos, sobre todo los que tienen que ver con el origen, procesamiento y almacenamiento de las muestras (Prescott et al., 2011) y que nosotros hemos controlado en la medida de lo posible en nuestro estudio.

1.2. Longitud telomérica en el cáncer colorrectal hereditario no polipósico

Con el fin de investigar el papel de la longitud de los telómeros en el riesgo a desarrollar cáncer en el contexto del cáncer colorrectal hereditario no polipósico hemos analizado mediante PCR cuantitativa monocroma multiplex un total de 592 muestras de ADN extraído de sangre periférica (leucocitos). De ellas, 388 correspondían a individuos de 96 familias diagnosticadas con SL: 144 portadores de mutación en un gen MMR afectos de cáncer, 100 portadores de mutación no afectos y 144 no portadores. También se han estudiado 114 individuos pertenecientes a 34 familias clasificadas como CCRf-X, 57 afectos de cáncer y 57 no afectos, y 90 muestras de individuos sanos no relacionados.

De todos los estudios publicados hasta el momento en que se analiza la longitud telomérica en casos de CCR, tan solo uno lo hace en pacientes con SL (Bozzao et al., 2011a), mientras que no existe ninguna publicación en la que se estudie esta variable en el contexto del CCRf-X.

1.2.1. Longitud telomérica en el síndrome de Lynch

En nuestra serie de familias con SL analizadas, hemos observado que los pacientes afectados presentan telómeros significativamente más cortos que sus familiares no afectados portadores de la mutación patogénica en los genes MMR. Como se ha comentado en el apartado anterior, la comparación entre estudios prospectivos y retrospectivos realizados con muestras de sangre periférica en CCR esporádico sugieren que los telómeros más cortos observados en afectados estudiados retrospectivamente son una consecuencia de la enfermedad y no un factor de riesgo (Zee et al., 2009; Lee et al., 2010; Pooley et al., 2010). Sin embargo, otro estudio prospectivo más reciente encuentra asociación tanto de telómeros cortos como largos con un riesgo incrementado de CCR (Cui et al., 2012). Por ello, según nuestros resultados obtenidos en individuos con deficiencias en la maquinaria de reparación de bases desapareadas, resultaría plausible pensar que o bien los telómeros más cortos de los individuos afectados son una consecuencia de la enfermedad, o bien suponen un factor modificador de riesgo en el SL.

En línea con esta segunda opción se han publicado recientemente unos resultados de nuestro laboratorio que involucran una variante en el gen de la telomerasa, *hTERT*, con el incremento del riesgo de cáncer en pacientes con SL jóvenes (menores de 45 años), asociada además con la presencia de telómeros cortos (Bellido et al., 2013). Reforzando esta hipótesis, hemos observado en nuestra serie que los portadores no afectados muestran telómeros más largos que el grupo de controles sanos, formado por familiares no portadores e individuos sanos no relacionados. Con estos resultados podríamos interpretar que, si consideramos los telómeros cortos como factor de riesgo adicional en el SL, la presencia de telómeros más largos protegería a los portadores de mutaciones en los genes MMR de desarrollar cáncer. Así, entre las personas con mutación germinal en los genes MMR, aquellas que tienen los telómeros más cortos tendrían una susceptibilidad mayor a desarrollar cáncer que aquellos con telómeros más largos. La longitud de los telómeros podría por tanto explicar parte de la heterogeneidad fenotípica observada en el SL.

Alternativamente, en el supuesto de que los telómeros más cortos observados en los afectados con SL sean consecuencia de la enfermedad, la presencia de telómeros más largos en los portadores no afectados podría sugerir un elongamiento de los telómeros debido a la haploinsuficiencia de la maquinaria de reparación de bases desapareadas. Esta idea resulta afín al modelo propuesto por Jones y colaboradores (2012), en el que se

sugiere que telómeros constitucionalmente más largos predispondrían a desarrollar CCR. Obviamente esta hipótesis requiere una validación así como la identificación del mecanismo por el cual la haploinsuficiencia de los genes MMR causaría el alargamiento de los telómeros.

Tal y como se esperaba (Lindsey et al., 1991; Iwama et al., 1998), hemos observado un acortamiento de los telómeros con la edad en todos los grupos estudiados. No obstante, examinando de forma exhaustiva la dinámica del acortamiento telomérico con la edad en portadores de mutación en los genes MMR, hemos observado que velocidad de acortamiento con la edad es mayor en los portadores que han desarrollado cáncer. Estos resultados están en consonancia con un estudio previo en el que se observó un acortamiento de los telómeros con la edad más rápido en pacientes con SL que en controles (Bozzao et al., 2011a). Aún así, en este estudio este comportamiento diferencial se limitaba a los portadores de mutación en *MSH2*, mientras que nosotros no hemos observado diferencias entre éstos y los portadores de mutaciones en *MLH1* (42 y 87 casos respectivamente). Ellos sugieren que la haploinsuficiencia del sistema de reparación de bases desapareadas en los individuos con SL puede provocar una tolerancia a los telómeros cortos, causando un acortamiento telomérico más rápido con el paso de los años. Sin embargo, nuestros datos, que incluyen el estudio de portadores de mutación sin cáncer, invalidan esta hipótesis ya que los portadores sanos no tienen una tasa de acortamiento acelerada respecto a los controles. Por tanto, podemos considerar que esta dinámica de acortamiento telomérico más pronunciado en los pacientes de SL con cáncer bien es el resultado de haber padecido cáncer, o bien se trata de un factor de riesgo para los portadores de mutación en los genes MMR.

1.2.2. Longitud telomérica en el cáncer colorrectal familiar tipo X

Paralelamente al estudio en el SL, hemos estudiado la longitud de los telómeros en familias clasificadas como CCRf-X. Si recordamos, estas familias cumplen los criterios clínicos de Ámsterdam para el CCR hereditario no polipósico pero no presentan defectos en la maquinaria de reparación de bases desapareadas como es el caso de los pacientes con SL. En nuestra serie, con 34 familias analizadas, hemos observado que los individuos afectados tenían los telómeros significativamente más largos que sus familiares no afectados y que los controles sin cáncer. Esta observación resulta llamativa en cuanto a que es opuesta a lo observado en SL. Parece por tanto que el estado de la maquinaria MMR es crucial para definir el papel la longitud de los telómeros como factor de riesgo en el CCR

hereditario no polipósico. Dando fuerza a esta idea, existen dos publicaciones que describen la presencia de telómeros más cortos en tumores colorrectales esporádicos con IMS comparado con tumores con EMS (Takagi et al., 2000; Rampazzo et al., 2010), mostrando que el estado de la vía de reparación de bases desapareadas influye en el estado de los telómeros.

Cabe destacar que aún siendo un estudio retrospectivo se han identificado telómeros más largos en los familiares con CCR que en familiares no afectos, individuos control y pacientes con cáncer de familias con SL. Por consiguiente, podemos concluir que telómeros largos están asociados con el riesgo de cáncer en casos de CCR hereditario no polipósico con la maquinaria de reparación de bases desapareadas intacta. Varios estudios han detectado también telómeros largos asociados al CCR. Por un lado, un estudio de carácter retrospectivo con un amplio tamaño muestral (2157 casos y 3912 controles) ha relacionado la presencia de telómeros largos con el riesgo incrementado a CCR de una manera indirecta. Tanto telómeros largos como una variante en *hTERT* mostraban asociaciones significativas con el CCR (Jones et al., 2012). En otro estudio de carácter prospectivo se ha asociado tanto telómeros cortos como largos con el riesgo de CCR (Cui et al., 2012). Finalmente, una publicación muy reciente ha diferenciado esta dualidad de telómeros largos y cortos asociada a pacientes de CCR entre jóvenes y de edad avanzada respectivamente (Boardman et al., 2014).

1.2.3. Equilibrio de la longitud telomérica en la viabilidad celular

En conjunto, nuestros resultados sugieren que tanto telómeros cortos como largos pueden contribuir a la susceptibilidad al CCR, dependiendo del fondo genético del individuo. En particular en los casos familiares, esta diferencia queda determinada por el estado de la maquinaria de reparación de bases desapareadas (Figura 15).

Aún singular, resulta plausible el hecho de que ambos rasgos antagónicos, telómeros cortos y largos, actúen como factores tumorigénicos. En un primer planteamiento, hay que tener en cuenta que el acortamiento de los telómeros puede representar dos funciones opuestas, por un lado la supresión tumoral y por el otro la inestabilidad genética, evento clave para el inicio de la carcinogénesis. El elemento decisorio que decanta en un sentido u otro resulta del trasfondo genético de cada célula. Existen numerosas evidencias que indican que la parada del ciclo celular en respuesta a factores de estrés celular, entre ellos las deficiencias teloméricas, supone una potente barrera anticancerígena (Blasco,

2005). Muestra de ello es el modelo de ratón deficiente para la telomerasa que presenta telómeros cortos y a su vez resistencia al cáncer (Gonzalez-Suarez et al., 2000). Sin embargo, al combinar telómeros cortos con mutaciones en *TP53*, se da un incremento en la incidencia de tumores (Attardi et al., 2005). Vinculándolo con la vía de reparación de bases desapareadas, se ha descrito un modelo de ratón deficiente tanto para la telomerasa como para *MSH2* en el que también se anula la actividad supresora de tumores en respuesta a telómeros cortos (Martinez et al., 2009). Por tanto, queda patente la participación de *MSH2* en la respuesta del organismo a anomalías teloméricas, y por consiguiente podríamos pensar que en individuos con deficiencias congénitas en los genes MMR puedan tener un riesgo incrementado a desarrollar cáncer en presencia de telómeros cortos.

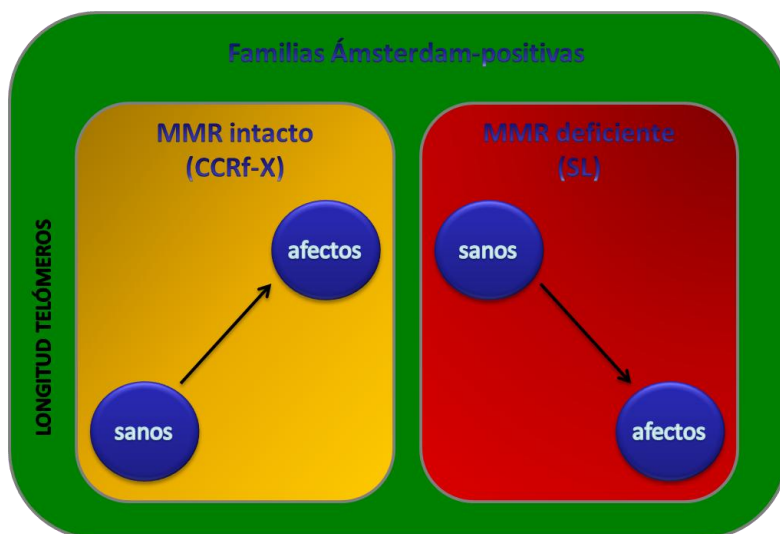


Figura 15. Comportamiento diferencial en la longitud telomérica entre familias que cumplen los criterios de Ámsterdam según el estado de la maquinaria de reparación de bases desapareadas.

Esta perspectiva formaría parte de la visión generalizada de las deficiencias teloméricas como promotoras de la iniciación y progresión tumoral, que consiste en el acortamiento telomérico y la posterior reactivación de la telomerasa tras una fase de crisis caracterizada por fusiones entre telómeros e inestabilidad cromosómica (Xu et al., 2013). Por otro lado, también es razonable pensar que telómeros largos en las células puedan incrementar el riesgo de cáncer. La presencia de telómeros largos puede provocar un retraso de la senescencia y apoptosis celular, incrementando así las posibilidades de acumular

anomalías genéticas y por tanto un mayor riesgo de transformación carcinogénica. Hay ciertas neoplasias, como el linfoma no-Hodgkin, el cáncer de mama e incluso el CCR, para las que se ha demostrado que telómeros largos son un factor de riesgo (Lan et al., 2009; Gramatges et al., 2010; Jones et al., 2012). Aunque la longitud de los telómeros en células somáticas principalmente refleja la proliferación celular, en células tumorales manifiesta el balance entre la elevada proliferación celular, con la consecuente pérdida de repeticiones teloméricas, y la actividad de la telomerasa, con síntesis *de novo* de estas secuencias. Además, múltiples estudios han mostrado que los telómeros son más cortos en tejido tumoral de CCR que en la mucosa normal adyacente (Hastie et al., 1990; Takagi et al., 1999; Gertler et al., 2004; Garcia-Aranda et al., 2006). Por ello, en nuestro caso sería interesante poder estudiar la longitud telomérica y el nivel de expresión de la telomerasa en tumores CCRf-X, y comprobar si se mantiene la relación de telómeros más largos observada en sangre periférica.

Además, comprender la dinámica de los telómeros en el cáncer es importante para poder enfocar posibles estrategias terapéuticas. Dado que la telomerasa está sobre-expresada en más del 85% de los tumores humanos, supone una de las dianas más atractivas y estudiadas desde un punto de vista farmacológico.

Con los resultados obtenidos podemos concluir que el estado de la maquinaria de reparación de bases desapareadas supone un factor diferencial que define el papel de la longitud telomérica en el CCR hereditario no polipósico. Así, la presencia de telómeros cortos o un acortamiento acelerado de éstos podría contribuir a un mayor riesgo de desarrollar cáncer en portadores de mutación en los genes MMR, hecho que podría explicar parte de las diferencias clínicas observadas entre familias diagnosticadas con SL y dentro de ellas mismas. De todas maneras estos resultados requieren validación en series independientes, si fuera posible prospectivas. Estudiando la longitud telomérica en los pacientes antes y después de desarrollarse el cáncer seremos capaces de determinar de forma concluyente si el acortamiento de los telómeros es consecuencia o contribuye al desarrollo de la enfermedad en el SL. Esto hace patente la importancia de realizar un estudio prospectivo, aun teniendo en cuenta las dificultades que ello conlleva.

Además, también sería interesante poder incluir en el estudio muestras de casos de CCR esporádico, tanto con y sin alteraciones en el sistema MMR, para ser comparadas con nuestras series de CCR hereditario. Desgraciadamente, no disponemos de muestras de casos esporádicos que hayan seguido un proceso de obtención, procesamiento y

almacenamiento equivalente a las muestras analizadas, factores que como hemos comentado afectan a la cuantificación de la longitud telomérica (Cunningham et al., 2013). Este estudio ayudaría a dilucidar si los telómeros cortos y largos son rasgos específicos o no del SL y del CCRf-X respectivamente, o si por el contrario, también son factores de riesgo de CCR en la población general.

1.3. Anticipación genética y longitud telomérica en síndrome de Lynch

Además de estudiar la longitud de los telómeros en relación al riesgo a desarrollar cáncer, también hemos evaluado su asociación con la anticipación de la edad de diagnóstico en las sucesivas generaciones descrita en las familias con SL, ya que se ha propuesto reiteradamente como posible mecanismo subyacente (Bozzao et al., 2011a; Bozzao et al., 2011b; Mendez-Bermudez y Royle, 2011).

1.3.1. Aspectos generales de la anticipación genética

La anticipación genética en el SL ha sido un tema controvertido durante muchos años, aunque el compendio de todos los estudios publicados en la última década señala esta asociación como verdadera (Bozzao et al., 2011b). Sin embargo, hasta que no se descubra el mecanismo molecular subyacente que pueda explicar genéticamente esta anticipación, no se pueden excluir sesgos o el trasfondo genético y ambiental común en las familias como posibles factores responsables.

En numerosos síndromes hay evidencias que explican la base molecular responsable de la anticipación genética. El primer mecanismo descrito fue la expansión de repeticiones trinucleotídicas durante sucesivas meiosis, que explica la anticipación genética que se observa en varias enfermedades neurodegenerativas como son la enfermedad de Huntington, el síndrome de X frágil, la distrofia miotónica y la ataxia de Friedreich, entre otras. En este caso el número de estas repeticiones nucleotídicas acumuladas es proporcional a la severidad con que se manifiesta la enfermedad (McInnis, 1996). Otro de los mecanismos asociados a la anticipación es el acortamiento gradual de los telómeros en generaciones sucesivas. Este fenómeno ocurre en la disqueratosis congénita donde mutaciones en *TERC*, gen que codifica para el componente RNA de la telomerasa, provocan el acortamiento progresivo de los telómeros al avanzar en las sucesivas generaciones, lo que provoca un incremento paulatino en la severidad de la clínica de la enfermedad (Marrone et al., 2005). Otro de los síndromes donde se ha asociado

anticipación con acortamiento telomérico es el síndrome de Li-Fraumeni, causado por mutaciones en la línea germinal en *TP53*. Se ha observado que p53 se une directamente a los bucles de los extremos de los telómeros (*t-loop*) y controla la respuesta celular en respuesta al daño en el ADN, por lo que deficiencias en esta proteína ayudan a las células a escapar de la apoptosis y seguir proliferando aún sobrepasada la longitud crítica de los telómeros. Así, telómeros cada vez más cortos en la línea germinal de los portadores se transmiten a la descendencia, lo que provoca una edad de desarrollo de la enfermedad cada vez más temprana (Tabori et al., 2007). En el caso de Li-Fraumeni, también hay indicios de que la degradación proteosómica de p53 potenciada por una variante en el gen *MDM2*, podría afectar a la anticipación genética de la enfermedad (Bougeard et al., 2006).

En relación a otros síndromes de predisposición al cáncer, además del SL, se ha descrito la presencia de anticipación genética en el cáncer de mama, páncreas, ovario, leucemia, linfoma y melanoma familiares (Hsu et al., 2000; McFaul et al., 2006; Goldberg et al., 1997; Horwitz et al., 1996; Wiernik et al., 2000; Goldstein et al., 1996). Pendiente de validación, la anticipación en el cáncer de mama hereditario se ha asociado al acortamiento telomérico en generaciones sucesivas (Martinez-Delgado et al., 2011). En el resto de síndromes hereditarios el mecanismo responsable permanece desconocido.

1.3.2. Valoración de la implicación de la longitud telomérica en la anticipación observada en el síndrome de Lynch

Estudiamos 59 familias para las que disponíamos datos de longitud telomérica en padres e hijos. De media, observamos una anticipación de la edad de diagnóstico en la segunda generación de afectos respecto a la primera de 12 años. Esta diferencia está muy probablemente sobre-estimada ya que no se tuvieron en cuenta en el análisis estadístico los factores confusores presentes en la serie. La falta de información sobre el tiempo de vigilancia en padres e hijos, la información limitada sobre la extirpación de pólipos en las colonoscopias de vigilancia, así como el tamaño muestral limitado, hacen que no se pueda realizar una evaluación precisa de la anticipación genética en estas familias. Estudios más exhaustivos de la anticipación utilizando modelos estadísticos que tienen en cuenta todas estas variables, indican que la anticipación en la edad de aparición del cáncer en sucesivas generaciones en el SL oscila entre los 3 y 9 años (Larsen et al., 2009; Boonstra et al., 2010; Boonstra et al., 2011; Bozzao et al., 2011b).

Sin embargo, el objetivo del estudio no era evaluar la presencia de anticipación genética, sino determinar si ésta estaba asociada al acortamiento telomérico. En este aspecto, nuestros datos, indican que no existe asociación entre estas dos variables. Existe un único estudio previo en el que se analiza esta relación y ellos sugieren que existe asociación entre la anticipación genética y la presencia de telómeros más largos y más cortos en los hijos para deficiencias en *MSH2* y *MLH1* respectivamente, pero esta observación está basada en tan solo el análisis de 7 parejas padres-hijos (6 para *MSH2* y 1 para *MLH1*) (Bozzao et al., 2011a). En nuestro estudio se analizaron 32, 20 y 7 familias con mutaciones en *MLH1*, *MSH2* y *MSH6* respectivamente.

Un análisis más detallado de nuestros resultados en las parejas padre/madre-hijo/a de forma individual, mostró una asociación de telómeros más cortos con anticipación en todas las parejas madre-hijo afectos portadores de mutación en los genes MMR. Es decir, todas las parejas madre-hijo (n=5) mostraban anticipación y en todos los casos los hijos tenían los telómeros más cortos que sus madres. Por el contrario, esta relación tan solo se daba en el 31% del resto de combinaciones. A pesar de que el tamaño muestral es muy reducido, esta observación abre las puertas a un estudio más completo en esta dirección en un mayor número de familias. En relación con nuestros resultados, se ha descrito una dependencia del riesgo de cáncer según el origen materno o paterno de las mutaciones en genes MMR. Se ha observado una mayor incidencia de CCR en hombres con SL en que la mutación ha sido heredada de la madre, lo que sugiere la existencia de un mecanismo modificador del riesgo de cáncer para los portadores de mutaciones en los genes MMR transmitido de la madre (van Vliet et al., 2011). En base a esta observación se ha abierto una nueva línea de investigación en nuestro grupo para estudiar la hipótesis de que la longitud telomérica pueda estar asociada al fenómeno de la dependencia de riesgo según el origen materno o paterno de la mutación en el gen MMR.

1.3.3. Otros posibles mecanismos que expliquen la anticipación genética en el síndrome de Lynch

Como alternativa al acortamiento telomérico, una posible explicación para la anticipación genética observada en SL es la acumulación de mutaciones debido a la capacidad constitutiva reducida de reparación de bases desapareadas del ADN en estos individuos. Esta idea se sugirió por primera vez en 1994 (Shibata et al., 1994), pero no ha sido hasta muy recientemente que se ha realizado el primer estudio evaluando esta hipótesis (Coolbaugh-Murphy et al., 2010). Se ha probado la existencia de un nivel de IMS en

sangre periférica de individuos heterocigotos para mutaciones en genes MMR mayor que en controles sanos y pacientes de CCR esporádico. Así, si estos niveles de inestabilidad genética observados en células normales también se diesen en la línea germinal, la carga mutacional iría consecuentemente aumentando de generación en generación. Si algunas de estas mutaciones en las células germinales aportasen una potencial ventaja tumorigénica, podrían transmitir a la descendencia un riesgo incrementado que se sumaría a la mutación en el sistema de reparación de bases desapareadas para el desarrollo del cáncer. De este modo, la acumulación progresiva de mutaciones en la línea germinal podría suponer un posible mecanismo molecular que explique la anticipación genética en estas familias. A diferencia de las enfermedades neuromusculares, en las que la base molecular de la anticipación es la expansión de repeticiones trinucleótidas en un solo *locus*, en el SL las mutaciones responsables ocurrirían en numerosos *loci* a lo largo de todo el genoma. El desarrollo de métodos más sensibles para cuantificar la IMS en células germinales y el análisis de más individuos será necesario para investigar en profundidad esta hipótesis.

2. ESTUDIO DE *GALNT12* COMO GEN CAUSAL EN CÁNCER COLORRECTAL FAMILIAR TIPO X

Uno de los objetivos que trabajamos en esta tesis doctoral es el estudio de la implicación de mutaciones en línea germinal en el gen *GALNT12* en el CCRf-X. Este estudio está fundamentado en evidencias previas que asociaban mutaciones en el gen con un incremento de riesgo a desarrollar CCR y que sugerían que este gen podría explicar la agregación de CCR en familias, incluidas las de alto riesgo.

2.1. Valoración de *GALNT12* como gen candidato

La glicosilación proteica es un proceso bioquímico esencial en todos los eucariotas mediante el que se añade un glúcido a una proteína durante o después de la traducción. Las funciones de los O-glicanos son muy diversas: Intervienen en la protección de superficies epiteliales, en el mantenimiento de la conformación de proteínas y participan en los procesos de adhesión celular y en el sistema inmune. Además, están relacionadas con la regulación del crecimiento, la embriogénesis, el desarrollo y la muerte celular (Brockhausen, 1999). Interviniendo en todas estas funciones, no es de extrañar que la estructura de los O-glicanos esté frecuentemente alterada en cáncer y contribuya en gran medida al fenotipo y a la biología de las células tumorales. La naturaleza patogénica de estas alteraciones en la estructura de la mucina es el resultado de que ligandos responsables de la interacción entre las células cancerígenas y su microambiente hayan cambiado. Esto influencia el crecimiento y la supervivencia de la célula, su habilidad invasiva y metastásica, y sus interacciones con receptores de membrana celular o células del sistema inmune. Puesto que los carbohidratos no pueden mutarse directamente, resulta evidente que sea la maquinaria enzimática responsable de la síntesis de estos complejos la que provoca estas alteraciones. En las células tumorales muchas de las enzimas que intervienen en la biosíntesis de los O-glicanos, como las N-acetilgalatosaminiltransferasas, están reguladas a la alta o a la baja (Brockhausen, 2006).

Para *GALNT12*, se ha correlacionado la presencia de su transcrito como índice negativo de la carcinogénesis y de la progresión del CCR (Guo et al., 2004). Se ha determinado por PCR cuantitativa a tiempo real que en condiciones normales *GALNT12* se expresa especialmente en órganos del sistema digestivo como el colon (Guo et al., 2002), mientras que su expresión es baja o despreciable en diferentes líneas celulares de cáncer de colon (Colo205, HCT15, LSC, LSB, SW480 y SW116) y en tejido tumoral respecto al normal de

pacientes de CCR. Además, la expresión de *GALNT12* muestra una correlación a la inversa con el estadio TNM y la presencia de metástasis (Guo et al., 2004). Por otra parte, recientemente se han asociado dos variantes genéticas en *GALNT12* con la supervivencia global en linfoma folicular (Gibson et al., 2012), y mutaciones en genes que codifican para otras N-acetilgalatosaminiltransferasas, con cáncer de mama y de ovario (Wood et al., 2007; Phelan et al., 2010).

Varios estudios de ligamiento a nivel genómico en familias con CCR realizados con el fin de identificar nuevos genes causales del CCR hereditario, han dado como resultado una señal positiva de ligamiento en la región 9q22.2-31.2, cerca de donde está situado *GALNT12* (9q21-33) (Wiesner et al., 2003; Skoglund et al., 2006; Kemp et al., 2006b; Gray-McGuire et al., 2010). El hecho de que no se hayan identificado mutaciones causales en los genes que alberga el pico de ligamiento, convertía a *GALNT12* en un buen candidato para realizar estudios genéticos. En 2009, Guda y colaboradores identificaron en este gen dos mutaciones somáticas en 30 líneas celulares de CCR con EMS, y siete mutaciones en línea germinal en 272 (2,6%) pacientes con cáncer de colon, ausentes en 192 controles sin cáncer. Demostraron que las dos variantes somáticas y seis de las germinales inactivan la función normal de la enzima, y la séptima mutación en la línea germinal reduce su actividad. Más recientemente, otro estudio analizó el gen en una cohorte de 118 familias con agregación de CCR (Clarke et al., 2012). Hallaron dos mutaciones de cambio de sentido en cuatro familias (3,4%) que cumplían los criterios de Bethesda. Una de las mutaciones, presente en tres individuos índice estudiados, coincidía con la variante en *GALNT12* con actividad reducida reportada en el estudio anterior.

Teniendo en cuenta todas estas premisas y asumiendo los resultados previos publicados en CCR, resultaba plausible plantearse que este gen pudiese estar detrás del riesgo de CCR en una fracción de las familias clasificadas como CCRf-X.

2.2. Evaluación de las variantes de *GALNT12* identificadas

En nuestro centro, contábamos con una serie de 43 familias CCRf-X atendidas en las unidades de consejo genético (34 ICO Hospitalet y 9 ICO Girona). En una primera aproximación estudiamos mediante la técnica del CSCE (del inglés, *Conformation Sensitive Capillary Electrophoresis*) los exones 4 y 6 del gen, ya que eran donde se localizaban la mayoría de las variantes patogénicas descritas anteriormente. No se identificaron más que dos SNP descritos en las bases de datos públicas (rs41306504 y

rs3216734). Tras estos resultados preliminares decidimos estudiar todo el gen mediante secuenciación. Asimismo, incluimos muestras provenientes de Castilla y León (n=27) y de la Comunidad Valenciana (n=33) para aumentar el tamaño muestral del estudio. Por consiguiente, analizamos un total de 103 familias CCRf-X mediante la secuenciación de las zonas codificantes y región colindante del gen en las muestras de ADN de sangre periférica de los individuos índice. Como resultado, identificamos un total de 20 variantes en 90 pacientes. Ocho de ellas estaban situadas en la región codificante de *GALNT12*, 7 en zonas intrónicas y 5 en la región 3'-UTR. Tras un análisis exhaustivo de las variantes, teniendo en cuenta datos de frecuencia en población sana, de cosegregación con la enfermedad y de predicción de función (*in silico*), pudimos concluir que ninguna de estas variantes era patogénica.

Las variantes no sinónimas son aquellas situadas en la región codificante de un gen que comportan el cambio de un aminoácido, lo que puede afectar a la función de la proteína y tener unas consecuencias fenotípicas en el individuo portador. La predicción de función de estas sustituciones aminoacídicas mediante métodos *in silico* basados en la estructura y la secuencia, ayudan a interpretar la patogenicidad de estos cambios. De las 8 variantes en regiones codificantes identificadas en nuestro estudio tres correspondían a mutaciones de cambio de aminoácido, c.136G>A (p.G46R), c.356A>T (p.E119V) y c.781G>A (p.D261N), observadas en uno, 18 y 2 individuos respectivamente. Todas ellas habían sido previamente descritas en las bases de datos públicas (dbSNP o NHLBI Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>)) y habían sido identificadas en los estudios genéticos previos de *GALNT12* en CCR (Guda et al., 2009; Clarke et al., 2012).

Los algoritmos *in silico* de predicción de función son una herramienta muy útil para interrogar variantes no sinónimas, y es la primera aproximación generalmente utilizada en los laboratorios de investigación y de diagnóstico genético. El 75-90% de los valores de predicción obtenidos por los diferentes algoritmos de clasificación *in silico* concuerdan favorablemente con otros análisis funcionales que están en uso (Chan et al., 2007). Aún así, estas clasificaciones no dejan de ser una predicción con una precisión de alrededor del 80%, por lo que para la aplicación clínica deberían acompañarse de otras herramientas de interpretación como ensayos funcionales o análisis de la cosegregación de la variante genética con la enfermedad en la familia. Además, dado que los diferentes algoritmos se basan en distintos atributos, puede haber incongruencias entre ellos (Tavtigian et al., 2008). En nuestra serie de resultados, para estos tres cambios localizados no se predecían efectos relevantes en la proteína por los algoritmos SIFT y

Condell, mientras que PolyPhen-2 predecía efectos funcionales para el cambio p.D261N (valor 1) y p.E119V (valor 0,8, en la modalidad HumanDiv). De todas maneras estos indicios fueron desestimados, ya que las tres variantes habían sido identificadas en controles sanos y testadas funcionalmente en el estudio de Guda y colaboradores, donde se demuestra que estos cambios no alteran la actividad enzimática de la proteína (Guda et al., 2009).

Los otros cinco cambios identificados en la región codificante del gen predecían cambios sinónimos. Dos de ellos eran variantes que no habían sido descritas previamente en las bases de datos públicas, c.237C>T (p.G79=) y c.399C>T (p.D133=), y otras dos tenían frecuencias alélicas en la población general menores del 2%, c.1392C>T (p.P464=) y c.1497C>T (p.N499=). Se analizaron todas las variantes con algoritmos *in silico* de predicción para posibles alteraciones en los sitios de *splicing*. Los programas NetGene2 y SoftBerry no predijeron alteraciones para ninguno de los cambios, pero el algoritmo NNSplice predecía un nuevo sitio donador para c.237C>T, aunque con una puntuación menor que el sitio donador salvaje (0,52 y 0,8 respectivamente). Un análisis de la mutación en el padre del probando, afecto de cáncer de recto a los 64 años y perteneciente a la rama familiar con casos de CCR, desveló que no era portador del cambio, por lo que se pudo descartar como potencial variante patogénica al no cosegregar con la enfermedad.

En cuanto a las variantes intrónicas identificadas, tan solo dos están situadas en proximidad al sitio donador de *splicing* consenso, por lo que más posiblemente podían alterarlo. Aún así, no se predijo ningún efecto en el proceso de *splicing* mediante los tres programas de predicción *in silico* utilizados. Para el cambio c.1036-4G>A se pudo realizar el análisis de cosegregación gracias a la disponibilidad de muestra de la madre del probando, afecta de dos tumores de colon metacrónicos. La mutación no estaba presente en la rama afectada de la familia, descartándose definitivamente como mutación causal de la agregación familiar de cáncer.

En conjunto, podemos concluir que no se han identificado cambios relevantes que afecten la actividad enzimática en las 103 familias CCRf-X analizadas. De igual modo, Clarke y colaboradores (2012) tampoco identificaron mutaciones patogénicas en 26 familias Ámsterdam-positivas estudiadas. Por tanto, no debemos considerar *GALNT12* como un gen de alta penetrancia en la predisposición al CCR. De las 20 variantes que hemos encontrado 7 son coincidentes con las identificadas en los dos estudios previos (Guda et

al., 2009; Clarke et al., 2012), pero ninguna de ellas ha sido descrita posteriormente como clínicamente patogénica en un estudio que, integrando las diferentes herramientas informáticas tanto a nivel funcional como estructural, ha valorado las variantes descritas en diversas N-acetilgalatosaminiltransferasas que aparecen en la literatura (Hussain et al., 2014).

2.3. Contribución de *GALNT12* a la predisposición al CCR

Considerando la presencia de mutaciones en *GALNT12* en casos de CCR familiar que no cumplen los criterios de Ámsterdam, Clarke y colaboradores (2012), identificaron dos mutaciones de cambio de sentido en cuatro familias que cumplían los criterios de Bethesda revisados. El cambio c.907G>A (p.D303N; rs145236923), identificado en tres familias, reduce la actividad enzimática de la proteína a un 37% pero no la inactiva, en contraste con las otras mutaciones encontradas por Guda y colaboradores (2009) para las que la inactivación es prácticamente total. Además, las evidencias de cosegregación de la variante con el CCR en una de las dos familias analizadas son bastante débiles (Clarke et al., 2012). En consonancia, el análisis en profundidad realizado en el estudio de Hussain y colaboradores (2014) para las variantes de *GALNT12* descritas, revela incongruencias para este cambio entre los distintos algoritmos utilizados, por lo que su patogenicidad no es del todo clara. El otro cambio identificado en un caso de CCR familiar, c.1187A>G (p.Y396C), no ha sido descrito en otros estudios. Esta variante está situada en el dominio catalítico de la proteína y los algoritmos de predicción *in silico* lo clasifican como deletéreo (Clarke et al., 2012). Por consiguiente, si aceptamos este segundo cambio como el único potencialmente patogénico, el porcentaje de mutaciones identificadas en CCR con agregación familiar quedaría reducido al 0,8%. La falta de información de la historia familiar sobre los portadores de mutaciones que inactivan la función del gen en Guda y colaboradores (2009), no permite valorar su participación en la predisposición al CCR familiar. Deberían realizarse más estudios con casos de CCR que cumplan criterios clínicos menos estrictos que los de Ámsterdam para evaluar el papel de *GALNT12* como gen de moderada o baja penetrancia en el CCR familiar.

Una de las razones por las que *GALNT12* podría actuar como un gen de baja-moderada penetrancia en vez de alta, es que pertenece a una familia de enzimas muy amplia con hasta 20 genes homólogos que las codifican. Existe una gran redundancia y complejidad en la familia de enzimas que controlan el inicio de la O-glicosilación, de manera que ésta es la familia de glicosiltransferasas más extensa que cubre un solo tipo de unión

glicosídica, demostrando que el concepto “una enzima-una unión” no es universal. No obstante, estudios recientes muestran que el gran número de genes *GALNT* en la familia no aportan funciones completamente redundantes. Existen ciertas diferencias en la especificidad por el péptido aceptor del sustrato, aunque mantienen un considerable solapamiento (Gerken et al., 2011). Además, hay que tener en cuenta que hay genes de la familia cuya expresión está restringida a unos tejidos determinados, como es el caso de *GALNT12*, pero que hay otros con expresión ubicua. Así, otro miembro de la familia podría estar cubriendo parte de la deficiencia provocada por mutaciones germinales en este gen.

Finalmente, cabe considerar que otros genes situados dentro del pico de ligamiento 9q22.2-31.2 podrían explicar algunos de los casos de CCR familiar. De hecho, los estudios de replicación del pico de ligamiento sugieren que el *locus* causal localizado en 9q es específico de CCR familiar con una edad temprana de diagnóstico (Kemp et al., 2006b; Gray-McGuire et al., 2010), lo que corresponde aproximadamente a un 20% de las familias estudiadas. En Guda y colaboradores (2009) la edad media de diagnóstico de los pacientes portadores de una mutación inactivante era de 71 años, y la edad de diagnóstico de los probandos en Clarke y colaboradores (2012) es de 21 (p.Y396C), 72, 35 y 72 años (p.D303N). Por tanto, tan solo dos encajarían con la característica típica de las familias asociadas al pico de ligamiento en cuestión, que es la presencia de cáncer a edad temprana.

Un gen situado en esta región cromosómica que inicialmente parecía un buen candidato es *TGF β 1*, ya que forma parte de la vía de señalización de TGF- β , típicamente alterada en cáncer. Aunque con cierta controversia, se había propuesto un polimorfismo común situado en el exón 1 del gen (rs11466445, *TGF β 1*6A/9A*) como posible alelo de riesgo en cáncer, incluido el CCR (Pasche et al., 1999; Pasche et al., 2004; Skoglund et al., 2007). Sin embargo, el compendio de estudios publicados en la última década desestiman la delección *TGF β 1*6A* como alelo de predisposición al CCR (revisado por Valle, 2012). De manera similar, estudios en ratones y un primer estudio caso-control de la expresión alelo-específica de *TGF β 1* sugerían la relevancia de la haploinsuficiencia del gen en el CCR (Valle et al., 2008; Zeng et al., 2009). Validaciones posteriores, entre ellas una llevada a cabo por nuestro grupo (Seguí et al., 2011), utilizando diferentes métodos de cuantificación, marcadores, tejidos analizados y poblaciones, parecen indicar que no existen diferencias significativas en la expresión alelo-específica de *TGF β 1* entre casos y controles (revisado por Valle, 2012). De todas maneras, con las técnicas actuales tan solo se puede determinar la expresión alelo-específica de *TGF β 1* del 25-60% de

individuos, por lo que son necesarias nuevas técnicas más precisas e informativas para poder obtener una respuesta definitiva.

Otros genes candidatos localizados en la región 9q son *ZNF367*, *HABP4* y *GABBR2*. *HABP4* es especialmente interesante a nivel funcional, ya que codifica para una proteína que interacciona con CHD3, una proteína nuclear involucrada en la regulación de la transcripción y la reorganización de la cromatina, con el receptor de la Kinasa 1 activa (*RACK1*), y que co-precipita con el promotor tumoral *PKC* (Hansen et al., 1989).

3. LA SECUENCIACIÓN DE EXOMAS COMO HERRAMIENTA PARA DESCIFRAR LAS CAUSAS DEL CÁNCER COLORRECTAL FAMILIAR TIPO X

En los últimos años el avance tecnológico de los métodos de genotipado y de secuenciación de alto rendimiento ha proporcionado potentes herramientas para explorar la base genética tanto de enfermedades monogénicas como poligénicas. Hoy en día se pueden obtener fácilmente grandes cantidades de datos, aunque interpretar esta información genómica aún supone un reto. El abaratamiento progresivo de la secuenciación paralela masiva, tanto para el estudio de exomas y genomas completos como para analizar grupos de genes concretos, ha acercado su uso tanto a la investigación como a la clínica. Sin embargo, los datos obtenidos tras este tipo de análisis conllevan un tratamiento especial desde el punto de vista clínico y ético, principalmente por la divulgación de hallazgos accidentales y la interpretación de variantes de significado desconocido (Rigter et al., 2013).

Para la búsqueda de genes de alta penetrancia mediante secuenciación masiva se pueden estudiar el exoma o el genoma completo de los individuos de interés. Para la identificación de nuevos genes de CCR hereditario nosotros hemos escogido la secuenciación de exomas completos. La gran mayoría de mutaciones patogénicas en enfermedades mendelianas se localizan en las zonas codificantes del genoma o en regiones colindantes a ellas, por lo que quedarían cubiertas con esta aproximación. Además con esta técnica se consigue una profundidad de secuencia mayor, favoreciendo una mejor detección de mutaciones. Asimismo, no hay que menospreciar que ésta es una tecnología más asequible a nivel económico y los datos obtenidos requieren menor espacio para su almacenamiento informático.

Como en todo experimento científico, para obtener un rendimiento óptimo de la secuenciación de exomas en la búsqueda de nuevos genes de susceptibilidad, es esencial un diseño experimental adecuado. Cuando las características fenotípicas de la enfermedad son únicas y muy homogéneas se esperaría que mutaciones en un único gen o unos pocos genes con funciones similares expliquen la mayoría de los casos. En este escenario, se secuenciarían individuos afectados de varias familias para buscar un gen que esté alterado en todas o la mayoría de ellas. No obstante, para el CCRf-X, dada la presupuesta heterogeneidad genética del grupo, hemos considerado preferible limitar el estudio a cada familia de forma individual. El CCRf-X es al fin y al cabo un cajón de sastre en el que se han agrupado todas aquellas familias que cumplen unas determinadas

características clínicas, pero cuyo fondo genético es muy probablemente heterogéneo. Asimismo, el tipo de herencia esperado para la variante patogénica juega un papel importante en el análisis de los datos. Se seleccionarán los genes candidatos bajo premisas diferentes si se espera una herencia autosómica dominante, autosómica recesiva o ligada al sexo. Teniendo en cuenta que los casos de CCRf-X cumplen los criterios clínicos de Ámsterdam para el CCR familiar, en un principio esperaríamos una herencia dominante.

Al analizar una familia determinada, resulta de suma importancia la selección de los familiares que se estudiarán. Para aumentar las posibilidades de homogeneidad genética, puede ser relevante establecer criterios de homogeneidad, como pueden ser las características patológicas y moleculares de los tumores: Tipo de tumor y localización, presencia/ausencia de inestabilidad de microsatélites, etc. Debido a la elevada frecuencia del CCR en la población general, es preferible estudiar a aquellos familiares que hayan desarrollado cáncer a edades tempranas, para evitar el estudio de casos esporádicos que hayan ocurrido dentro de la misma familia, es decir, de fenocopias. La inclusión de individuos no afectados puede ayudar a delimitar mejor las variantes candidatas, especialmente si se trata de familiares que a edades avanzadas no han desarrollado cáncer y que además tengan descendientes sin cáncer, también con edades relativamente avanzadas. Por desgracia, en la práctica el diseño experimental está condicionado a la accesibilidad de muestras. Remarcar en este sentido la importancia de la unidad de consejo genético, encargada de que se realice una buena recogida de muestras, información y seguimiento de las familias.

Las bases de datos públicas como dbSNP, *1000Genomes* o *NHLBI Exome Sequencing Project* son muy útiles para filtrar aquellas variantes que están presentes con mayor frecuencia en la población, y por tanto que son menos probablemente patogénicas. Por lo general, buscaremos cambios con frecuencias poblacionales menores del 1%. Sin embargo, debemos tener presente la posibilidad de que existan mutaciones patogénicas recurrentes que hayan sido descritas en las bases de datos, así como dar importancia a la creación de bases de datos locales que recopilen la variación genética específica de la población de donde proceden las muestras a estudio.

Una vez se han filtrado los resultados obtenidos de la secuenciación en función de las premisas deseadas, se obtiene una lista más o menos extensa de genes candidatos. A continuación, se ha de valorar la naturaleza de las variantes candidatas: Creación de

codón de parada nuevo, cambio de sentido de aminoácido, alteración de los sitios de *splicing*, cambio de pauta de lectura, sinónima. Los algoritmos de predicción *in silico* pueden ayudar a seleccionar las variantes más probablemente patogénicas, pero no ha de ser un factor eliminatorio. Finalmente, la búsqueda bibliográfica para inferir la participación del gen según su función o vías en las que participa puede ser muy útil para refinar la selección de genes en los que realizar estudios más exhaustivos.

En esta tesis se presentan los resultados obtenidos del estudio de dos familias clasificadas como CCRf-X mediante secuenciación de exomas. En ambos casos creemos haber identificado el gen responsable de la susceptibilidad al CCR en la familia, por lo que consideramos que la secuenciación de exomas completos es una buena herramienta para interrogar y esclarecer la base genética del cáncer en este tipo de familias.

4. MUTACIONES EN *MUTYH* ENMASCARADAS BAJO UN FENOTIPO INUSUAL

Una de las dos familias CCRf-X que seleccionamos para el estudio mediante secuenciación de exomas ha resultado ser un caso asociado a mutaciones en *MUTYH*. Esta familia estaba formada por 6 miembros afectados de CCR repartidos en dos generaciones, cumplía los criterios de Ámsterdam I y, en el momento de diagnóstico del cáncer, tan solo se habían detectado 2 pólipos en uno de los pacientes, por lo que se consideró como una familia con CCR hereditario no polipósico. Tres de los individuos de los que se disponía muestra tumoral mostraban intacta la maquinaria de reparación de bases desapareadas, mientras que un cuarto presentaba IMS. No obstante, no se identificaron mutaciones en la línea germinal en los genes MMR. Por ello esta familia fue clasificada como CCRf-X y seleccionada para el estudio de exomas con el objetivo de identificar el gen causal de la agregación de CCR.

4.1. Valoración de las variantes identificadas por secuenciación de exomas

Se realizó la secuenciación de exomas completos en ADN extraído de sangre periférica de 4 afectados y un miembro no afecto de la familia, padre de 5 hijos sanos. No se incluyó el paciente con el tumor con IMS por ser una característica molecular discordante con el resto de casos. Con el patrón de herencia que mostraba la familia se esperaba un gen de predisposición de carácter autosómico dominante. Así, se enfocó el análisis de los resultados para la identificación de variantes poco frecuentes presentes en todos los miembros afectados y no en el no afecto. Se localizaron 11 variantes que cumplían estos requisitos, entre ellas una delección en el exón 12 de *MUTYH* (c.1147delC; p.Ala385Profs*23), conocido gen causal de la poliposis asociada a *MUTYH*. Ésta es una mutación patogénica recurrente en individuos de origen europeo que representa el 5,14% de todas las mutaciones identificadas en un estudio multicéntrico en poblaciones de Holanda, Reino Unido y Alemania (Nielsen et al., 2009; Nielsen et al. 2011). Dos estudios han demostrado que dificulta severamente tanto la unión de la proteína al ADN como su actividad reparadora (Parker et al., 2005; Ali et al., 2008).

Como *MUTYH* es un gen de predisposición al cáncer y/o poliposis colónica de carácter recesivo, se realizó un análisis más exhaustivo de su secuencia, en busca de mutaciones bialélicas. Encontramos que los 3 afectados de la segunda generación de los que disponíamos muestra eran portadores de una segunda variante novel en el gen: c.43A>G;

p.M15V, incluido el familiar con el tumor con IMS. Por el contrario, no se identificó ninguna variante adicional en los afectos de la primera generación

4.1.1. Riesgo de cáncer en portadores heterocigotos de mutaciones en *MUTYH*

La tasa de mutaciones monoalélicas en *MUTYH* en controles es del 1,5-2%, y la gran mayoría corresponden a dos variantes comunes (p.G396D y p.Y179C) (Cleary et al., 2009). Está bien documentado que mutaciones bialélicas en *MUTYH* confieren un riesgo elevado de CCR (43-100%) (Sampson et al., 2003; Farrington et al., 2005), pero ha habido mucho debate acerca de si los portadores de mutaciones monoalélicas tienen también un riesgo incrementado (Nielsen et al., 2011). Diversos meta-análisis que incluyen los estudios caso-control que se han realizado indican un riesgo bajo de CCR para los portadores heterocigotos, con *odds ratios* entre 1,11 y 1,27 (Webb et al., 2006; Tenesa et al., 2006; Avezzu et al., 2008). Pero estudios en familias estiman que los familiares de primer grado heterocigotos tienen un riesgo de CCR entre 2 y 3 veces superior que el riesgo en la población general (Jenkins et al., 2006; Jones et al., 2009b). Además, al estratificar por edad se ha observado un incremento del riesgo en los portadores monoalélicos mayores de 55 años (Farrington et al., 2005; Jenkins et al., 2006).

En base a estos datos resulta plausible pensar que variantes de *MUTYH* en heterocigosis pudiesen estar asociadas a un riesgo moderado a desarrollar CCR. En estos individuos heterocigotos la susceptibilidad al cáncer llegaría con una segunda mutación somática en *MUTYH*, que convergería en células con defectos en la reparación por escisión de bases que podrían inducir a la tumorigénesis. Consecuentemente, tiene sentido pensar que en estos casos el CCR se desarrolle a una edad más tardía, lo que concuerda con lo observado en los dos individuos afectos monoalélicos de la primera generación afecta de la familia estudiada (II.2 y II.6), con edades al diagnóstico de 85 y 73 años. Para valorar más en profundidad esta posibilidad habría que estudiar las características clínicas asociadas a la mutación c.1147delC en concreto en un número mayor de individuos, si fuera posible, en portadores homocigotos, portadores heterocigotos y en portadores heterocigotos compuestos (portadores también de otra mutación distinta en *MUTYH*); ya que se ha especulado que la penetrancia de las diferentes mutaciones patogénicas de *MUTYH* puede ser distinta (Cleary et al., 2009).

Alternativamente, se podría considerar que el riesgo moderado aportado por la variante c.1147delC esté incrementado por factores genéticos de riesgo adicionales, o incluso por

factores ambientales. En este caso encajaría el hecho de que el tumor del afecto II.6 no posea la transversión c.34G>T (p.G12C) en *KRAS*, comúnmente encontrada en los tumores asociados a *MUTYH*.

4.1.2. Estudio funcional de la variante novel

La mutación c.43A>G (p.M15V) identificada en los individuos de la segunda generación afecta no ha sido descrita previamente, ni en las bases de datos públicas (dbSNP, *1000Genomes* o *NHLBI Exome Variant Server*), ni en 500 alelos de individuos de origen español (Quesada et al., 2011). Esta mutación supone un cambio de la metionina 15 por una valina para la isoforma mitocondrial (α), clasificado por todos los programas de predicción *in silico* utilizados como funcionalmente relevante. La metionina afectada está muy conservada a lo largo de la evolución. Por otro lado, este cambio también inactiva el codón de inicio para los transcritos que codifican las isoformas nucleares de *MUTYH* (β y γ). *MUTYH* es un gen que presenta un alto grado de *splicing* alternativo, con expresión diferencial dependiente del órgano (Plotz et al., 2012). Se ha visto que las isoformas nucleares de la proteína son especialmente relevantes en el colon derecho, localización de al menos 5 de los tumores desarrollados en la familia estudiada.

En colaboración con el laboratorio del Dr. Plotz de la Universidad de Frankfurt, se estudió la expresión de los diferentes transcritos de *MUTYH* en presencia del cambio c.43A>G. El análisis de expresión se realizó mediante PCR cuantitativa en ARN extraído de linfocitos de varios miembros de la familia, portadores de una o las dos variantes identificadas, así como en varios controles. Se observó una gran variabilidad inter-individual en la expresión global de *MUTYH* y de las diferentes isoformas por separado. Al cultivar los linfocitos una semana para inducir la expresión de las isoformas nucleares, se detectó una reducción estadísticamente significativa de los transcritos β y γ en los individuos portadores del cambio c.43A>G. Sin embargo, esta reducción de la expresión no representa la pérdida completa del alelo c.43A>G, resultados que se confirmaron por secuenciación directa del ARN mensajero. A pesar de no mostrar un efecto muy relevante en la transcripción de las isoformas, la mutación, afecta a la Metionina 1 de las isoformas nucleares y por tanto altera el codón de iniciación de la traducción. Teóricamente, el codón de iniciación de las formas nucleares se trasladaría del exón 2 al exón 4, dando lugar a una proteína exactamente igual que la que se traduce a partir del transcrito BAA89339, una proteína muy inestable que además no tiene elementos funcionales relevantes (Plotz et al., 2012).

Todos estos resultados apuntan a que el cambio c.43A>G es una mutación patogénica que impide la correcta traducción de las isoformas nucleares (β y γ) de *MUTYH*.

4.2. Características clínicas y moleculares de la familia

Las características clínicas que actualmente se tienen en cuenta para realizar el estudio genético de *MUTYH* (Brand et al., 2012) pueden dejar fuera del marco de consideración a familias, como la nuestra, con fenotipos poco habituales.

En el momento del asesoramiento genético realizado a la familia en estudio, ésta mostraba unas características fenotípicas que condujeron a su clasificación como CCRf-X, lo que implica un modelo de herencia aparentemente dominante y la ausencia de poliposis.

Primero de todo, quizás el rasgo de confusión más claro es el patrón de herencia aparentemente dominante que se observa en la familia estudiada. Si los portadores heterocigotos de mutaciones en *MUTYH* realmente tienen un riesgo incrementado de CCR tal y como se ha sugerido (Jones et al., 2009b), puede provocar que aparezcan en una misma familia más individuos afectados de los que se esperaría en un modelo de herencia autosómico recesivo, confundiéndolo en consecuencia con un patrón dominante. La predisposición incrementada en estos casos vendría dada por la deficiencia monoalélica de *MUTYH*, probablemente junto a otros factores de riesgo, ya sean ambientales o genéticos. Un indicio que podría ayudarnos a desenmascarar la naturaleza recesiva de estas variantes es la extrema anticipación en la edad de diagnóstico que se da entre la primera (85 y 73 años) y la segunda (43, 45 y 46 años) generación afectada.

Por otro lado, aunque la mayoría de individuos con mutaciones bialélicas en *MUTYH* desarrollan pólipos (<100), la poliposis no es una característica siempre presente en este síndrome. Se han descrito individuos con poliposis asociada a *MUTYH* que han desarrollado CCR sin o muy escasos pólipos (Croitoru et al., 2004; Farrington et al., 2005; Balaguer et al., 2007; Cleary et al., 2009). Recopilando los resultados de ocho estudios poblacionales, se ha observado que el 35% de portadores de mutaciones bialélicas en *MUTYH* con CCR no presentaban pólipos (Nielsen et al., 2011). Por ello, hay que considerar la posible presencia de individuos afectados de CCR sin pólipos en el momento de diagnóstico, incluso a edades relativamente avanzadas (en la década de los 50 años). Si esto sucede en varios individuos afectados de una misma familia, como en el caso en

estudio, puede confundirse con un caso de CCR hereditario no polipósico. En este sentido, queda patente la importancia de realizar un buen seguimiento de los miembros de las familias cuya causa genética de predisposición permanece desconocida. En la familia estudiada no se habían detectado pólipos ni adenomas en el momento de diagnóstico del CCR ni cuando fueron asesoradas en la consulta de consejo genético, sin embargo años más tarde se detectaron 34 (21 adenomas) y 7 (3 adenomas) pólipos en dos de los afectos portadores de mutaciones bialélicas de la segunda generación.

Otra característica que hemos observado en nuestra familia es la coexistencia de tumores con estabilidad e inestabilidad de microsatélites. La gran mayoría de los CCR de personas con esta enfermedad presentan estabilidad de microsatélites, mientras que se ha descrito IMS elevada en tan solo una minoría (de media 4%, con rango del 0 al 18%) (Nielsen et al., 2011). Aunque parece claro que la vía de la inestabilidad de microsatélites no es muy común en los tumores colorrectales asociados a *MUTYH*, se han descrito ciertas similitudes histológicas entre estas neoplasias y las caracterizadas por IMS (SL y esporádicos con elevada IMS), tales como una localización preferentemente en el colon proximal, presencia de tumores sincrónicos, rasgos mucinosos y una alta frecuencia de linfocitos infiltrantes de tumor (Nielsen et al., 2009). La presencia de IMS no debería ser por tanto un rasgo de exclusión para la sospecha de enfermedad asociada a *MUTYH*.

Finalmente, comentar que los resultados obtenidos con el estudio de esta familia apoyan la utilización en la práctica diagnóstica del análisis del codón 12 de *KRAS* en tumores colorrectales como cribado de individuos, sobretudo de aquellos con CCR y sin poliposis, para realizar el estudio genético de *MUTYH* en todos aquellos casos en los que se identifique *KRAS* c.34G>T (p.G12C), tal y como ya se ha propuesto anteriormente (van Puijenbroek et al., 2008). *MUTYH* reconoce y elimina los desapareamientos 8-oxo-G:A del ADN, por lo que individuos con deficiencias en este gen tienden a acumular transversiones G>T a nivel somático (Nielsen et al., 2011). Por ellos, el cambio c.34G>T en *KRAS* (p.G12C) es un rasgo molecular característico de los tumores asociados a *MUTYH* que se identifica en el 64% de estos casos, frente a un 29% de los CCR esporádicos (Lipton et al., 2003; Nielsen et al., 2011).

Los avances en los últimos años en la NGS han facilitado el diseño de ensayos con paneles de genes de predisposición al CCR, o al cáncer en general, que permiten su estudio simultáneo en un conjunto de muestras (Pritchard et al., 2012; Stadler et al., 2014). Poco a poco estos avances se van aplicando cada vez más en la clínica, lo que

permitirá detectar más fácilmente y describir más en profundidad estos casos con fenotipos solapados entre síndromes de predisposición al cáncer.

5. IDENTIFICACIÓN DE UN NUEVO GEN DE CÁNCER COLORRECTAL HEREDITARIO

5.1. Hallazgos de la secuenciación de exomas en una familia con cáncer colorrectal familiar tipo X

La segunda familia clasificada como CCRf-X que seleccionamos para ser estudiada mediante secuenciación de exomas completos, consta de tres individuos diagnosticados de CCR a los 42, 67 y 72 años, repartidos en dos generaciones. Se secuenció el ADN extraído de sangre periférica de todos ellos, identificando un total de 32 variantes genéticas no descritas previamente o con frecuencias alélicas poblacionales menores al 1%, compartidas por los tres individuos afectados. De todas las variantes, cuatro correspondían a mutaciones sin sentido. Este tipo de mutaciones generalmente son consideradas como potencialmente no funcionales, ya que crean un codón de parada prematuro de manera que se traduce en una proteína truncada. En tres de los genes (*DACT2*, *DARC* y *FAM55B*) la variante identificada está situada en la zona final del transcrito (codones 264 de 280, 256 de 338 y 529 de 559 respectivamente). Contrariamente, el cuarto cambio sin sentido identificado, en el gen *GEN1X*, está localizado cerca del inicio de traducción (codón 47), lo que aumenta las posibilidades de que la función de la proteína esté drásticamente afectada. Además, los tres individuos afectados compartían 5 variantes de cambio de sentido clasificadas por los programas *in silico* de predicción de función como posiblemente patogénicas, pero el estudio bibliográfico de los genes identificados no desveló ningún candidato especialmente sugestivo de ser el causal del CCR en la familia.

Centrándonos en la función de los genes seleccionados tras el análisis de la secuenciación de exomas, tres de ellos parecían especialmente interesantes: *BRCA1*, *GEN1X* y *MCC*. *BRCA1* es un gen supresor de tumores que participa en la reparación del ADN, el control del ciclo celular y mantenimiento de la estabilidad genómica. Mutaciones en este gen en la línea germinal predisponen a desarrollar cáncer de mama, ovario, próstata y páncreas (Antoniou et al., 2003; Al-sukhni et al., 2008). En el caso de *MCC*, se han descrito mutaciones somáticas en este gen en tumores colorrectales, con lo que ha sido propuesto como potencial gen supresor de tumores (Kinzler et al., 1991). Aún así, los programas *in silico* de predicción de función clasifican las mutaciones de cambio de sentido identificadas en nuestra familia para estos dos genes como benignas. Finalmente, el gen *GEN1X* codifica una proteína implicada en la reparación de uniones cruzadas entre hebras de ADN (en inglés, *DNA interstrand crosslinks*), formando parte de la vía de

reparación de la Anemia de Fanconi. Deficiencias en distintos componentes de esta vía se han relacionado con el cáncer familiar de mama, páncreas y ovario (Seal et al., 2006; Rahman et al., 2007; Jones et al., 2009a; Coulet et al., 2013).

Dada la naturaleza de la variante identificada en *GEN1X* en los tres individuos afectados de la familia, junto con su participación en la reparación del ADN, consideramos este gen como el mejor candidato para profundizar en su estudio. Aún así, no podemos descartar completamente el papel de las mutaciones identificadas en otros genes en el incremento a desarrollar cáncer en la familia.

5.2. Búsqueda de mutaciones en otras familias con CCR hereditario

Con el fin de valorar la implicación del gen en el CCR hereditario, analizamos los individuos índice de un total de 247 familias sin defectos en la maquinaria de reparación de bases desapareadas. De ellas, 176 cumplían los criterios de Ámsterdam I y II y las 71 restantes cumplían los criterios de Bethesda. Este segundo grupo estaba formado por familias que cumplían los siguientes requisitos: Presencia de dos familiares de primer grado con CCR u otros tumores relacionados, uno de ellos diagnosticado antes de los 50 años (criterios de Bethesda 1 y 4); o tres familiares afectados independientemente de la edad de diagnóstico (criterio de Bethesda 5). Se identificaron cuatro variantes adicionales en la serie de familias Ámsterdam positivas, y ninguna en el grupo de individuos que cumplían los criterios más laxos. Los cambios identificados corresponden a una segunda mutación truncante (c.2854C>T, p.R952*) y tres variantes de cambio de sentido (c.418G>T; p.D140Y, c.1018C>T; p.P340S y c.1771C>T; p.R591W).

5.3. Evaluación de las mutaciones encontradas en *GEN1X*

5.3.1. Frecuencias poblacionales

La primera mutación identificada mediante secuenciación de exomas, c.141C>A; p.C47*, no había sido descrita previamente en las bases de datos públicas (Exome Sequencing Project, 1000 genomes, dbSNP), ni tampoco se había hallado en 500 alelos de origen español (Quesada et al., 2011). En una primera evaluación del cambio, analizamos mediante la técnica del CSCE 288 controles y 286 casos de CCR recogidos consecutivamente de la misma región geográfica que la familia estudiada, sin identificar la mutación en ningún individuo.

De las variantes genéticas identificadas posteriormente en familias Ámsterdam positivas, dos no han sido descritas anteriormente y dos aparecen en las bases de datos públicas con una frecuencia menor del 1%. Por un lado, el cambio c.2854C>T ha sido previamente identificado en un individuo de origen colombiano (0,0458%, fuente 1000 Genomes), y la variante c.1771C>T en dos genotipos afro-americanos (0,0154%, fuente Exome Sequencing Project).

5.3.2. Estudio de segregación con la enfermedad

El análisis de cosegregación de las distintas mutaciones con el CCR en las familias respalda la patogenicidad de las variantes en cuatro de ellas (familias 1, 2, 3 y 5). Para el cambio c.1771C>T (familia 4) tan solo se ha podido examinar una hermana sana del individuo índice, que ha resultado ser no portadora de la mutación correspondiente. En el caso de las familias donde los datos de segregación son informativos, se han analizado entre dos y cuatro muestras, siendo todas de individuos afectados de CCR o cáncer de mama en un caso, y resultando portadores de la correspondiente mutación en *GEN1X*. Idealmente se debería analizar un número mayor de familiares para conseguir un estudio de cosegregación más completo, pero desgraciadamente este tipo de análisis está estrictamente condicionado a la disponibilidad de muestras y a la predisposición de las familias a colaborar en el estudio.

5.3.3. Evidencias de afectación de la función y posible patogenicidad de las variantes identificadas

Las mutaciones que crean un codón de parada, p.C47* y p.R952*, además de generar una proteína truncada, se sitúan ambas en dominios funcionales de la proteína (Figura 16). De hecho, el estudio de expresión de la proteína mediante Western Blot muestra que la expresión de la proteína en mucosa normal de un paciente portador de p.C47* es menor que en mucosa normal de un paciente control (no portador).

En cuanto a las mutaciones de cambio de sentido, el estudio *in silico* de predicción de función clasifica como deletéreo el cambio p.R591W con máximas puntuaciones por todos los programas utilizados. En el caso de p.D140Y, dos de los tres algoritmos utilizados la clasifican como deletérea y la variante p.P340S es clasificada como neutra por todos ellos. No se detectaron alteraciones en el *splicing* para ninguna de las tres variantes de cambio de sentido utilizando los métodos de predicción *in silico* habituales. Estos programas

generalmente consideran que las mutaciones que afectan al mecanismo de corte y empalme se sitúan en las inmediaciones de la unión intrón-exón. Sin embargo, recientemente se ha desarrollado un algoritmo de predicción que tiene en cuenta hexámeros de nucleótidos como elementos reguladores del *splicing* (Ke et al., 2011), del que se ha visto una buena correlación con datos experimentales (Di Giacomo et al., 2013). Utilizando este algoritmo de predicción, se obtiene un valor de -2,374 para el cambio c.418G>T (p.D140Y), altamente sugestivo de alteraciones en el mecanismo de corte y empalme para esta variante. Queda pendiente estudiar la presencia del correspondiente transcrito aberrante en ARN de los portadores de esta mutación.

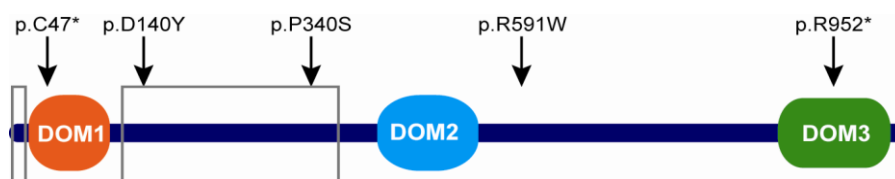


Figura 16. Dominios funcionales y regiones desordenadas de GEN1X. Las flechas indican la posición de las variantes identificadas y los recuadros grises las regiones desordenadas.

Las mutaciones patogénicas a menudo están situadas en zonas que son importantes para la estabilidad tridimensional (3D) de las estructuras proteicas. Realizamos un estudio de predicción del modelo 3D de la proteína para analizar la posible alteración de la estructura por las variantes de cambio de sentido identificadas. Esta proteína contiene tres dominios funcionales descritos, y el análisis de la secuencia mediante meta-servidores (DisMeta, Drip-Pred y MetaDisorder) concluyó que la proteína posee dos regiones desordenadas, la primera incluye los residuos 1-30 y la segunda 95-370. Estas regiones suelen adoptar una conformación más estable cuando la proteína interactúa con otros elementos, por lo que no se puede realizar la predicción del efecto de los cambios utilizando el modelo 3D de la proteína (Figura 16). Dos de las variantes de cambio de sentido identificadas (p.D140Y y p.P340S) están localizadas en una de estas zonas, por lo que no se puede predecir su efecto sobre la estabilidad de la proteína. En cambio el residuo R591 está situado en un lazo que conecta dos hélices α cerca de un dominio funcional (Figura 16). Esta secuencia hélice-lazo-hélice está altamente conservada en la evolución, y los diferentes modelos de predicción 3D para GEN1X concuerdan en que el residuo R591 siempre se localiza en el lazo expuesto al solvente, de manera que el cambio c.1771C>T desestabiliza la estructura helicoidal de este segmento de la proteína. Estas observaciones están en consonancia

con el análisis *in silico* de predicción de función, con resultados consistentes tan solo para la variante p.R591W.

5.3.4. Ensayo de viabilidad celular en respuesta a agentes de daño celular

Defectos en la vía de reparación de la anemia de Fanconi causan sensibilidad a agentes intercalantes del ADN como la MMC y el cisplatino (D'Andrea y Grompe, 2003). Puesto que nuestro gen de interés se ha relacionado con esta vía, realizamos un ensayo de viabilidad celular en respuesta a MMC utilizando linfocitos inmortalizados de un portador de la mutación c.141C>A (p.C47*) y como controles, un paciente con Anemia de Fanconi y mutación bialélica en *FANCA* (*FANCA*^{-/-}), un individuo sano y dos pacientes con mutaciones bialélicas en nuestro gen de interés (*GEN1X*^{-/-}). Este experimento mostró que células con un solo alelo funcional del gen presentan una sensibilidad mayor que células salvajes a concentraciones relativamente altas de MMC (10-70 nM), pero menor que aquellas con dos copias anómalas del gen. Estos resultados sugieren que defectos de una copia de *GEN1X* ya provocan un fallo en la reparación del ADN, siendo consistente con un modelo de haploinsuficiencia.

Para demostrar de manera definitiva que las mutaciones de cambio de sentido identificadas tienen un efecto funcional estamos desarrollando actualmente un sistema celular deficiente en nuestro gen de interés donde poder transfectar vectores con cada una de las mutaciones identificadas. Esto nos permitiría realizar ensayos *in vitro* de sensibilidad a la MMC y analizar si los distintos cambios tienen un efecto similar o menor que la delección completa del gen.

5.4. Papel del gen en la carcinogénesis colorrectal

5.4.1. Gen supresor de tumores: Búsqueda del segundo evento a nivel somático

En los síndromes de cáncer hereditario causados por mutaciones en un gen supresor de tumores se hereda una copia anómala del gen de uno de los progenitores, mientras que la otra copia se inactiva en el tejido diana (inactivación somática) (Knudson, 1976). En base a esta hipótesis, estudiamos los tumores de portadores heterocigotos de mutación en nuestro gen de interés en busca del evento que inactivase a nivel somático el segundo alelo del gen.

Para la familia portadora de la mutación truncante c.141C>A (familia 1) pudimos estudiar muestras de los tres individuos afectados. Analizando el ADN tumoral no se detectaron mutaciones adicionales en *GEN1X* en ninguno de los tumores. Tampoco se observó pérdida de heterocigosidad, determinada mediante el estudio de microsatélites cercanos al gen para los individuos II.6 y II.5, y con la secuenciación de exomas completos del tumor del probando. Además, considerando otros mecanismos de silenciamiento génico, como la hipermetilación del promotor, se estudió la expresión alelo-específica de *GEN1X* utilizando la técnica de SNaPshot, pero no se detectó pérdida de expresión del alelo salvaje en el ARN. De hecho, en el caso del afecto III.5 se observó una ratio entre tejido normal y tumoral consistente con la ganancia de una copia del alelo salvaje o pérdida del mutado. Tampoco se observó pérdida de expresión proteica de *GEN1X* en tejido tumoral del probando, al contrario, la proteína mostraba niveles mayores de expresión, consistente con la ganancia del cromosoma donde el gen está localizado, que ya habíamos observado gracias al estudio del tumor mediante arrays de número de copia (datos no mostrados) (T14, Bellido et al., 2014).

Igualmente, pudimos comprobar que los tumores pertenecientes a dos afectados portadores de la mutación c.418G>T en el gen de interés (familia 3) tampoco presentaban pérdida de heterocigosidad, es decir delección del alelo salvaje.

Estos resultados hacen pensar que *GEN1X* no actúa como un gen supresor de tumores clásico. Sugieren que existe un efecto dominante negativo de la mutación de manera que aún con un alelo salvaje del gen la función de la proteína en las células se podría ver afectada.

5.4.2. Gen implicado en la reparación del daño genético: ¿Acumulación de errores en el ADN?

Existen evidencias de la gran heterogeneidad mutacional que hay en el cáncer. Esta variabilidad incluye el tipo de cáncer, el espectro de mutaciones y la frecuencia mutacional a lo largo del genoma, resultado de la expresión génica y de los tiempos de replicación del ADN (Lawrence et al., 2013). De este modo, la prevalencia de mutaciones somáticas es muy variable entre tipos de cáncer y también dentro de ellos, moviéndose la tasa mutacional en un rango de 0,001 mutaciones por megabase (Mb) a más de 400 por Mb (Alexandrov et al., 2013). En el CCR hay casos ocupando toda la amplitud del espectro, aunque se puede diferenciar una mayoría (84%) con una tasa de mutación menor de 8,24

por Mb, de aquellos con tasas mayores de 12 mutaciones por Mb (16%), denominados hipermutados (Network Cancer Genome Atlas (TCGA), 2012). Gran parte de estos tumores se caracterizan por presentar niveles altos de IMS con frecuente silenciamiento epigenético de *MLH1*, pero una porción considerable (aproximadamente un 25%) y correspondiente a los tumores con mayor tasa mutacional, son portadores de mutaciones en otros genes MMR o en la subunidad catalítica de *POLE*.

La secuenciación de exomas completos del tumor del individuo índice de la familia portadora de la mutación en la línea germinal c.141C>A en *GEN1X*, permitió identificar un total de 236 mutaciones somáticas. Destaca la presencia de mutaciones en genes típicamente alterados en el CCR, tales como *APC*, *KRAS* o *PIK3CA*. La tasa mutacional en el tumor analizado es de 5 mutaciones por Mb, reduciéndose a 1,3 por Mb si se consideran solo los cambios no sinónimos, por lo que encaja dentro del rango observado para los tumores colorrectales no hipermutados (Network TCGA, 2012).

Existen seis tipos de sustitución de base nucleotídica: C:G>G:C, C:G>A:T, T:A>C:G, T:A>G:C, T:A>A:T, C:G>T:A. La diversidad y complejidad del desarrollo de la carcinogénesis conlleva a que se generen distintas firmas genéticas representativas del proceso mutacional somático en los distintos tipos de cánceres (Alexandrov et al., 2013). El análisis de los cambios identificados en el tumor estudiado reveló un patrón singular caracterizado por un exceso de las sustituciones T:A>G:C y C:G>G:C, que difiere de lo identificado en los CCR del TCGA, tanto hipermutados como no hipermutados. La frecuencia de ambas transversiones, 10,5% para T:A>G:C y 12,5% para C:G>G:C, supera en ambos casos el percentil 95 en CCR no hipermutados. Por otro lado, comparando el espectro mutacional asociado a nuestro gen de interés con los tumores colorrectales hipermutados, observamos que el cambio C:G>G:C sigue siendo más frecuente, pero que éstos también presentan un incremento similar en la frecuencia de la sustitución T:A>G:C (Figura 17). Esta mayor proporción de cambios T:A>G:C también se ha observado en los tumores hipermutantes desarrollados en el CCR asociado a la actividad reparadora de la polimerasa, causada por mutaciones germinales en *POLE* y *POLD1* (Briggs y Tomlinson, 2013). Estos resultados sugieren que la haploinsuficiencia de *GEN1X* puede causar un determinado tipo de errores debido a defectos en la reparación del ADN. No obstante, para validar esta hipótesis es necesario estudiar a nivel genético y genómico más tumores asociados a deficiencias en este gen. Si el espectro mutacional se confirma, estaríamos ante una nueva firma genética, resultado del proceso mutacional producto de los errores en la reparación del ADN mediada por *GEN1X*.

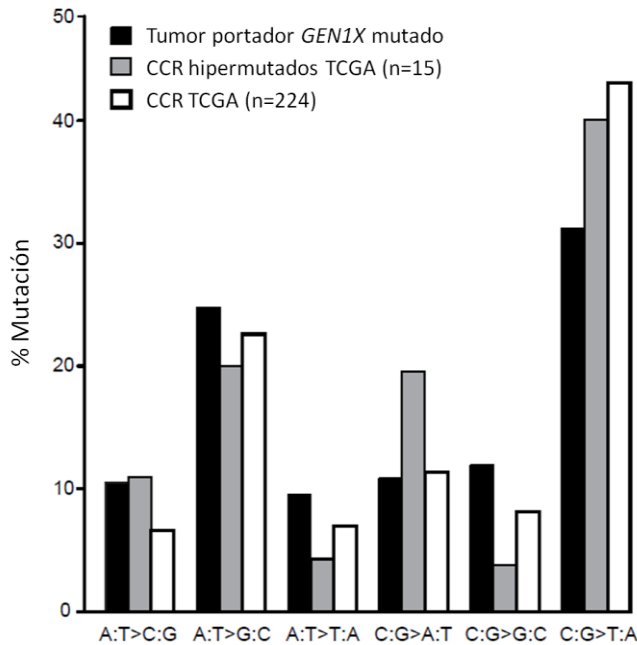


Figura 17. Comparación del espectro mutacional somático del tumor CCR asociado a *GEN1X* y los 239 tumores de CCR del TCGA. El tumor analizado corresponde al desarrollado por el portador de la mutación c.141C>A (p.C47*) en *GEN1X*. Los tumores del TCGA se han considerado separadamente los hipermutados de los no hipermutados.

5.4.3. Participación de *GEN1X* en la progresión tumoral

Además de su papel como gen de susceptibilidad, evaluamos la posible implicación de *GEN1X* como promotor de la carcinogénesis y de la progresión tumoral. Para ello realizamos el silenciamiento estable del gen mediante ARN de horquilla pequeña (shRNA, del inglés, *short hairpin RNA*) en la línea celular Caco-2. Éstas son células intestinales humanas procedentes de un cáncer de colon de un individuo caucásico. Se inyectaron subcutáneamente células Caco-2 con o sin expresión de *GEN1X* en ratones atímicos. Como no se observaron diferencias entre los tumores generados por un grupo de ratones o el otro, se implantaron ortotópicamente fragmentos equivalentes de los tumores en el ciego de nuevos ratones atímicos. Teóricamente estos tumores ya serían portadores de las mutaciones características de la deficiencia de *GEN1X*. La implantación ortotópica de tumores, directamente en el órgano de origen, se considera un modelo con mejor valor predictivo que los modelos subcutáneos convencionales, ya que estos tumores crecen en condiciones celulares y tisulares más similares al del tumor original, en este caso en el colon. Aunque sin alcanzar significancia estadística, observamos una tendencia de los

tumores derivados de Caco-2 con el gen de interés silenciado a ser más grandes que los del grupo control. Además, mientras que no se observó metástasis en ningún ejemplar control, se detectaron dos tumores macroscópicos en el diafragma de dos ratones del grupo con silenciamiento de *GEN1X*. Esto hace pensar que los tumores derivados de Caco-2 con deficiencias en esta proteína puedan ser ligeramente más agresivos que los del grupo control. A pesar de la tendencia observada, detectamos una considerable variabilidad en el tamaño de los tumores resultantes de los dos grupos experimentales (con o sin silenciamiento de *GEN1X*). Esto podría explicarse debido a la selección de diferentes sub-clones celulares en cada tumor. De hecho, hemos corroborado esta hipótesis tras secuenciar los exomas de dos tumores provenientes de líneas con el gen silenciado y dos tumores provenientes de líneas con el gen sin silenciar.

Además, dada la naturaleza celular de Caco-2, es posible que el potencial carcinogénico basal de esta línea celular sea más intenso que las capacidades tumorigénicas generadas con la pérdida de *GEN1X*. Para evitar este posible factor de confusión, cabría repetir el experimento con el silenciamiento del gen en una línea celular normal, sin propiedades carcinogénicas intrínsecas

5.4.4. *GEN1X* como gen de alta penetrancia en el cáncer colorrectal

Los hallazgos presentados en esta tesis doctoral señalan que mutaciones en *GEN1X* pueden estar implicadas en la susceptibilidad al CCR hereditario, por lo que debería tenerse en cuenta y ser integrado en los programas de consejo genético en familias con CCR no polipósico.

El hecho de que se hayan detectado mutaciones en este gen en familias que cumplen los criterios más estrictos de agregación familiar, los criterios de Ámsterdam, pero no los de Bethesda, sugiere que éste es un gen de alta penetrancia. Según los resultados obtenidos en nuestra serie de muestras, con 176 familias analizadas de diferentes zonas geográficas de la península Ibérica, mutaciones patogénicas en *GEN1X* podrían explicar cerca del 3% de las familias Ámsterdam positivas sin causa genética conocida. Sin embargo será necesario analizar un número mayor de casos para determinar con mayor precisión la prevalencia y penetrancia de las mutaciones en este gen en el CCR hereditario.

6. IDENTIFICACIÓN DE LAS CAUSAS GENÉTICAS DEL CÁNCER COLORRECTAL FAMILIAR TIPO X

A pesar del conocimiento acumulado en las últimas décadas, el CCRf-X junto con la poliposis serrada agrupan la mayor parte de casos de CCR familiar cuya etiología genética permanece desconocida. Determinar la causa genética subyacente en estas familias es de suma importancia, ya que permitiría distinguir aquellos individuos de las familias con mayor riesgo a desarrollar cáncer de los que presentan un riesgo similar al de la población general. De esta forma, los individuos a riesgo podrían someterse a una vigilancia periódica y beneficiarse de la detección precoz de posibles tumores, cuya eficacia ha sido probada para disminuir considerablemente la incidencia y la mortalidad del CCR (Levin et al., 2008). Asimismo, los familiares sin la alteración genética subyacente dejarían de someterse a una vigilancia médica tan intensa, y a menudo invasiva, como los portadores de mutación en un gen de alta penetrancia.

Tan solo un 2-3% de los casos diagnosticados de CCR cumplen los criterios de Ámsterdam, de los que al menos la mitad corresponden a casos con SL. Prácticamente el resto, aproximadamente el 40% de los casos que cumplen los criterios de Ámsterdam, se agrupan dentro del CCRf-X. Dada la baja incidencia del CCRf-X, resulta complicado reunir un número elevado de casos, en ocasiones necesario para poder llegar a conclusiones consistentes. En nuestro estudio, gracias a la colaboración con otros centros del país se ha conseguido una colección de muestras que agrupa 176 familias con CCRf-X.

El CCRf-X es una entidad heterogénea. Recordemos que incluye aquellos casos que comparten unas mismas características clínicas, pero que pueden enmascarar bases genéticas muy diferentes. Dilucidar esta forma de CCR incluye la paradoja de la naturaleza monogénica o poligénica de la enfermedad, pudiendo estar implicadas una sola mutación de alta penetrancia o múltiples variantes de bajo o moderado riesgo. El enfoque inicial del que se parte condicionará el planteamiento del estudio a realizar. El trabajo realizado en esta tesis para la búsqueda de nuevos genes de susceptibilidad en el CCRf-X ha tratado la enfermedad bajo un enfoque monogénico. Además de nuestros resultados que implican a *GEN1X* como gen causal en un número de familias, recientemente se han identificado nuevos alelos poco frecuentes de susceptibilidad para el CCR familiar en otros genes, tales como *OGG1*, *PTPRJ*, *BUB1* y *BUB3*, entre otros (Garre et al., 2011; Venkatachalam et al., 2010; de Voer et al., 2011). Sería interesante

estudiar estos genes en nuestras familias diagnosticadas con CCRf-X, ya que podrían explicar el riesgo incrementado en alguna de ellas.

En la misma línea, se ha de considerar que la asociación genotipo-fenotipo no siempre es directa. Puede que algunas de las familias clasificadas como CCRf-X estén causadas por defectos en genes típicos de CCR, pero que no presentan el fenotipo habitual para alteraciones de ese gen. Este es el caso de una de las familias que hemos estudiado en nuestro laboratorio y que constituye parte del trabajo presentado en esta tesis, donde se han identificado afectos con mutaciones monoalélicas y bialélicas en *MUTYH*. De igual modo, en otro estudio se han descrito mutaciones en *BMPR1A*, gen clásicamente asociado a la poliposis juvenil, en una familia con CCRf-X (Nieminen et al., 2011). El análisis generalizado de todos los genes de riesgo de CCR ayudará a detectar estos casos con fenotipos que difieren del clásicamente asociado a mutaciones en un gen determinado. Esto comienza a ser actualmente una realidad gracias al estudio completo de paneles de genes de susceptibilidad al cáncer utilizado técnicas que combinan la captura de regiones del genoma y la secuenciación paralela masiva.

En el otro extremo, existe la posibilidad de que algunos casos de CCRf-X concuerden con un modelo poligénico de la enfermedad, con la posible participación del ambiente o estilo de vida, lo que hace más complejo su estudio. Aunque generalmente se ha considerado que los alelos de bajo riesgo desempeñan un papel relevante en las formas de cáncer esporádico, también se ha propuesto su participación en el cáncer familiar. Se ha sugerido que la agrupación de variantes de baja penetrancia pueden explicar parte del exceso de riesgo observado en familias con CCR (Middeldorp et al., 2009; Niittymäki et al., 2010). Actualmente se han descrito cerca de 50 variantes de riesgo a CCR (<http://www.genome.gov/gwastudies>), cuya implicación en la agregación de cáncer en las familias X sería interesante evaluar.

Finalmente, también existe la opción de que algunas familias con CCRf-X hayan sido clasificadas como tal debido a la coexistencia de múltiples casos esporádicos, agregados en la familia como resultado de un ambiente compartido, el estilo de vida o incluso simplemente por azar.

CONCLUSIONES

- Los pacientes con CCR hereditario no polipósico muestran un patrón de longitud telomérica en sangre distinto según el estado de la maquinaria de reparación de bases desapareadas.
- El acortamiento telomérico podría actuar como factor de riesgo a desarrollar cáncer en el SL, mientras que en el CCRf-X lo haría una longitud mayor de los telómeros.
- Descartamos el acortamiento de los telómeros como mecanismo molecular que explique el fenómeno de la anticipación genética observada en el SL.
- Datos preliminares sugieren que el acortamiento telomérico podría tener relación con la dependencia del riesgo a desarrollar cáncer según origen materno-paterno de la mutación causal en el SL.
- Mutaciones en *GALNT12* no contribuyen al CCRf-X, sugiriendo que no es un gen de alta penetrancia para el CCR.
- La secuenciación de exomas es una técnica útil para la identificación del gen causal de la predisposición al cáncer en el CCRf-X.
- En el diagnóstico genético de la poliposis asociada a *MUTYH* se deben considerar fenotipos clínicos que no cumplen los criterios establecidos, incluyendo: Herencias aparentemente dominantes debido al riesgo que aportan las variantes monoalélicas en *MUTYH*, ausencia de pólipos, o presencia de tumores con inestabilidad de microsatélites. El análisis universal de la mutación *KRAS* c.34G>T (p.G12C) en los tumores colorectales o la utilización de paneles completos de genes en el diagnóstico genético de rutina del cáncer familiar ayudarán a la identificación de familias *MUTYH* con fenotipos atípicos.
- La variante c.43A>G (p.Met15Val) en *MUTYH* es una mutación patogénica que actúa a través de la inactivación del codón de inicio de la traducción de las isoformas nucleares (β y γ), especialmente relevantes en el epitelio colorrectal.
- Hemos identificado un nuevo gen de predisposición al CCR que podría explicar aproximadamente un 3% de las familias que cumplen los criterios de Ámsterdam sin defectos en maquinaria de reparación de bases desapareadas del ADN. La haploinsuficiencia de este gen podría causar una reparación defectuosa del ADN con la consiguiente acumulación de un tipo específico de errores.

BIBLIOGRAFÍA

A

- Abdel-Rahman WM, Ollikainen M, Kariola R, Jarvinen HJ, Mecklin JP, Nystrom-Lahti M, Knuutila S, Peltomaki P. Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. *Oncogene* 2005;24:1542-51.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Illicic T, Imbeaud S, Imielinski M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichten P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. Signatures of mutational processes in human cancer. *Nature* 2013;500:415-21.
- Ali M, Kim H, Cleary S, Cupples C, Gallinger S, Bristow R. Characterization of mutant MUTYH proteins associated with familial colorectal cancer. *Gastroenterology* 2008;135:499-507.
- Al-Sukhni W, Rothenmund H, Borgida AE, Zogopoulos G, O'Shea AM, Pollett A, Gallinger S. Germline BRCA1 mutations predispose to pancreatic adenocarcinoma. *Hum Genet* 2008;124:271-8.
- Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* 2002;30:227-32.
- Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjakoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum*

- Genet 2003;72:1117-30.
- Artandi SE, Chang S, Lee SL, Alson S, Gottlieb GJ, Chin L, DePinho RA. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* 2000;406:641-5.
- Attardi LD. The role of p53-mediated apoptosis as a crucial anti-tumor response to genomic instability: lessons from mouse models. *Mutat Res* 2005;569:145-57.
- Avezzu A, Agostini M, Pucciarelli S, Lise M, Urso ED, Mammi I, Maretto I, Enzo MV, Pastrello C, Lise M, Nitti D, Viel A. The role of MYH gene in genetic predisposition to colorectal cancer: another piece of the puzzle. *Cancer Lett* 2008;268:308-13.
- Aviv A. Telomeres and human somatic fitness. *J Gerontol A Biol Sci Med Sci* 2006;61:871-3.
- Aviv A, Hunt SC, Lin J, Cao X, Kimura M, Blackburn E. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. *Nucleic Acids Res* 2011;39:e134.

B

- Baglietto L, Lindor NM, Dowty JG, White DM, Wagner A, Gomez Garcia EB, Vriends AH, Cartwright NR, Barnetson RA, Farrington SM, Tenesa A, Hampel H, Buchanan D, Arnold S, Young J, Walsh MD, Jass J, Macrae F, Antill Y, Winship IM, Giles GG, Goldblatt J, Parry S, Suthers G, Leggett B, Butz M, Aronson M, Poynter JN, Baron JA, Le Marchand L, Haile R, Gallinger S, Hopper JL, Potter J, de la Chapelle A, Vasen HF, Dunlop MG, Thibodeau SN, Jenkins MA. Risks of Lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst* 2010;102:193-201.
- Bagnardi V, Blangiardo M, La Vecchia C, Corrao G. A meta-analysis of alcohol drinking and cancer risk. *Br J Cancer* 2001;85:1700-5.
- Balaguer F, Castellvi-Bel S, Castells A, Andreu M, Munoz J, Gisbert JP, Llor X, Jover R, de Cid R, Gonzalo V, Bessa X, Xicola RM, Pons E, Alenda C, Paya A, Pique JM. Identification of MYH mutation carriers in colorectal cancer: a multicenter, case-control, population-based study. *Clin Gastroenterol Hepatol* 2007;5:379-87.
- Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. *Gut* 2013;62:933-47.
- Beeharry N, Broccoli D. Telomere dynamics in response to chemotherapy. *Curr Mol Med* 2005;5:187-96.
- Bellido F, Guino E, Jagmohan-Changur S, Segui N, Pineda M, Navarro M, Lazaro C, Blanco I, Vasen HF, Moreno V, Capella G, Wijnen JT, Valle L. Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome. *Eur J Hum Genet*

- 2013;21:511-6.
- Bellido F, Pineda M, Sanz-Pamplona R, Navarro M, Nadal M, Lazaro C, Blanco I, Moreno V, Capella G, Valle L. Comprehensive molecular characterization of hereditary non-polyposis colorectal tumors with mismatch repair proficiency. 2014; en prensa.
- Biegler KA, Anderson AK, Wenzel LB, Osann K, Nelson EL. Longitudinal change in telomere length and the chronic stress response in a randomized pilot biobehavioral clinical study: implications for cancer prevention. *Cancer Prev Res (Phila)* 2012;5:1173-82.
- Bisgaard ML, Fenger K, Bulow S, Niebuhr E, Mohr J. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. *Hum Mutat* 1994;3:121-5.
- Bisoffi M, Heaphy CM, Griffith JK. Telomeres: prognostic markers for solid tumors. *Int J Cancer* 2006;119:2255-60.
- Blackburn EH. Structure and function of telomeres. *Nature* 1991;350:569-73.
- Blasco MA. Mice with bad ends: mouse models for the study of telomeres and telomerase in cancer and aging. *Embo J* 2005;24:1095-103.
- Boardman LA, Litzelman K, Seo S, Johnson RA, Vanderboom RJ, Kimmel GW, Cunningham JM, Gangnon RE, Engelman CD, Riegert-Johnson DL, Potter J, Haile R, Buchanan D, Jenkins MA, Rider DN, Thibodeau SN, Petersen GM, Skinner HG. The association of telomere length with colorectal cancer differs by the age of cancer onset. *Clin Transl Gastroenterol* 2014;5:e52.
- Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 1998;58:5248-57.
- Boonstra PS, Gruber SB, Raymond VM, Huang SC, Timshel S, Nilbert M, Mukherjee B. A review of statistical methods for testing genetic anticipation: looking for an answer in Lynch syndrome. *Genet Epidemiol* 2010;34:756-68.
- Boonstra PS, Mukherjee B, Taylor JM, Nilbert M, Moreno V, Gruber SB. Bayesian modeling for genetic anticipation in presence of mutational heterogeneity: a case study in Lynch syndrome. *Biometrics* 2011;67:1627-37.
- Bougeard G, Baert-Desurmont S, Tournier I, Vasseur S, Martin C, Brugieres L, Chompret A, Bressac-de Paillerets B, Stoppa-Lyonnet D, Bonaiti-Pellie C, Frebourg T. Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome. *J Med Genet* 2006;43:531-3.

- Boyle P, Langman JS. ABC of colorectal cancer: Epidemiology. *Bmj* 2000;321:805-8.
- Bozzao C, Lastella P, Ponz de Leon M, Pedroni M, Di Gregorio C, D'Ovidio FD, Resta N, Prete F, Guanti G, Stella A. Analysis of telomere dynamics in peripheral blood cells from patients with Lynch syndrome. *Cancer* 2011 a;117:4325-35.
- Bozzao C, Lastella P, Stella A. Anticipation in lynch syndrome: where we are where we go. *Curr Genomics* 2011 b;12:451-65.
- Brand R, Nielsen M, Lynch H, E. I. MUTYH-Associated Polyposis. In: Pagon RA AM, Bird TD, et al., ed. *GeneReviews™* [Internet]. Seattle (WA): University of Washington, Seattle, 2012.
- Briggs S, Tomlinson I. Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. *J Pathol* 2013;230:148-53.
- Brockhausen I. Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta* 1999;1473:67-95.
- Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep* 2006;7:599-604.
- Burn J, Gerdes AM, Macrae F, Mecklin JP, Moeslein G, Olschwang S, Eccles D, Evans DG, Maher ER, Bertario L, Bisgaard ML, Dunlop MG, Ho JW, Hodgson SV, Lindblom A, Lubinski J, Morrison PJ, Murday V, Ramesar R, Side L, Scott RJ, Thomas HJ, Vasen HF, Barker G, Crawford G, Elliott F, Movahedi M, Pylvanainen K, Wijnen JT, Fodde R, Lynch HT, Mathers JC, Bishop DT. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet* 2011;378:2081-7.

C

- Capezzone M, Cantara S, Marchisotta S, Filetti S, De Santi MM, Rossi B, Ronga G, Durante C, Pacini F. Short telomeres, telomerase reverse transcriptase gene amplification, and increased telomerase activity in the blood of familial papillary thyroid cancer patients. *J Clin Endocrinol Metab* 2008;93:3950-7.
- Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 2009;37:e21.
- Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* 2010;11:319-30.
- Chan AT, Giovannucci EL. Primary prevention of colorectal cancer. *Gastroenterology* 2010;138:2029-2043 e10.

- Chan PA, Duraisamy S, Miller PJ, Newell JA, McBride C, Bond JP, Raevaara T, Ollila S, Nystrom M, Grimm AJ, Christodoulou J, Oetting WS, Greenblatt MS. Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat* 2007;28:683-93.
- Cherkas LF, Hunkin JL, Kato BS, Richards JB, Gardner JP, Surdulescu GL, Kimura M, Lu X, Spector TD, Aviv A. The association between physical activity in leisure time and leukocyte telomere length. *Arch Intern Med* 2008;168:154-8.
- Church DN, Briggs SE, Palles C, Domingo E, Kearsley SJ, Grimes JM, Gorman M, Martin L, Howarth KM, Hodgson SV, Kaur K, Taylor J, Tomlinson IP. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* 2013;22:2820-8.
- Clarke E, Green RC, Green JS, Mahoney K, Parfrey PS, Younghusband HB, Woods MO. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. *Hum Mutat* 2012;33:1056-8.
- Cleary SP, Cotterchio M, Jenkins MA, Kim H, Bristow R, Green R, Haile R, Hopper JL, LeMarchand L, Lindor N, Parfrey P, Potter J, Younghusband B, Gallinger S. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. *Gastroenterology* 2009;136:1251-60.
- Clendenning M, Baze ME, Sun S, Walsh K, Liyanarachchi S, Fix D, Schunemann V, Comeras I, Deacon M, Lynch JF, Gong G, Thomas BC, Thibodeau SN, Lynch HT, Hampel H, de la Chapelle A. Origins and prevalence of the American Founder Mutation of MSH2. *Cancer Res* 2008;68:2145-53.
- Coissieux MM, Tomsic J, Castets M, Hampel H, Tuupanen S, Andrieu N, Comeras I, Drouet Y, Lasset C, Liyanarachchi S, Mazelin L, Puisieux A, Saurin JC, Scoazec JY, Wang Q, Aaltonen L, Tanner SM, de la Chapelle A, Bernet A, Mehlen P. Variants in the netrin-1 receptor UNC5C prevent apoptosis and increase risk of familial colorectal cancer. *Gastroenterology* 2011;141:2039-46.
- Comino-Mendez I, Gracia-Aznarez FJ, Schiavi F, Landa I, Leandro-Garcia LJ, Leton R, Honrado E, Ramos-Medina R, Caronia D, Pita G, Gomez-Grana A, de Cubas AA, Inglada-Perez L, Maliszewska A, Taschin E, Bobisse S, Pica G, Loli P, Hernandez-Lavado R, Diaz JA, Gomez-Morales M, Gonzalez-Neira A, Roncador G, Rodriguez-Antona C, Benitez J, Mannelli M, Opocher G, Robledo M, Cascon A. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet* 2011;43:663-7.

- Coolbaugh-Murphy MI, Xu JP, Ramagli LS, Ramagli BC, Brown BW, Lynch PM, Hamilton SR, Frazier ML, Siciliano MJ. Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. *Hum Mutat* 2010;31:317-24.
- Coulet F, Fajac A, Colas C, Eyries M, Dion-Miniere A, Rouzier R, Uzan S, Lefranc JP, Carbonnel M, Cornelis F, Cortez A, Soubrier F. Germline RAD51C mutations in ovarian cancer susceptibility. *Clin Genet* 2013;83:332-6.
- Croitoru ME, Cleary SP, Di Nicola N, Manno M, Selander T, Aronson M, Redston M, Cotterchio M, Knight J, Gryfe R, Gallinger S. Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *J Natl Cancer Inst* 2004;96:1631-4.
- Cui Y, Cai Q, Qu S, Chow WH, Wen W, Xiang YB, Wu J, Rothman N, Yang G, Shu XO, Gao YT, Zheng W. Association of leukocyte telomere length with colorectal cancer risk: nested case-control findings from the Shanghai Women's Health Study. *Cancer Epidemiol Biomarkers Prev* 2012;21:1807-13.
- Cunningham JM, Johnson RA, Litzelman K, Skinner HG, Seo S, Engelman CD, Vanderboom RJ, Kimmel GW, Gangnon RE, Riegert-Johnson DL, Baron JA, Potter JD, Haile R, Buchanan DD, Jenkins MA, Rider DN, Thibodeau SN, Petersen GM, Boardman LA. Telomere length varies by DNA extraction method: implications for epidemiologic research. *Cancer Epidemiol Biomarkers Prev* 2013;22:2047-54.

D

- D'Andrea AD, Grompe M. The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer* 2003;3:23-34.
- Davies RJ, Miller R, Coleman N. Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer* 2005;5:199-209.
- de la Chapelle A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer* 2004;4:769-80.
- de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005;19:2100-10.
- de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ, Smeets D, Fu L, Vreede L, Kamping EJ, Verwiel ET, Hahn MM, Ariaans M, Spruijt L, van Essen T, Houge G, Schackert HK, Sheng JQ, Venselaar H, van Ravenswaaij-Arts CM, van Krieken JH, Hoogerbrugge N, Kuiper RP. Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer.

- Gastroenterology 2013;145:544-7.
- de Voer RM, Hoogerbrugge N, Kuiper RP. Spindle-assembly checkpoint and gastrointestinal cancer. *N Engl J Med* 2011;364:1279-80.
- de Vos tot Nederveen Cappel WH, Nagengast FM, Griffioen G, Menko FH, Taal BG, Kleibeuker JH, Vasen HF. Surveillance for hereditary nonpolyposis colorectal cancer: a long-term study on 114 families. *Dis Colon Rectum* 2002;45:1588-94.
- Desai DC, Lockman JC, Chadwick RB, Gao X, Percesepe A, Evans DG, Miyaki M, Yuen ST, Radice P, Maher ER, Wright FA, de La Chapelle A. Recurrent germline mutation in MSH2 arises frequently de novo. *J Med Genet* 2000;37:646-52.
- Di Giacomo D, Gaildrat P, Abuli A, Abdat J, Frebourg T, Tosi M, Martins A. Functional analysis of a large set of BRCA2 exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum Mutat* 2013;34:1547-57.
- Ding Z, Wu CJ, Jaskelioff M, Ivanova E, Kost-Alimova M, Protopopov A, Chu GC, Wang G, Lu X, Labrot ES, Hu J, Wang W, Xiao Y, Zhang H, Zhang J, Zhang J, Gan B, Perry SR, Jiang S, Li L, Horner JW, Wang YA, Chin L, DePinho RA. Telomerase reactivation following telomere dysfunction yields murine prostate tumors with bone metastases. *Cell* 2012;148:896-907.
- Djureinovic T, Skoglund J, Vandrovcova J, Zhou XL, Kalushkova A, Iselius L, Lindblom A. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. *Gut* 2006;55:362-6.
- Dominguez-Valentin M, Therkildsen C, Veerla S, Jonsson M, Bernstein I, Borg A, Nilbert M. Distinct gene expression signatures in lynch syndrome and familial colorectal cancer type x. *PLoS One* 2013;8:e71755.
- Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, Liu B, Kinzler KW, Vogelstein B. Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum Mol Genet* 1997;6:105-10.

E

- Engel C, Rahner N, Schulmann K, Holinski-Feder E, Goecke TO, Schackert HK, Kloor M, Steinke V, Vogelsang H, Moslein G, Gorgens H, Dechant S, von Knebel Doeberitz M, Ruschoff J, Friedrichs N, Buttner R, Loeffler M, Propping P, Schmiegel W. Efficacy of annual colonoscopic surveillance in individuals with hereditary nonpolyposis colorectal cancer. *Clin Gastroenterol Hepatol* 2010;8:174-82.

F

- Farrington SM, Tenesa A, Barnetson R, Wiltshire A, Prendergast J, Porteous M, Campbell H, Dunlop MG. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* 2005;77:112-9.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759-67.
- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893-917.
- Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;49:1374-403.
- Francisco I, Albuquerque C, Lage P, Belo H, Vitoriano I, Filipe B, Claro I, Ferreira S, Rodrigues P, Chaves P, Leitao CN, Pereira AD. Familial colorectal cancer type X syndrome: two distinct molecular entities? *Fam Cancer* 2011;10:623-31.
- Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 1986;323:643-6.

G

- Gallinger S, Aronson M, Shayan K, Ratcliffe EM, Gerstle JT, Parkin PC, Rothenmund H, Croitoru M, Baumann E, Durie PR, Weksberg R, Pollett A, Riddell RH, Ngan BY, Cutz E, Lagarde AE, Chan HS. Gastrointestinal cancers and neurofibromatosis type 1 features in children with a germline homozygous MLH1 mutation. *Gastroenterology* 2004;126:576-85.
- Garcia-Aranda C, de Juan C, Diaz-Lopez A, Sanchez-Pernaute A, Torres AJ, Diaz-Rubio E, Balibrea JL, Benito M, Iniesta P. Correlations of telomere length, telomerase activity, and telomeric-repeat binding factor 1 expression in colorectal carcinoma. *Cancer* 2006;106:541-51.
- Garre P, Briceno V, Xicola RM, Doyle BJ, de la Hoya M, Sanz J, Llovet P, Pescador P, Puente J, Diaz-Rubio E, Llor X, Caldes T. Analysis of the oxidative damage repair genes NUDT1, OGG1, and MUTYH in patients from mismatch repair proficient HNPCC families (MSS-HNPCC). *Clin Cancer Res* 2011;17:1701-12.
- Gerken TA, Jamison O, Perrine CL, Collette JC, Moinova H, Ravi L, Markowitz SD, Shen W, Patel H, Tabak LA. Emerging paradigms for the initiation of mucin-type protein

- O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J Biol Chem* 2011;286:14493-507.
- Gertler R, Rosenberg R, Stricker D, Friederichs J, Hoos A, Werner M, Ulm K, Holzmann B, Nekarda H, Siewert JR. Telomere length and human telomerase reverse transcriptase expression as markers for progression and prognosis of colorectal carcinoma. *J Clin Oncol* 2004;22:1807-14.
- Giardiello FM, Brensinger JD, Petersen GM. AGA technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology* 2001;121:198-213.
- Gibson TM, Wang SS, Cerhan JR, Maurer MJ, Hartge P, Habermann TM, Davis S, Cozen W, Lynch CF, Severson RK, Rothman N, Chanock SJ, Morton LM. Inherited genetic variation and overall survival following follicular lymphoma. *Am J Hematol* 2012;87:724-6.
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011;12:228.
- Goel A, Xicola RM, Nguyen TP, Doyle BJ, Sohn VR, Bandipalliam P, Rozek LS, Reyes J, Cordero C, Balaguer F, Castells A, Jover R, Andreu M, Syngal S, Boland CR, Llor X. Aberrant DNA methylation in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 2010;138:1854-62.
- Goldberg JM, Piver MS, Jishi MF, Blumenson L. Age at onset of ovarian cancer in women with a strong family history of ovarian cancer. *Gynecol Oncol* 1997;66:3-9.
- Goldstein AM, Clark WH, Jr., Fraser MC, Tucker MA. Apparent anticipation in familial melanoma. *Melanoma Res* 1996;6:441-6.
- Gonzalez-Suarez E, Samper E, Flores JM, Blasco MA. Telomerase-deficient mice with short telomeres are resistant to skin tumorigenesis. *Nat Genet* 2000;26:114-7.
- Gramatges MM, Telli ML, Balise R, Ford JM. Longer relative telomere length in blood from women with sporadic and familial breast cancer compared with healthy controls. *Cancer Epidemiol Biomarkers Prev* 2010;19:605-13.
- Gray-McGuire C, Guda K, Adrianto I, Lin CP, Natale L, Potter JD, Newcomb P, Poole EM, Ulrich CM, Lindor N, Goode EL, Fridley BL, Jenkins R, Le Marchand L, Casey G, Haile R, Hopper J, Jenkins M, Young J, Buchanan D, Gallinger S, Adams M, Lewis S, Willis J, Elston R, Markowitz SD, Wiesner GL. Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. *Cancer Res* 2010;70:5409-18.
- Gruber SB, Mukherjee B. Anticipation in lynch syndrome: still waiting for the answer. *J Clin Oncol* 2009;27:326-7.

- Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, Parnigiani G, Barnholtz-Sloan J, Dawson DW, Velculescu VE, Kinzler KW, Papadopoulos N, Vogelstein B, Willis J, Gerken TA, Markowitz SD. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci U S A* 2009;106:12921-5.
- Guo JM, Chen HL, Wang GM, Zhang YK, Narimatsu H. Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-12 in gastric and colonic cancer cell lines and in human colorectal cancer. *Oncology* 2004;67:271-6.
- Guo JM, Zhang Y, Cheng L, Iwasaki H, Wang H, Kubota T, Tachibana K, Narimatsu H. Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12. *FEBS Lett* 2002;524:211-8.
- Gylfe AE, Sirkia J, Ahlsten M, Jarvinen H, Mecklin JP, Karhu A, Aaltonen LA. Somatic mutations and germline sequence variants in patients with familial colorectal cancer. *Int J Cancer* 2010;127:2974-80.

H

- Hamilton SR, Liu B, Parsons RE, Papadopoulos N, Jen J, Powell SM, Krush AJ, Berk T, Cohen Z, Tetu B, et al. The molecular basis of Turcot's syndrome. *N Engl J Med* 1995;332:839-47.
- 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. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J Clin Oncol* 2008;26:5783-8.
- Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100:57-70.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646-74.
- Hansen H, Lemke H, Bredfeldt G, Konnecke I, Havsteen B. The Hodgkin-associated Ki-1 antigen exists in an intracellular and a membrane-bound form. *Biol Chem Hoppe Seyler* 1989;370:409-16.
- Harley CB. Telomere loss: mitotic clock or genetic time bomb? *Mutat Res* 1991;256:271-82.
- Hastie ND, Dempster M, Dunlop MG, Thompson AM, Green DK, Allshire RC. Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 1990;346:866-8.

- Heald B, Plesec T, Liu X, Pai R, Patil D, Moline J, Sharp RR, Burke CA, Kalady MF, Church J, Eng C. Implementation of universal microsatellite instability and immunohistochemistry screening for diagnosing lynch syndrome in a large academic medical center. *J Clin Oncol* 2013;31:1336-40.
- Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G, Warren W, Aminoff M, Hoglund P, Jarvinen H, Kristo P, Pelin K, Ridanpaa M, Salovaara R, Toro T, Bodmer W, Olschwang S, Olsen AS, Stratton MR, de la Chapelle A, Aaltonen LA. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* 1998;391:184-7.
- Hes FJ, Nielsen M, Bik EC, Konvalinka D, Wijnen JT, Bakker E, Vasen HF, Breuning MH, Tops CM. Somatic APC mosaicism: an underestimated cause of polyposis coli. *Gut* 2008;57:71-6.
- Horwitz M, Goode EL, Jarvik GP. Anticipation in familial leukemia. *Am J Hum Genet* 1996;59:990-8.
- Hou L, Zhang X, Gawron AJ, Liu J. Surrogate tissue telomere length and cancer risk: shorter or longer? *Cancer Lett* 2012;319:130-5.
- Houlston RS. COGENT (COlorectal cancer GENEtics) revisited. *Mutagenesis* 2012;27:143-51.
- 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. The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations. *J Med Genet* 2004;41:484-91.
- Hsu L, Zhao LP, Malone KE, Daling JR. Assessing changes in ages at onset over successive generation: an application to breast cancer. *Genet Epidemiol* 2000;18:17-32.
- Hussain MR, Nasir J, Al-Aama JY. Clinically significant missense variants in human GALNT3, GALNT8, GALNT12, and GALNT13 genes: intriguing in silico findings. *J Cell Biochem* 2014;115:313-27.
- Hutter RV, Sobin LH. A universal staging system for cancer of the colon and rectum. Let there be light. *Arch Pathol Lab Med* 1986;110:367-8.
- I
- Iwama H, Ohyashiki K, Ohyashiki JH, Hayashi S, Yahata N, Ando K, Toyama K, Hoshika A, Takasaki M, Mori M, Shay JW. Telomeric length and telomerase activity vary with age in peripheral blood cells obtained from normal individuals. *Hum Genet*

1998;102:397-402.

J

Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinemann K, Howarth K, East J, Taylor J, Thomas H, Tomlinson I. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat Genet* 2012;44:699-703.

Jaeger EE, Woodford-Richens KL, Lockett M, Rowan AJ, Sawyer EJ, Heinemann K, Rozen P, Murday VA, Whitelaw SC, Ginsberg A, Atkin WS, Lynch HT, Southey MC, Debinski H, Eng C, Bodmer WF, Talbot IC, Hodgson SV, Thomas HJ, Tomlinson IP. An ancestral Ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. *Am J Hum Genet* 2003;72:1261-7.

Jarvinen HJ, Mecklin JP, Sistonen P. Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 1995;108:1405-11.

Jarvinen HJ, Renkonen-Sinisalo L, Aktan-Collan K, Peltomaki P, Aaltonen LA, Mecklin JP. Ten years after mutation testing for Lynch syndrome: cancer incidence and outcome in mutation-positive and mutation-negative family members. *J Clin Oncol* 2009;27:4793-7.

Jass JR. Hereditary Non-Polyposis Colorectal Cancer: the rise and fall of a confusing term. *World J Gastroenterol* 2006;12:4943-50.

Jass JR, Cottier DS, Jeevaratnam P, Pokos V, Holdaway KM, Bowden ML, Van de Water NS, Browett PJ. Diagnostic use of microsatellite instability in hereditary non-polyposis colorectal cancer. *Lancet* 1995;346:1200-1.

Jenkins MA, Croitoru ME, Monga N, Cleary SP, Cotterchio M, Hopper JL, Gallinger S. Risk of colorectal cancer in monoallelic and biallelic carriers of MYH mutations: a population-based case-family study. *Cancer Epidemiol Biomarkers Prev* 2006;15:312-4.

Jenne DE, Reimann H, Nezu J, Friedel W, Loff S, Jeschke R, Muller O, Back W, Zimmer M. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* 1998;18:38-43.

Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;96:2992-3003.

- Jones AM, Beggs AD, Carvajal-Carmona L, Farrington S, Tenesa A, Walker M, Howarth K, Ballereau S, Hodgson SV, Zauber A, Bertagnoli M, Midgley R, Campbell H, Kerr D, Dunlop MG, Tomlinson IP. TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. *Gut* 2012;61:248-54.
- 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. Increased colorectal cancer incidence in obligate carriers of heterozygous mutations in MUTYH. *Gastroenterology* 2009 b;137:489-94, 494 e1; quiz 725-6.
- Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Parmigiani G, Kern SE, Velculescu VE, Kinzler KW, Vogelstein B, Eshleman JR, Goggins M, Klein AP. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 2009 a;324:217.

K

- Ke S, Shang S, Kalachikov SM, Morozova I, Yu L, Russo JJ, Ju J, Chasin LA. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res* 2011;21:1360-74.
- Kemp Z, Carvajal-Carmona L, Spain S, Barclay E, Gorman M, Martin L, Jaeger E, Brooks N, Bishop DT, Thomas H, Tomlinson I, Papaemmanuil E, Webb E, Sellick GS, Wood W, Evans G, Lucassen A, Maher ER, Houlston RS. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. *Hum Mol Genet* 2006 a;15:2903-10.
- Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, Bishop T, Thomas H, Houlston R, Tomlinson I. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. *Cancer Res* 2006 b;66:5003-6.
- Kim S, Parks CG, DeRoo LA, Chen H, Taylor JA, Cawthon RM, Sandler DP. Obesity and weight gain in adulthood and telomere length. *Cancer Epidemiol Biomarkers Prev* 2009;18:816-20.
- Kinzler KW, Nilbert MC, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hamilton SR, Hedge P, Markham A, et al. Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* 1991;251:1366-70.
- Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996;87:159-70.

- Knudsen AL, Bulow S, Tomlinson I, Moslein G, Heinimann K, Christensen IJ. Attenuated familial adenomatous polyposis: results from an international collaborative study. *Colorectal Dis* 2010;12:e243-9.
- Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971;68:820-3.
- Knudson AG. Cancer genetics. *Am J Med Genet* 2002;111:96-102.
- Knudson AG, Jr., Meadows AT, Nichols WW, Hill R. Chromosomal deletion and retinoblastoma. *N Engl J Med* 1976;295:1120-3.
- Ku CS, Cooper DN, Wu M, Roukos DH, Pawitan Y, Soong R, Iacopetta B. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. *Mod Pathol* 2012;25:1055-68.
- Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011;129:351-70.

L

- Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel H, Markowitz A, Klimstra D, Jhanwar S, Winawer S, Offit K, Luce MC, Kinzler KW, Vogelstein B. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet* 1997;17:79-83.
- Lan Q, Cawthon R, Shen M, Weinstein SJ, Virtamo J, Lim U, Hosgood HD, 3rd, Albanes D, Rothman N. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of non-Hodgkin lymphoma. *Clin Cancer Res* 2009;15:7429-33.
- Larsen K, Petersen J, Bernstein I, Nilbert M. A parametric model for analyzing anticipation in genetically predisposed families. *Stat Appl Genet Mol Biol* 2009;8:Article26.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortes ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CW, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G. Mutational heterogeneity in cancer and the search

- for new cancer-associated genes. *Nature* 2013;499:214-8.
- Lee IM, Lin J, Castonguay AJ, Barton NS, Buring JE, Zee RY. Mean leukocyte telomere length and risk of incident colorectal carcinoma in women: a prospective, nested case-control study. *Clin Chem Lab Med* 2010;48:259-62.
- Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, Dash C, Giardiello FM, Glick S, Levin TR, Pickhardt P, Rex DK, Thorson A, Winawer SJ. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008;58:130-60.
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78-85.
- Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ, Tsui WY, Kong CK, Brunner HG, van Kessel AG, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 2009;41:112-7.
- Lindor NM, Rabe K, Petersen GM, Haile R, Casey G, Baron J, Gallinger S, Bapat B, Aronson M, Hopper J, Jass J, LeMarchand L, Grove J, Potter J, Newcomb P, Terdiman JP, Conrad P, Moslein G, Goldberg R, Ziogas A, Anton-Culver H, de Andrade M, Siegmund K, Thibodeau SN, Boardman LA, Seminara D. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *Jama* 2005;293:1979-85.
- Lindsey J, McGill NI, Lindsey LA, Green DK, Cooke HJ. In vivo loss of telomeric repeats with age in humans. *Mutat Res* 1991;256:45-8.
- Lipton L, Halford SE, Johnson V, Novelli MR, Jones A, Cummings C, Barclay E, Sieber O, Sadat A, Bisgaard ML, Hodgson SV, Aaltonen LA, Thomas HJ, Tomlinson IP. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res* 2003;63:7595-9.
- Llor X, Pons E, Xicola RM, Castells A, Alenda C, Pinol V, Andreu M, Castellvi-Bel S, Paya A, Jover R, Bessa X, Giros A, Roca A, Gassull MA. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. *Clin Cancer Res* 2005;11:7304-10.

- Londono-Vallejo JA. Telomere instability and cancer. *Biochimie* 2008;90:73-82.
- Lynch HT. Hereditary nonpolyposis colorectal cancer (HNPCC). *Cytogenet Cell Genet* 1999;86:130-5.
- Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 2009;76:1-18.

M

- Ma X, Zhang B, Zheng W. Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Gut* 2013.
- Macklin MT. Inheritance of cancer of the stomach and large intestine in man. *J Natl Cancer Inst* 1960;24:551-71.
- Marrone A, Walne A, Dokal I. Dyskeratosis congenita: telomerase, telomeres and anticipation. *Curr Opin Genet Dev* 2005;15:249-57.
- Martinez P, Siegl-Cachedenier I, Flores JM, Blasco MA. MSH2 deficiency abolishes the anticancer and pro-aging activity of short telomeres. *Aging Cell* 2009;8:2-17.
- Martinez-Delgado B, Gallardo M, Tanic M, Yanowsky K, Inglada-Perez L, Barroso A, Rodriguez-Pinilla M, Canamero M, Blasco MA, Benitez J. Short telomeres are frequent in hereditary breast tumors and are associated with high tumor grade. *Breast Cancer Res Treat* 2013;141:231-42.
- Martinez-Delgado B, Yanowsky K, Inglada-Perez L, de la Hoya M, Caldes T, Vega A, Blanco A, Martin T, Gonzalez-Sarmiento R, Blasco M, Robledo M, Urioste M, Song H, Pharoah P, Benitez J. Shorter telomere length is associated with increased ovarian cancer risk in both familial and sporadic cases. *J Med Genet* 2012;49:341-4.
- Martinez-Delgado B, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M, Osorio A, Benitez J. Genetic anticipation is associated with telomere shortening in hereditary breast cancer. *PLoS Genet* 2011;7:e1002182.
- Maxwell F, McGlynn LM, Muir HC, Talwar D, Benzeval M, Robertson T, Roxburgh CS, McMillan DC, Horgan PG, Shiels PG. Telomere attrition and decreased fetuin-A levels indicate accelerated biological aging and are implicated in the pathogenesis of colorectal cancer. *Clin Cancer Res* 2011;17:5573-81.
- McFaul CD, Greenhalf W, Earl J, Howes N, Neoptolemos JP, Kress R, Sina-Frey M, Rieder H, Hahn S, Bartsch DK. Anticipation in familial pancreatic cancer. *Gut*

- 2006;55:252-8.
- McInnis MG. Anticipation: an old idea in new genes. *Am J Hum Genet* 1996;59:973-9.
- Meeker AK, Hicks JL, Iacobuzio-Donahue CA, Montgomery EA, Westra WH, Chan TY, Ronnett BM, De Marzo AM. Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. *Clin Cancer Res* 2004;10:3317-26.
- Mendez-Bermudez A, Royle NJ. Deficiency in DNA mismatch repair increases the rate of telomere shortening in normal human cells. *Hum Mutat* 2011;32:939-46.
- Mester J, Eng C. When overgrowth bumps into cancer: the PTEN-opathies. *Am J Med Genet C Semin Med Genet* 2013;163C:114-21.
- Middeldorp A, Jagmohan-Changur S, van Eijk R, Tops C, Devilee P, Vasen HF, Hes FJ, Houlston R, Tomlinson I, Houwing-Duistermaat JJ, Wijnen JT, Morreau H, van Wezel T. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. *Cancer Epidemiol Biomarkers Prev* 2009;18:3062-7.
- Middeldorp A, van Eijk R, Oosting J, Forte GI, van Puijenbroek M, van Nieuwenhuizen M, Corver WE, Ruano D, Caldes T, Wijnen J, Morreau H, van Wezel T. Increased frequency of 20q gain and copy-neutral loss of heterozygosity in mismatch repair proficient familial colorectal carcinomas. *Int J Cancer* 2012;130:837-46.
- Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet* 1997;17:271-2.
- Morak M, Laner A, Bacher U, Keiling C, Holinski-Feder E. MUTYH-associated polyposis - variability of the clinical phenotype in patients with biallelic and monoallelic MUTYH mutations and report on novel mutations. *Clin Genet* 2010;78:353-63.
- Morak M, Laner A, Scholz M, Madorf T, Holinski-Feder E. Report on de-novo mutation in the MSH2 gene as a rare event in hereditary nonpolyposis colorectal cancer. *Eur J Gastroenterol Hepatol* 2008;20:1101-5.
- Mueller-Koch Y, Vogelsang H, Kopp R, Lohse P, Keller G, Aust D, Muders M, Gross M, Daum J, Schieman U, Grabowski M, Scholz M, Kerker B, Becker I, Henke G, Holinski-Feder E. Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer. *Gut* 2005;54:1733-40.

N

- Neklason DW, Kerber RA, Nilson DB, Anton-Culver H, Schwartz AG, Griffin CA, Lowery JT, Schildkraut JM, Evans JP, Tomlinson GE, Strong LC, Miller AR, Stopfer JE,

- Finkelstein DM, Nadkarni PM, Kasten CH, Mineau GP, Burt RW. Common familial colorectal cancer linked to chromosome 7q31: a genome-wide analysis. *Cancer Res* 2008;68:8993-7.
- Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330-7.
- Nielsen M, Joerink-van de Beld MC, Jones N, Vogt S, Tops CM, Vasen HF, Sampson JR, Aretz S, Hes FJ. Analysis of MUTYH genotypes and colorectal phenotypes in patients With MUTYH-associated polyposis. *Gastroenterology* 2009;136:471-6.
- Nielsen M, Morreau H, Vasen HF, Hes FJ. MUTYH-associated polyposis (MAP). *Crit Rev Oncol Hematol* 2011;79:1-16.
- Nieminen TT, Abdel-Rahman WM, Ristimaki A, Lappalainen M, Lahermo P, Mecklin JP, Jarvinen HJ, Peltomaki P. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 2011;141:e23-6.
- Niittymaki I, Kaasinen E, Tuupanen S, Karhu A, Jarvinen H, Mecklin JP, Tomlinson IP, Di Bernardo MC, Houlston RS, Aaltonen LA. Low-penetrance susceptibility variants in familial colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2010;19:1478-83.
- Nilbert M, Timshel S, Bernstein I, Larsen K. Role for genetic anticipation in Lynch syndrome. *J Clin Oncol* 2009;27:360-4.
- Nystrom-Lahti M, Kristo P, Nicolaides NC, Chang SY, Aaltonen LA, Moisio AL, Jarvinen HJ, Mecklin JP, Kinzler KW, Vogelstein B, et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat Med* 1995;1:1203-6.

O

- Olsen MT, Berube M, Robbins J, Palsboll PJ. Empirical evaluation of humpback whale telomere length estimates; quality control and factors causing variability in the singleplex and multiplex qPCR methods. *BMC Genet* 2012;13:77.

P

- Palles C, Cazier JB, 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, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsley SE, Thomas HJ, McVean G, Houlston

- RS, Tomlinson I. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013;45:136-44.
- Parker AR, Sieber OM, Shi C, Hua L, Takao M, Tomlinson IP, Eshleman JR. Cells with pathogenic biallelic mutations in the human MUTYH gene are defective in DNA damage binding and repair. *Carcinogenesis* 2005;26:2010-8.
- Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. 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. *J Med Genet* 2012;49:151-7.
- Pasche B, Kaklamani V, Hou N, Young T, Rademaker A, Peterlongo P, Ellis N, Offit K, Caldes T, Reiss M, Zheng T. TGFBR1*6A and cancer: a meta-analysis of 12 case-control studies. *J Clin Oncol* 2004;22:756-8.
- Pasche B, Kolachana P, Nafa K, Satagopan J, Chen YG, Lo RS, Brenner D, Yang D, Kirstein L, Oddoux C, Ostrer H, Vineis P, Varesco L, Jhanwar S, Luzzatto L, Massague J, Offit K. TbetaR-I(6A) is a candidate tumor susceptibility allele. *Cancer Res* 1999;59:5678-82.
- Pellatt AJ, Wolff RK, Lundgreen A, Cawthon R, Slattery ML. Genetic and lifestyle influence on telomere length and subsequent risk of colon cancer in a case control study. *Int J Mol Epidemiol Genet* 2012;3:184-94.
- Petersen GM, Slack J, Nakamura Y. Screening guidelines and premorbid diagnosis of familial adenomatous polyposis using linkage. *Gastroenterology* 1991;100:1658-64.
- Phelan CM, Tsai YY, Goode EL, Vierkant RA, Fridley BL, Beesley J, Chen XQ, Webb PM, Chanock S, Cramer DW, Moysich K, Edwards RP, Chang-Claude J, Garcia-Closas M, Yang H, Wang-Gohrke S, Hein R, Green AC, Lissowska J, Carney ME, Lurie G, Wilkens LR, Ness RB, Pearce CL, Wu AH, Van Den Berg DJ, Stram DO, Terry KL, Whitman DC, Whittemore AS, DiCioccio RA, McGuire V, Doherty JA, Rossing MA, Anton-Culver H, Ziogas A, Hogdall C, Hogdall E, Kruger Kjaer S, Blaakaer J, Quaye L, Ramus SJ, Jacobs I, Song H, Pharoah PD, Iversen ES, Marks JR, Pike MC, Gayther SA, Cunningham JM, Goodman MT, Schildkraut JM, Chenevix-Trench G, Berchuck A, Sellers TA. Polymorphism in the GALNT1 gene and epithelial ovarian cancer in non-Hispanic white women: the Ovarian Cancer Association Consortium. *Cancer Epidemiol Biomarkers Prev* 2010;19:600-4.
- Pineda M, Gonzalez S, Lazaro C, Blanco I, Capella G. Detection of genetic alterations in

- hereditary colorectal cancer screening. *Mutat Res* 2010;693:19-31.
- Plotz G, Casper M, Raedle J, Hinrichsen I, Heckel V, Brieger A, Trojan J, Zeuzem S. MUTYH gene expression and alternative splicing in controls and polyposis patients. *Hum Mutat* 2012;33:1067-74.
- Pooley KA, Sandhu MS, Tyrer J, Shah M, Driver KE, Luben RN, Bingham SA, Ponder BA, Pharoah PD, Khaw KT, Easton DF, Dunning AM. Telomere length in prospective and retrospective cancer case-control studies. *Cancer Res* 2010;70:3170-6.
- Potter JD. Colorectal cancer: molecules and populations. *J Natl Cancer Inst* 1999;91:916-32.
- Prescott J, Wentzensen IM, Savage SA, De Vivo I. Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutat Res* 2011;730:75-84.
- 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 MC, Tait JF, Walsh T. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn* 2012;14:357-66.

Q

- Qin Q, Sun J, Yin J, Liu L, Chen J, Zhang Y, Li T, Shi Y, Wei S, Nie S. Telomere length in peripheral blood leukocytes is associated with risk of colorectal cancer in Chinese population. *PLoS One* 2014;9:e88135.
- Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, Lopez-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Gine E, Hernandez JM, Gonzalez-Diaz M, Puente DA, Velasco G, Freije JM, Tubio JM, Royo R, Gelpi JL, Orozco M, Pisano DG, Zamora J, Vazquez M, Valencia A, Himmelbauer H, Bayes M, Heath S, Gut M, Gut I, Estivill X, Lopez-Guillermo A, Puente XS, Campo E, Lopez-Otin C. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet* 2011;44:47-52.

R

- Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D, Easton DF, Stratton MR. PALB2, which encodes a BRCA2-interacting protein, is a breast

- cancer susceptibility gene. *Nat Genet* 2007;39:165-7.
- Rampazzo E, Bertorelle R, Serra L, Terrin L, Candiotto C, Pucciarelli S, Del Bianco P, Nitti D, De Rossi A. Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers. *Br J Cancer* 2010;102:1300-5.
- Raynaud CM, Sabatier L, Philipot O, Olaussen KA, Soria JC. Telomere length, telomeric proteins and genomic instability during the multistep carcinogenic process. *Crit Rev Oncol Hematol* 2008;66:99-117.
- Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, Aittomaki K, Jarvinen HJ, Mecklin JP, Lindblom A, Peltomaki P. Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. *J Clin Oncol* 2003;21:3629-37.
- Reya T, Clevers H. Wnt signalling in stem cells and cancer. *Nature* 2005;434:843-50.
- Rigter T, Henneman L, Kristoffersson U, Hall A, Yntema HG, Borry P, Tonnies H, Waisfisz Q, Elting MW, Dondorp WJ, Cornel MC. Reflecting on earlier experiences with unsolicited findings: points to consider for next-generation sequencing and informed consent in diagnostics. *Hum Mutat* 2013;34:1322-8.
- Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, Srivastava S. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J Natl Cancer Inst* 1997;89:1758-62.

S

- Salk JJ, Fox EJ, Loeb LA. Mutational heterogeneity in human cancers: origin and consequences. *Annu Rev Pathol* 2010;5:51-75.
- 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. Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* 2003;362:39-41.
- Sanchez-de-Abajo A, de la Hoya M, van Puijenbroek M, Tosar A, Lopez-Asenjo JA, Diaz-Rubio E, Morreau H, Caldes T. Molecular analysis of colorectal cancer tumors from patients with mismatch repair proficient hereditary nonpolyposis colorectal cancer suggests novel carcinogenic pathways. *Clin Cancer Res* 2007;13:5729-35.
- Sayed MG, Ahmed AF, Ringold JR, Anderson ME, Bair JL, Mitros FA, Lynch HT, Tinley ST, Petersen GM, Giardiello FM, Vogelstein B, Howe JR. Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis. *Ann Surg Oncol*

- 2002;9:901-6.
- Schwartz RA, Torre DP. The Muir-Torre syndrome: a 25-year retrospect. *J Am Acad Dermatol* 1995;33:90-104.
- Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR, Rahman N. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 2006;38:1239-41.
- Segui N, Stevens KN, Guino E, Rozek LS, Moreno VR, Capella G, Gruber SB, Valle L. No association between germline allele-specific expression of TGFBR1 and colorectal cancer risk in Caucasian and Ashkenazi populations. *Br J Cancer* 2011;104:735-40.
- Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, Lindblom A, Lagerstedt K, Thibodeau SN, Lindor NM, Young J, Winship I, Dowty JG, White DM, Hopper JL, Baglietto L, Jenkins MA, de la Chapelle A. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology* 2008;135:419-28.
- Shibata D, Peinado MA, Ionov Y, Malkhosyan S, Perucho M. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet* 1994;6:273-81.
- Sieber OM, Lamlum H, Crabtree MD, Rowan AJ, Barclay E, Lipton L, Hodgson S, Thomas HJ, Neale K, Phillips RK, Farrington SM, Dunlop MG, Mueller HJ, Bisgaard ML, Bulow S, Fidalgo P, Albuquerque C, Scarano MI, Bodmer W, Tomlinson IP, Heinemann K. Whole-gene APC deletions cause classical familial adenomatous polyposis, but not attenuated polyposis or "multiple" colorectal adenomas. *Proc Natl Acad Sci U S A* 2002;99:2954-8.
- Skoglund J, Djureinovic T, Zhou XL, Vandrovicova J, Renkonen E, Iselius L, Bisgaard ML, Peltomaki P, Lindblom A. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. *J Med Genet* 2006;43:e7.
- Skoglund J, Song B, Dalen J, Dedorson S, Edler D, Hjern F, Holm J, Lenander C, Lindfors U, Lundqvist N, Olivecrona H, Olsson L, Pahlman L, Rutegard J, Smedh K, Tornqvist A, Houlston RS, Lindblom A. Lack of an association between the TGFBR1*6A variant and colorectal cancer risk. *Clin Cancer Res* 2007;13:3748-52.
- Stadler ZK, Schrader KA, Vijai J, Robson ME, Offit K. Cancer genomics and inherited risk.

J Clin Oncol 2014;32:687-98.

Stella A, Surdo NC, Lastella P, Barana D, Oliani C, Tibiletti MG, Viel A, Natale C, Piepoli A, Marra G, Guanti G. Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. Clin Genet 2007;71:130-9.

Stupart D, Goldberg P, Algar U, Vorster A, Ramesar R. No evidence of genetic anticipation in a large family with Lynch syndrome. Fam Cancer 2013.

Syngal S, Fox EA, Eng C, Kolodner RD, Garber JE. Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. J Med Genet 2000;37:641-5.

T

Tabori U, Nanda S, Druker H, Lees J, Malkin D. Younger age of cancer initiation is associated with shorter telomere length in Li-Fraumeni syndrome. Cancer Res 2007;67:1415-8.

Takagi S, Kinouchi Y, Hiwatashi N, Chida M, Nagashima F, Takahashi S, Negoro K, Shimosegawa T, Toyota T. Telomere shortening and the clinicopathologic characteristics of human colorectal carcinomas. Cancer 1999;86:1431-6.

Takagi S, Kinouchi Y, Hiwatashi N, Nagashima F, Chida M, Takahashi S, Negoro K, Shimosegawa T, Toyota T. Relationship between microsatellite instability and telomere shortening in colorectal cancer. Dis Colon Rectum 2000;43:S12-7.

Tan MH, Mester J, Peterson C, Yang Y, Chen JL, Rybicki LA, Milas K, Pederson H, Remzi B, Orloff MS, Eng C. A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. Am J Hum Genet 2011;88:42-56.

Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB. In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat 2008;29:1327-36.

Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet 2010;19:R145-51.

Ten Hagen KG, Fritz TA, Tabak LA. All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. Glycobiology 2003;13:1R-16R.

Tenesa A, Campbell H, Barnetson R, Porteous M, Dunlop M, Farrington SM. Association of MUTYH and colorectal cancer. Br J Cancer 2006;95:239-42.

Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, Dotson WD, Douglas MP, Berg AO. The Evaluation of Genomic Applications in Practice and

- Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med* 2009;11:3-14.
- Therkildsen C, Jonsson G, Dominguez-Valentin M, Nissen A, Rambech E, Halvarsson B, Bernstein I, Borg K, Nilbert M. Gain of chromosomal region 20q and loss of 18 discriminates between Lynch syndrome and familial colorectal cancer. *Eur J Cancer* 2013;49:1226-35.
- Thompson ER, Doyle MA, Ryland GL, Rowley SM, Choong DY, Tothill RW, Thorne H, Barnes DR, Li J, Ellul J, Philip GK, Antill YC, James PA, Trainer AH, Mitchell G, Campbell IG. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet* 2012;8:e1002894.
- Tomlinson I, Rahman N, Frayling I, Mangion J, Barfoot R, Hamoudi R, Seal S, Northover J, Thomas HJ, Neale K, Hodgson S, Talbot I, Houlston R, Stratton MR. Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology* 1999;116:789-95.
- Tomsic J, Senter L, Liyanarachchi S, Clendenning M, Vaughn CP, Jenkins MA, Hopper JL, Young J, Samowitz W, de la Chapelle A. Recurrent and founder mutations in the PMS2 gene. *Clin Genet* 2013;83:238-43.
- Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* 1999;96:8681-6.
- Tsai YY, Petersen GM, Booker SV, Bacon JA, Hamilton SR, Giardiello FM. Evidence against genetic anticipation in familial colorectal cancer. *Genet Epidemiol* 1997;14:435-46.

U

- Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff 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 HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 2004;96:261-8.

V

- Valle L. Debate about TGFBR1 and the susceptibility to colorectal cancer. *World J*

- Gastrointest Oncol 2012;4:1-8.
- Valle L, Hernandez-Illan E, Bellido F, Aiza G, Castillejo A, Castillejo MI, Navarro M, Segui N, Vargas G, Guarinos C, Juarez M, Sanjuan X, Iglesias S, Alenda C, Egoavil C, Segura A, Juan MJ, Rodriguez-Soler M, Brunet J, Gonzalez S, Jover R, Lazaro C, Capella G, Pineda M, Soto JL, Blanco I. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. *Hum Mol Genet* 2014a.
- Valle L. Genetic predisposition to colorectal cancer: where do we stand and future perspectives. *World J Gastroenterol* 2014b; en prensa.
- Valle L, Perea J, Carbonell P, Fernandez V, Dotor AM, Benitez J, Urioste M. Clinicopathologic and pedigree differences in amsterdam I-positive hereditary nonpolyposis colorectal cancer families according to tumor microsatellite instability status. *J Clin Oncol* 2007;25:781-6.
- Valle L, Serena-Acedo T, Liyanarachchi S, Hampel H, Comeras I, Li Z, Zeng Q, Zhang HT, Pennison MJ, Sadim M, Pasche B, Tanner SM, de la Chapelle A. Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. *Science* 2008;321:1361-5.
- van Puijenbroek M, Nielsen M, Tops CM, Halfwerk H, Vasen HF, Weiss MM, van Wezel T, Hes FJ, Morreau H. Identification of patients with (atypical) MUTYH-associated polyposis by KRAS2 c.34G > T prescreening followed by MUTYH hotspot analysis in formalin-fixed paraffin-embedded tissue. *Clin Cancer Res* 2008;14:139-42.
- van Vliet CM, Dowty JG, van Vliet JL, Smith L, Mead LJ, Macrae FA, St John DJ, Giles GG, Southey MC, Jenkins MA, Velan GM, Hopper JL. Dependence of colorectal cancer risk on the parent-of-origin of mutations in DNA mismatch repair genes. *Hum Mutat* 2011;32:207-12.
- van Wezel T, Middeldorp A, Wijnen JT, Morreau H. A review of the genetic background and tumour profiling in familial colorectal cancer. *Mutagenesis* 2012;27:239-45.
- Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 1991;34:424-5.
- Vasen HF, Taal BG, Griffioen G, Nagengast FM, Cats A, Menko FH, Oskam W, Kleibeuker JH, Offerhaus GJ, Khan PM. Clinical heterogeneity of familial colorectal cancer and its influence on screening protocols. *Gut* 1994;35:1262-6.
- Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the

- International Collaborative group on HNPCC. *Gastroenterology* 1999;116:1453-6.
- Venkatachalam R, Ligtenberg MJ, Hoogerbrugge N, Schackert HK, Gorgens H, Hahn MM, Kamping EJ, Vreede L, Hoenselaar E, van der Looij E, Goossens M, Churchman M, Carvajal-Carmona L, Tomlinson IP, de Bruijn DR, Van Kessel AG, Kuiper RP. Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 2010;139:2221-4.
- Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988;319:525-32.
- Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004;10:789-99.
- Vogt S, Jones N, Christian D, Engel C, Nielsen M, Kaufmann A, Steinke V, Vasen HF, Propping P, Sampson JR, Hes FJ, Aretz S. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. *Gastroenterology* 2009;137:1976-85 e1-10.
- von Zglinicki T. Oxidative stress shortens telomeres. *Trends Biochem Sci* 2002;27:339-44.
- Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* 2001;413:432-5.

W

- Ward RL, Dobbins T, Lindor NM, Rapkins RW, Hitchins MP. Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. *Genet Med* 2013 a;15:25-35.
- Ward RL, Hicks S, Hawkins NJ. Population-based molecular screening for Lynch syndrome: implications for personalized medicine. *J Clin Oncol* 2013 b;31:2554-62.
- Webb EL, Rudd MF, Houlston RS. Colorectal cancer risk in monoallelic carriers of MYH variants. *Am J Hum Genet* 2006;79:768-71; author reply 771-2.
- Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. The association of telomere length and cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2011;20:1238-50.
- Westphalen AA, Russell AM, Buser M, Berthod CR, Hutter P, Plasilova M, Mueller H, Heinimann K. Evidence for genetic anticipation in hereditary non-polyposis colorectal cancer. *Hum Genet* 2005;116:461-5.

- Whitelaw SC, Murday VA, Tomlinson IP, Thomas HJ, Cottrell S, Ginsberg A, Bukofzer S, Hodgson SV, Skudowitz RB, Jass JR, Talbot IC, Northover JM, Bodmer WF, Solomon E. Clinical and molecular features of the hereditary mixed polyposis syndrome. *Gastroenterology* 1997;112:327-34.
- Wiernik PH, Wang SQ, Hu XP, Marino P, Paietta E. Age of onset evidence for anticipation in familial non-Hodgkin's lymphoma. *Br J Haematol* 2000;108:72-9.
- Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, Markowitz SD. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proc Natl Acad Sci U S A* 2003;100:12961-5.
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyansky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318:1108-13.
- Wolf CM, Richards RC, Gardner EJ. Occasional discrete polyps of the colon and rectum showing an inherited tendency in a kindred. *Cancer* 1955;8:403-8.

X

- Xie Y, Yang H, Miller JH, Shih DM, Hicks GG, Xie J, Shiu RP. Cells deficient in oxidative DNA damage repair genes *Myh* and *Ogg1* are sensitive to oxidants with increased G2/M arrest and multinucleation. *Carcinogenesis* 2008;29:722-8.
- Xu L, Li S, Stohr BA. The role of telomere biology in cancer. *Annu Rev Pathol* 2013;8:49-78.

Y

- Yamamoto H, Higasa K, Sakaguchi M, Shien K, Soh J, Ichimura K, Furukawa M, Hashida S, Tsukuda K, Takigawa N, Matsuo K, Kiura K, Miyoshi S, Matsuda F, Toyooka S. Novel Germline Mutation in the Transmembrane Domain of HER2 in Familial Lung Adenocarcinomas. *J Natl Cancer Inst* 2013.

Z

- Zee RY, Castonguay AJ, Barton NS, Buring JE. Mean telomere length and risk of incident colorectal carcinoma: a prospective, nested case-control approach. *Cancer Epidemiol Biomarkers Prev* 2009;18:2280-2.
- Zeng Q, Phukan S, Xu Y, Sadim M, Rosman DS, Pennison M, Liao J, Yang GY, Huang CC, Valle L, Di Cristofano A, de la Chapelle A, Pasche B. *Tgfb1* haploinsufficiency is a potent modifier of colorectal cancer development. *Cancer Res* 2009;69:678-86.

ANEXOS

ANEXO I: Otras publicaciones

No association between germline allele-specific expression of *TGFBR1* and colorectal cancer risk in Caucasian and Ashkenazi populations

Nuria Seguí, Kristen N Stevens, Elisabet Guinó, Laura S Rozek, Victor Moreno, Gabriel Capellá, Setephen B Gruber y Laura Valle.

British Journal of Cancer 104(4):735-40; 2011

Resumen:

La expresión alelo-específica de *TGFBR1* se había presentado en un estudio anterior como un factor genético de riesgo a desarrollar CCR (*odds ratio* cercana a 9). Sin embargo, los resultados obtenidos requerían validación en una serie independiente para poder evaluar su posible aplicación en la evaluación clínica de la susceptibilidad al CCR.

Para ello estudiamos 426 pacientes con CCR y 433 controles de origen judío Asquenazi y 178 pacientes con CCR españoles. Para el estudio de expresión alelo específica del gen utilizamos la pirosecuenciación, una técnica cuantitativa que mostró mayor robustez que el SNaPshot, método utilizado para la evaluación del desequilibrio alélico en el estudio original.

De la colección de muestras de origen judío Ashkenazi el 26,5% de los individuos fueron informativos para al menos uno de los marcadores alélicos estudiados. Se estudió la expresión alelo-específica en ARN extraído de sangre periférica, obteniendo valores en un rango de 0,76 a 1,67 (mediana: 1,00) para los casos y de 0,67 a 1,39 (mediana: 1,00) para los controles (diferencia media: -0,003; intervalo de confianza 95%: -0,032 a 0,025; $p=0,81$). En la serie española, 50% de los individuos resultaron informativos para los marcadores alélicos analizados. Se evaluó la expresión alelo-específica del gen en ARN extraído de mucosa colónica normal. Los valores de expresión alelo-específica para este grupo oscilaron entre 0,68 y 1,40 (mediana: 1,01), sin detectar diferencias estadísticamente significativas al comparar con los casos de CCR judío Ashkenazis y/o controles. En conjunto, nuestros resultados no muestran diferencias en el grado de expresión alelo-específica de *TGFBR1* entre pacientes con CCR y controles, sugiriendo que no contribuye a la predisposición genética del CCR.

Full Paper

No association between germline allele-specific expression of *TGFBR1* and colorectal cancer risk in Caucasian and Ashkenazi populations

N Seguí^{1,6}, KN Stevens^{2,6}, E Guinó³, LS Rozek⁴, VR Moreno^{3,5}, G Capellá¹, SB Gruber^{*,2} and L Valle^{*,1}

¹Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Av. Gran Vía 199-203, Barcelona 08908, Spain; ²Departments of Internal Medicine, Epidemiology, and Human Genetics, Division of Molecular Medicine and Genetics, University of Michigan, 1524 BSRB, 109 Zina Pitcher, Ann Arbor, MI 48109, USA; ³Bioinformatics and Biostatistics Unit, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Barcelona 08908, Spain; ⁴Environmental Health Sciences, School of Public Health, University of Michigan Medical School, Ann Arbor, MI 48109, USA; ⁵Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain

BACKGROUND: Germline allele-specific expression (ASE) of the *TGFBR1* gene has been reported as a strong risk factor for colorectal cancer (CRC) with an odds ratio close to 9. Considering the potential implications of the finding, we undertook the task of validating the initial results in this study.

METHODS: Allele-specific expression was measured using the highly quantitative and robust technique of pyrosequencing. Individuals from two different populations were studied, one Caucasian-dominated and the other of Ashkenazi Jewish descent, with different sources of non-tumoral genetic material in each.

RESULTS: Our results showed no statistically significant differences in the degree of ASE between CRC patients and controls, considering ASE as either a quantitative or a binary trait. Using defined cutoff values to categorise ASE, 1.0% of blood lymphocytes from informative Israeli cases (total $n = 96$) were ASE positive (median 1.00; range 0.76–1.31) and 2.2% of informative matched controls (total $n = 90$) were ASE positive (median 1.00; range 0.76–1.87). Likewise, normal mucosae from Spanish patients (median 1.03; range: 0.68–1.43; $n = 75$) did not show significant differences in the degree of ASE when compared with the Israeli patients or controls.

CONCLUSIONS: Taken together, these results suggest that ASE of *TGFBR1* does not confer an increased risk of CRC.

British Journal of Cancer advance online publication, 11 January 2011; doi:10.1038/sj.bjc.6606079 www.bjcancer.com
© 2011 Cancer Research UK

Keywords: allele-specific expression; allelic imbalance; *TGFBR1*; colorectal cancer predisposition; pyrosequencing; cancer-predisposing gene

Colorectal cancer (CRC) is among the four most common cancers in industrialised countries, and is one of the leading causes of cancer-related deaths. Although familial clustering of CRC occurs in 20–30% of all cases, the known highly penetrant autosomal-dominant and -recessive forms of the disease account for less than 5% of all CRC cases (de la Chapelle, 2004; Lynch *et al*, 2009). Although additional low-penetrance alleles have been proposed in the last few years, the underlying genetic risk factors for CRC predisposition remain largely unknown (Hemminki *et al*, 2009).

On the basis of previous evidence that pointed towards the importance of downstream signalling elements of the transforming growth factor β (TGF- β) pathway in CRC (Wood *et al*, 2007), and the known linkage peak for familial CRC in 9q22–31 where *TGFBR1* is located (Wiesner *et al*, 2003; Kemp *et al*, 2006;

Skoglund *et al*, 2006), we undertook the task of studying the role of *TGFBR1* in CRC predisposition. Although risk-conferring germline genetic variants in this gene had not been identified, we reported that germline allele-specific expression (ASE) of *TGFBR1*, measured with the SNaPshot technique, occurred in ~20% of informative CRC patients and ~3% of informative controls, thus conferring a substantially increased risk of CRC (odds ratio 8.7, 95% confidence interval (CI): 2.6–29.1) (Valle *et al*, 2008). This differential allele-specific expression was suggested to be dominantly inherited and to alter the downstream SMAD-mediated TGF- β signalling (Valle *et al*, 2008). A subsequent report showed that *APC^{Min/+};Tgfbr1^{+/-}* mice developed twice as many intestinal tumours and colonic carcinomas as *APC^{Min/+};Tgfbr1^{+/+}*, supporting the role of *TGFBR1* gene haploinsufficiency in CRC development (Zeng *et al*, 2009). Also, *TGFBR1*6A*, a common variant in exon 1 of the gene, has been weakly associated with CRC (Pasche *et al*, 2004; Skoglund *et al*, 2007).

As allele-specific expression of *TGFBR1* has the potential to be used in the clinical evaluation of CRC risk, the aim of this study was to further investigate the extent of ASE of *TGFBR1* in CRC using the robust and specific pyrosequencing technique for ASE

*Correspondence: Dr L Valle; E-mail: lvalle@iconcologia.net or Dr SB Gruber; E-mail: sgruber@med.umich.edu

⁶These authors contributed equally to this work.

Received 22 October 2010; revised 3 December 2010; accepted 6 December 2010

determination. In addition, we studied two different populations with different biological sources of non-tumour genetic material to evaluate ASE frequency in a variety of populations.

MATERIALS AND METHODS

Patients and controls

Uncultured blood lymphocytes from a total of 426 Ashkenazi Jewish CRC patients and 433 Ashkenazi Jewish controls were obtained from a collection of Israeli CRC patients and matched controls. This series corresponds to a population-based case-control study (Molecular Epidemiology of Colorectal Cancer; MECC) of incident CRC, including histopathologically confirmed cases of all incident CRC diagnosed in northern Israel beginning 31 March 1998 (Poynter *et al*, 2005). Informed consent was obtained from all of the subjects who participated in the study. All 426 CRC patients showed tumour microsatellite stability and did not carry germline mutations in known cancer-predisposing genes.

A total of 178 normal mucosae from Spanish CRC patients (6% showed tumour microsatellite instability) were obtained from a hospital-based case-control study (Bellvitge Colorectal Cancer Study; BCCS). Cases were consecutive patients with a new diagnosis of colorectal adenocarcinoma attending a University Hospital in Barcelona. Details about the study population, interviews and collection of biological samples were published elsewhere (Landi *et al*, 2003).

Nucleic acid extraction and cDNA synthesis

Genomic DNA from purified blood lymphocytes and frozen normal colon mucosa was extracted using standard phenol-chloroform procedures. For total RNA extraction, the different tissue sources were processed with TRIzol reagent (Invitrogen, Carlsbad, CA, USA). In all cases, nucleic acid concentrations and purities were analyzed with the NanoDrop spectrophotometer, and the level of degradation of the RNA (RIN number) was checked by using the RNA Nano assay on the Agilent 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, CA, USA) when necessary.

Total RNA was treated with DNase (DNasefree, Ambion, Austin, TX, USA) before cDNA synthesis (Transcriptor First Strand cDNA Synthesis Kit, Roche Diagnostics GmbH, Mannheim, Germany).

Genotyping of transcribed SNPs

Current techniques for ASE determination require heterozygous markers in the transcribed regions of the gene to discriminate between its two alleles. The transcribed markers we used were four SNPs located in the 3'-UTR region: rs334349, rs420549, rs7850895 and rs1590. Owing to the fact that rs334349 and rs1590 are in strong linkage disequilibrium, only the latter was genotyped to determine the informativeness of both. Three commercially available TaqMan SNP genotyping assays were used to genotype rs420549 (C_662618_1), rs7850895 (C_29248567_20) and rs1590 (C_2945143_10) (Applied Biosystems Inc., Foster City, CA, USA). Reactions were performed following the instructions provided by the manufacturer.

ASE determination by pyrosequencing

PCR and pyrosequencing reactions for rs334349, rs1590 and rs420549 were performed as described previously (Guda *et al*, 2009). For rs7850895, PCR and sequencing primers were designed using the PSQ Assay Design software provided by the manufacturer: PCR-fw-5'-TCATGCCATATGTAGTTGCTGTAG-3'; biotinylated PCR-rv-5'-ACACCCTAAGCATGTGGAGA-3'; and SEQ-5'-CTAGTGCAAGTTACAATAT-3'. After PCR, DNA and cDNA

amplification products were sequenced on a PyroMark MD pyrosequencing instrument (Qiagen, Chatsworth, CA, USA).

The proportions of individual alleles for each SNP were obtained from the PyroMark MD software calculations. The medians and standard deviations (s.d.) of the triplicates for both DNA and cDNA were calculated for each SNP. To obtain an ASE value, the ratio of the common vs the rare allele in the cDNA was normalised to the respective ratio in the DNA: cDNA (median common allele/median rare allele)/DNA (median common allele/median rare allele). The final ASE value was calculated as the median of the ASE values obtained for the SNPs studied in each sample.

Before the complete analysis of all informative samples, we tested the robustness and reproducibility of pyrosequencing compared with SNaPshot, the technique used in the original report (Valle *et al*, 2008), by randomly choosing 10 informative samples and measuring ASE using both techniques. SNaPshot was carried out as described previously (Valle *et al*, 2008) and ASE value calculations were performed as described above for both techniques. Pyrosequencing yielded lower variability in ASE among different SNP markers and it was able to obtain valuable results in situations when SNaPshot was not able to assess ASE (Supplementary Table 1). Guda *et al* (2009) previously reported additional information showing that ASE results obtained by pyrosequencing can be reproduced by SNaPshot, supporting the ability of pyrosequencing to detect allelic imbalances.

Statistical analyses

Pairwise comparisons between cases and controls were performed using the Wilcoxon's rank-sum test and Bonferroni correction was applied to account for the two comparisons performed: MECC cases vs MECC controls and MECC cases vs BCCS cases.

To dichotomise the ASE variable, cutoff points were established based on the ASE values obtained in cancer-free controls (median \pm (2 \times s.d.)). When ASE was considered as a binary variable, comparisons of proportions between cases and controls were performed using a likelihood ratio test derived from logistic regression adjusting for population source.

RESULTS

Among a total of 426 Ashkenazi Jewish CRC patients from the Israeli MECC study, 115 (27%) were informative for at least one SNP tested: 88 (20.7%) for rs334349 and rs1590, 65 (15.3%) for rs420549 and 18 (4.2%) for rs7850895. Of a total of 433 Ashkenazi Jewish MECC controls, 112 (25.9%) were informative: 81 (18.7%) for rs334349 and rs1590, 59 (13.6%) for rs420549 and 19 (4.4%) for rs7850895. Of 178 normal mucosae from a Spanish collection of Caucasian CRC patients (BCCS), 88 (49.4%) were heterozygous for at least one SNP genotyped: 64 (36%) for rs334349 and rs1590, 64 (36%) for rs420549 and 18 (10%) for rs7850895. Three of the 88 had no RNA available.

ASE values were obtained for 96 (83.5%) informative MECC CRC patients, 90 (80.4%) informative MECC controls and 75 (85.2%) informative BCCS CRC patients. The ASE values obtained for cases and controls are shown in Figure 1. For the MECC series alone, values range from 0.76 to 1.31 (median: 1.00) in cases, and from 0.76 to 1.87 (median: 1.00) in controls (Figure 1A). When ASE was considered as a continuous variable, no differences were detected between cases and controls (median difference -0.002 ; 95% CI: -0.027 to 0.032 ; $P=0.86$). Although observed data suggest that ASE is a quantitative trait, ASE was transformed into a binary trait (ASE vs non-ASE) to facilitate the interpretation of the results. For this purpose, cutoff points were defined based on the results obtained in controls (median \pm (2 \times s.d.) = $1.00 \pm$ (2 \times 0.157)). Under that criterion, 1.0% (1 out of 96) of informative CRC patients and 2.2% (2 out of 90) of informative controls showed ASE of *TGFBR1* ($P=0.52$) (Figure 2).

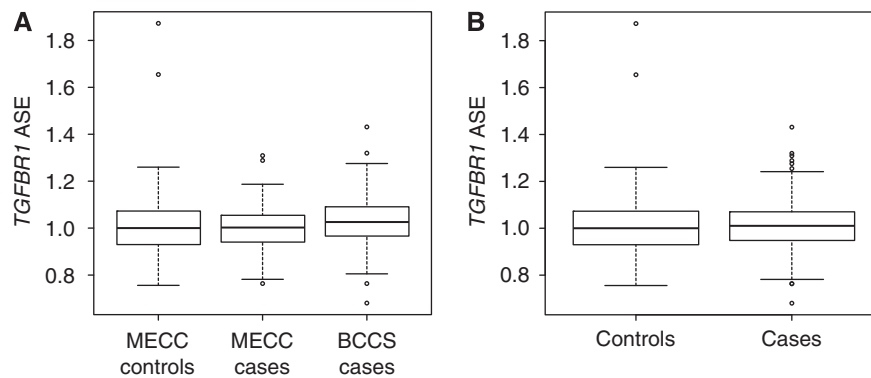


Figure 1 *TGFBR1* ASE distributions in cases and controls. **(A)** *TGFBR1* ASE in MECC controls ($n=90$), MECC CRC patients ($n=96$) and BCCS CRC patients ($n=75$). **(B)** *TGFBR1* ASE in CRC patients ($n=171$) and controls ($n=90$). The boxes represent the inter-quartile range of distributions (25–75th percentile); the horizontal lines within the boxes represent the medians; and the vertical lines represent the 5 and 95th percentiles.

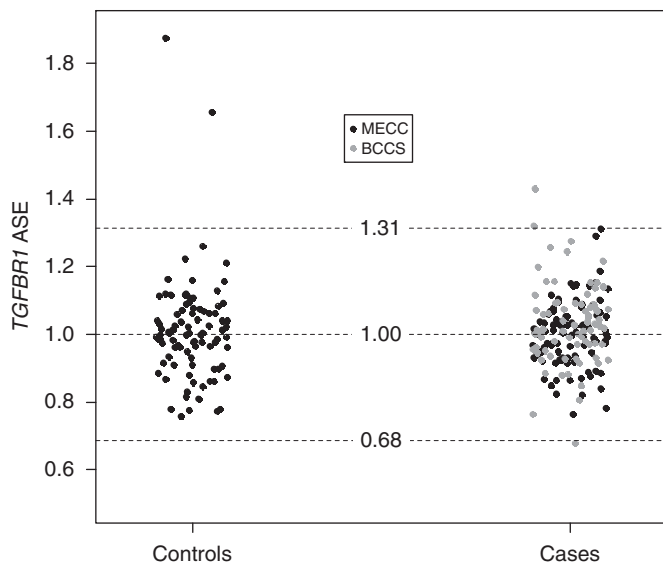


Figure 2 *TGFBR1* ASE distribution in 171 CRC patients (96 MECC (black dots) and 75 BCCS (grey dots) CRC patients) and 90 controls. The median and cutoff points, defined as the median $\pm 2 \times$ s.d. of controls, used to categorise ASE are indicated as discontinuous lines.

No differences in ASE levels were identified between MECC CRC patients (median 1.00; range 0.76–1.31) and BCCS patients (median 1.03; range 0.68–1.43) (median difference 0.026; 95% CI: -0.001 to 0.059 ; $P=0.06$). Consistent with this analysis, no differences were identified between MECC and BCCS CRC patients when ASE was treated as a binary variable ($P=0.20$), suggesting that ethnic origin (Ashkenazi Jewish and Caucasian) and the source of biological material assessed (uncultured blood lymphocytes and normal colon mucosae) do not have a major influence on *TGFBR1* ASE assessment. This allowed us to combine the two groups of CRC patients and compare them with the available group of controls (Figures 1B and 2). Combining all data from MECC and BCCS subjects, no differences were detected between CRC patients and controls when considering ASE as either a quantitative (median difference -0.010 ; 95% CI: -0.037 to 0.017 ; $P=0.48$) or a binary variable ($P=0.52$, adjusted by population).

The RNA quality of subjects with ASE in both the BCCS and MECC series was checked to confirm that they had been classified as such owing to the presence of real allelic imbalances and not owing to technical artefacts caused by poor RNA quality. In all samples where the source RNA was available, the RIN value was above 6. To ensure that our results were not affected by poor

performance of the PCR/pyrosequencing reaction owing to low RNA quality, a more stringent analysis was carried out, including only those samples whose s.d. among pyrosequencing triplicates was below 0.20. Similar to the results using the entire sample, no differences were detected between CRC patients and controls (Supplementary Figure 1).

DISCUSSION

The presence of allelic imbalances is well known to be widespread throughout the transcriptome and has been associated with cancer risk in some instances (Yan *et al*, 2002; Raval *et al*, 2007; Chen *et al*, 2008). In a previous report, we suggested that ASE of *TGFBR1* confers a substantially increased risk of CRC (odds ratio 8.7), potentially placing ASE of *TGFBR1* among the major contributors to the genetic predisposition to both familial and sporadic CRC (Valle *et al*, 2008). The main significance of those findings pertains to early detection and prevention of CRC, therefore requiring validation in larger series and different populations for future implementation in clinical practice. Here, using a more robust technique for ASE determination, studying Ashkenazi Jewish and Caucasian populations, and using different sources of non-tumoral genetic material, we identified no differences in the degree or frequency of ASE of *TGFBR1* between CRC patients and controls, discarding its role in CRC predisposition. Our results are supported by the study by Guda *et al* (2009), in which it was concluded that ASE of *TGFBR1* is unlikely to be the major driver of linkage in some colon neoplasia families to the 9q22.2–31.2 region, in which *TGFBR1* is located, and that ASE is not associated with sporadic CRC ($n=44$). Recently, Carvajal-Carmona *et al* (2010) reported no evidence of genetic variation at *TGFBR1* as a predisposing factor for CRC and found no increased level of *TGFBR1* ASE in 24 familial CRC patients compared with 45 informative controls. In fact, ASE turned out to be more prevalent among controls than among cases.

Very recently, two additional studies on ASE of *TGFBR1* were published. In the first study, ASE of *TGFBR1*, assessed by SNaPshot, was found in approximately 10% of CRC patients (15% of informative patients), agreeing with our initial report; however, no controls were included for comparison (Pasche *et al*, 2010). In the second study, where ASE was measured by pyrosequencing, 109 informative cases and 125 informative controls were studied. No differences were identified when ASE was considered as a binary variable; however, when treated with a continuous variable, ASE was significantly higher in cases than in controls. However, the differences identified between CRC patients and controls were very subtle and definitely not useful for cancer risk assessment (Tomsic *et al*, 2010). Table 1 shows the main

Table 1 Summary of the characteristics and results obtained in different studies on ASE of *TGFBR1* and CRC risk

Study	Population	Nucleic acid source	Methodology	Allelic markers	Informative cases; controls	ASE: Cases vs controls ^a	
						Quantitative	Binary (ASE vs non-ASE)
Valle et al (2008)	C ~90% O ~10%	Peripheral blood	SNaPshot	rs334348 rs7871490 rs334349 rs1590	138; 105	$P=0.12$ Wilcoxon test $P=0.02$ Permutation test	Cutoff: $<0.67, \geq 1.5$ (ROC analysis) ASE: 29/138; 3/105 $P=7.66 \times 10^{-5}$
Guda et al (2009)	C ~89% AA ~9% O ~2%	Lymphoblastoid cell line (EBV) Normal colon mucosa	Pyrosequencing	rs868 rs334348 rs334349 rs420549 rs1590	Familial: 46; 17 Sporadic: 44; 0	N/A	Cutoff: $<0.67, \geq 1.5$ (Valle et al, 2008) Familial ASE: 2/46; 0/17 Sporadic ASE: 0/44
Carvajal-Carmona et al (2010)	C ~100%	Lymphoblastoid cell line (EBV)	Genescan SNaPshot	*6A/9A rs1590	Familial: 24; 45	$P=0.09$ Wilcoxon test ^b $P=0.13$ t-test ^b	Cutoff: $<0.67, \geq 1.5$ (Valle et al, 2008) ASE: 7/24; 12/45 $P=0.83$
Pasche et al (2010)	C ~83% AA ~12% O ~5%	Lymphoblastoid cell line (EBV)	SNaPshot	rs334348 rs7871490 rs334349 rs1590	74; 0	N/A	Cutoff: $<0.67, \geq 1.5$ (Valle et al, 2008) ASE: 11/74
Tomsic et al (2010)	C ~92% AA ~8%	Peripheral blood	Pyrosequencing	rs868 rs334348 rs334349 rs420549 rs1590	109; 125 ^c	$P=0.009; 0.006$ Wilcoxon test $P=0.081; 0.077$ Permutation test $P=0.003; 0.007$ Permutation (median)	Cutoff: $<0.67, \geq 1.5$ (Valle et al, 2008) ASE: 2/109; 2/125 Cutoff: $<0.9, \geq 1.1$ (ROC analysis) ASE: 51/109; 39/125 $P=0.06$
This study	Ash 100% C 100%	Uncultured lymphocytes (Ash) Normal colon mucosa (C)	Pyrosequencing	rs334349 rs7850895 rs420549 rs1590	Ash: 96; 90 C: 75; 0	Ash: $P=0.86$ Wilcoxon (median) Ash+C: $P=0.48$ Wilcoxon (median)	Cutoff: $<0.67, \geq 1.5$ (Valle et al, 2008) Ash ASE: 0/96; 2/90 Ash+C ASE: 0/171; 2/90 Cutoff: $<0.68, \geq 1.31$ (median $\pm 2 \times$ s.d.) Ash ASE: 1/96; 2/90 $P=0.52$ Ash+C ASE: 4/171; 2/90 $P=0.52$

Abbreviations: AA = African-American; ASE = allele-specific expression; Ash = Ashkenazi Jewish; C = Caucasian; CRC = colorectal cancer; EBV = Epstein–Barr virus; NA = not available; O = Others; ROC = receiver operating characteristic. ^aASE measured as $(A_{cDNA}/B_{cDNA})/(A_{gDNA}/B_{gDNA})$, A being the common allele and B the rare allele. ^bASE trend was more pronounced in the controls than in the cases. ^cForty-nine cases were the same as in Valle et al (2008).

characteristics and results of previous studies focused on the role of ASE of *TGFBR1* in CRC risk.

Several features differentiate the original and present studies, and important consequences might have derived from these differences. The use of different assays for ASE determination and the exclusion of cases that showed high variability among replicates (exclusion of samples with low-quality source RNA shown in Supplementary Material) have likely increased the robustness of our study. Tomsic et al (2010) also found that the SNaPshot technology used for ASE determination gave inconsistent results, as evidenced by considerably larger standard deviations compared with pyrosequencing, and that high RNA quality is essential for reproducibility of ASE.

Some SNP markers and sources of nucleic acids used were also different. The rs7871490 SNP, located in the 3'-UTR of *TGFBR1*, was used as a marker in the original study (Valle et al, 2008), but not in the present one. We previously found that the marker was very useful because it allowed us to significantly increase the number of informative individuals from 40 to 55–60%. This SNP is located in a region of repetitive sequence, 5'-GGGGGTTTTTTT TTGTTTTTTTTT[G/T]TTGTTGTTGTTTTGGGCCATTTCT-3', which might have affected the correct performance of SNaPshot owing to the design and molecular basis of the technique. When

excluding all individuals from the original study whose ASE value was only based on the results obtained from the analysis of rs7871490 (16 out of 29 CRC patients and 3 out of 3 controls with ASE values >1.5), the proportion of ASE in informative CRC patients drops from 21% (29 out of 138) to 13% (13 out of 97) and in informative controls from 3% (3 out of 105) to 0% (0 out of 76). Likely because of the repetitive sequences in the flanking region of rs7871490, we were not able to design a pyrosequencing assay, which precluded a direct comparison between SNaPshot and pyrosequencing. A subset of individuals within this group belonged to the so-called 'group 2', which was characterised by a particular haplotype significantly over-represented among ASE CRC patients (Valle et al, 2008). To increase the number of informative individuals, the rs420549 and rs7850895 allelic markers were included in this study. This resulted in an increase of 22 out of 75 (29%) informative BCCS CRC patients (three of which showed ASE), 22 out of 96 (23%) informative MECC CRC patients (one showed ASE) and 28 out of 90 (31%) informative MECC controls (two showed ASE). In short, all ASE individuals were informative for only either rs420549 or rs7850895. This observation, together with what has been discussed above about rs7871490, suggests that ASE might be more common among individuals who carry minor alleles for specific *TGFBR1* SNPs, and

therefore might be more or less frequent depending on the panel of SNP markers used to define informative individuals. Nevertheless, our results point to a similar frequency of ASE among cases and controls.

The possibility of ASE being tissue specific has been suggested previously (Cowles *et al*, 2002; Wilkins *et al*, 2007). This was one concern that arose in the paper from Guda *et al* (2009), in which the sources of nucleic acids for ASE determination were EBV-transformed cultured lymphocytes and normal mucosae from CRC patients, in contrast to the total blood used in our initial study (Valle *et al*, 2008). Carvajal-Carmona *et al* (2010) also employed lymphoblastoid cell lines. It is still unknown whether EBV transformation and/or cell culture alter the allelic expression of genes. Here, we obtained uncultured blood lymphocytes from the MECC series, which may well correlate with the total blood used in our previous study (Valle *et al*, 2008) or by Tomsic *et al* (2010), and normal mucosae from BCCS patients, which can be compared with the sporadic cases reported in the series from Guda *et al* (2009). These results suggest that different (uncultured) biological sources of genetic material for the determination of *TGFBR1* ASE can be used without distinction.

In the original report, a mostly Caucasian population from Central Ohio was evaluated, whereas Ashkenazi Israeli and Caucasian Spanish populations were studied here. The fact that ASE first seemed to be heavily dependent on allele frequencies left open the possibility of inter-ethnic variation. The degree of SNP informativity was different between the two populations (20% MECC vs 36% BCCS for rs334349 and rs1590; 14% MECC vs 36% BCCS for rs420549; and 4% MECC vs 10% BCCS for rs7850895); however, no differences were detected in the level of ASE between the two populations. Unfortunately, we did not have access to all types of samples (normal mucosae and lymphocytes) from the same individuals or populations; therefore, there remains a certain degree of uncertainty about tissue and ethnic variability.

REFERENCES

- Carvajal-Carmona LG, Churchman M, Bonilla C, Walther A, Lefevre JH, Kerr D, Dunlop M, Houlston R, Bodmer WF, Tomlinson I. Comprehensive assessment of variation at the transforming growth factor beta type 1 receptor locus and colorectal cancer predisposition. *Proc Natl Acad Sci USA* (2010); **107**: 7858–7862
- Chen X, Weaver J, Bove BA, Vanderveer LA, Weil SC, Miron A, Daly MB, Godwin AK (2008) Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. *Hum Mol Genet* **17**: 1336–1348
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. *Nat Genet* **32**: 432–437
- de la Chapelle A (2004) Genetic predisposition to colorectal cancer. *Nat Rev Cancer* **4**: 769–780
- Guda K, Natale L, Lutterbaugh J, Wiesner GL, Lewis S, Tanner SM, Tomsic J, Valle L, de la Chapelle A, Elston RC, Willis J, Markowitz SD (2009) Infrequent detection of germline allele-specific expression of *TGFBR1* in lymphoblasts and tissues of colon cancer patients. *Cancer Res* **69**: 4959–4961
- Hemminki K, Forsti A, Lorenzo Bermejo J (2009) Surveying the genomic landscape of colorectal cancer. *Am J Gastroenterol* **104**: 789–790
- Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, Bishop T, Thomas H, Houlston R, Tomlinson I (2006) Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. *Cancer Res* **66**: 5003–5006
- Landi S, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, Capella G, Canzian F (2003) Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer. *Cancer Res* **63**: 3560–3566
- Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR (2009) Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* **76**: 1–18
- Pasche B, Kaklamani V, Hou N, Young T, Rademaker A, Peterlongo P, Ellis N, Offit K, Caldes T, Reiss M, Zheng T (2004) *TGFBR1**6A and cancer: a meta-analysis of 12 case-control studies. *J Clin Oncol* **22**: 756–758
- Pasche B, Wisinski KB, Sadim M, Kaklamani V, Pennison MJ, Zeng Q, Bellam N, Zimmerman J, Yi N, Zhang K, Baron J, Stram DO, Hayes MG (2010) Constitutively decreased *TGFBR1* allelic expression is a common finding in colorectal cancer and is associated with three *TGFBR1* SNPs. *J Exp Clin Cancer Res* **29**: 57
- Poynter JN, Gruber SB, Higgins PD, Almog R, Bonner JD, Rennert HS, Low M, Greenson JK, Rennert G (2005) Statins and the risk of colorectal cancer. *N Engl J Med* **352**: 2184–2192
- Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen SS, 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**: 879–890
- Skoglund J, Djureinovic T, Zhou XL, Vandrovцова J, Renkonen E, Iselius L, Bisgaard ML, Peltomaki P, Lindblom A (2006) Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32–31.1. *J Med Genet* **43**: e7
- Skoglund J, Song B, Dalen J, Dedorson S, Edler D, Hjern F, Holm J, Lenander C, Lindfors U, Lundqvist N, Olivecrona H, Olsson L, Pahlman L, Rutegard J, Smedh K, Tornqvist A, Houlston RS,

In conclusion, the improved determination of ASE of *TGFBR1* achieved by pyrosequencing revealed no differences between CRC cases and controls, in both Caucasian and Ashkenazi populations. The sample size in ASE studies is highly relevant owing to their dependence on marker informativity and to the difficulties associated with collection of high-quality germline RNA. Finally, the use of different sources of non-tumour nucleic acids for ASE determination adds consistency to our results. However, the lack of informativity for transcribed SNPs in a substantial proportion of individuals complicates the task to assess the extent of germline ASE of *TGFBR1* in CRC. New technological advances that allow the measurement of allelic imbalances in a more precise and informative manner will be of substantial importance to provide a definitive answer to the real extent of ASE at *TGFBR1* in CRC patients.

ACKNOWLEDGEMENTS

We thank Dr Albert de la Chapelle for helpful discussions and Gemma Aiza for technical support. This work was partly funded by the Spanish Ministry of Science and Innovation (BFU2009-10281 to LV and SAF2009-07319 to GC); the Scientific Foundation of Asociación Española Contra el Cáncer; the National Institutes of Health (R01 CA81488-04 to SBG); the University of Michigan's Cancer Center (5 P30 CA46592 to SBG); Spanish Networks RTICCC (RD06/0020/1050); Gastroenterology Foundation Dr Francisco Vilardeell (F05-01); and Carlos III Health Institute (through Acción Transversal del Cáncer, CIBERESP and Grants FIS PI08/1635, FIS PI08/1359, FIS 06/0545 and FIS 05/1006). NS is supported by a fellowship and LV holds a contract 'Ramón y Cajal', both from the Spanish Ministry of Science and Innovation.

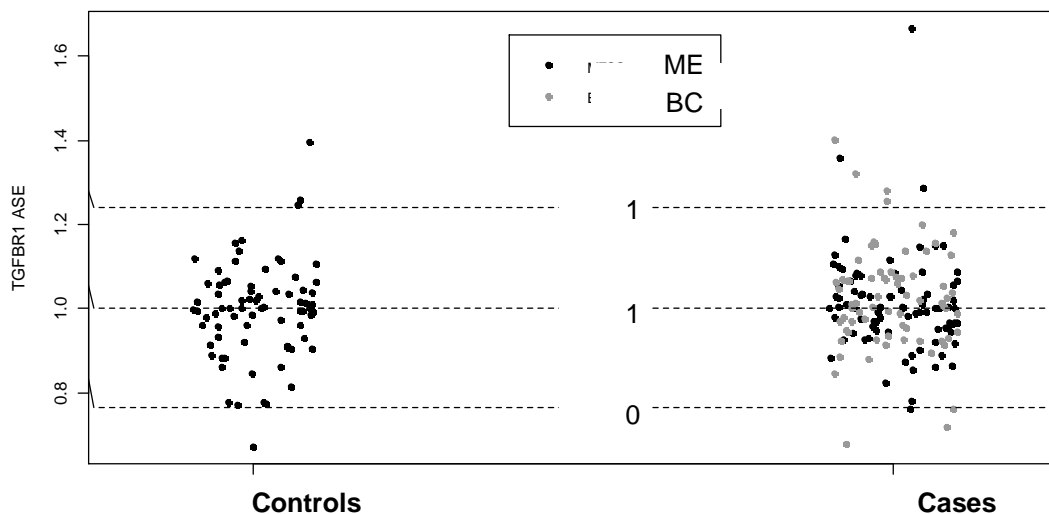
Supplementary Information accompanies the paper on British Journal of Cancer website (<http://www.nature.com/bjc>)

- Lindblom A (2007) Lack of an association between the *TGFBR1**6A variant and colorectal cancer risk. *Clin Cancer Res* **13**: 3748–3752
- Tomsic J, Guda K, Liyanarachchi S, Hampel H, Natale L, Markowitz SD, Tanner SM, de la Chapelle A. Allele-specific expression of *TGFBR1* in colon cancer patients. *Carcinogenesis* (2010); **31**(10): 1800–1804
- Valle L, Serena-Acedo T, Liyanarachchi S, Hampel H, Comeras I, Li Z, Zeng Q, Zhang HT, Pennison MJ, Sadim M, Pasche B, Tanner SM, de la Chapelle A (2008) Germline allele-specific expression of *TGFBR1* confers an increased risk of colorectal cancer. *Science* **321**: 1361–1365
- Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, Markowitz SD (2003) A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2–31.2. *Proc Natl Acad Sci USA* **100**: 12961–12965
- Wilkins JM, Southam L, Price AJ, Mustafa Z, Carr A, Loughlin J (2007) Extreme context specificity in differential allelic expression. *Hum Mol Genet* **16**: 537–546
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B (2007) The genomic landscapes of human breast and colorectal cancers. *Science* **318**: 1108–1113
- Yan H, Dobbie Z, Gruber SB, Markowitz S, Romans K, Giardiello FM, Kinzler KW, Vogelstein B (2002) Small changes in expression affect predisposition to tumorigenesis. *Nat Genet* **30**: 25–26
- Zeng Q, Phukan S, Xu Y, Sadim M, Rosman DS, Pennison M, Liao J, Yang GY, Huang CC, Valle L, Di Cristofano A, de la Chapelle A, Pasche B (2009) *Tgfr1* haploinsufficiency is a potent modifier of colorectal cancer development. *Cancer Res* **69**: 678–686

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Results of the pilot study with samples from 10 informative CRC patients. ASE results obtained with the four 3'UTR SNPs used, an overall ASE value calculated as the median of the ASE values for the informative SNPs, and a standard deviation (s.d) value that indicates the variability among different SNPs, are shown for both SNaPshot and pyrosequencing assays.

Samples	BCCS1	BCCS2	BCCS3	BCCS4	BCCS5	BCCS6	BCCS7	BCCS8	MECC1	MECC2
SNaPshot										
rs334349	1.37	NV	0.92		0.79		0.91	1.06		0.95
rs1590	0.94	NV	0.61		0.54		0.74	0.72		0.65
rs420549	1.06	NV	1.05	1.06	1.29			1.35	1.31	1.05
rs7850895						NV				
ASE (median)	1.06		0.92	1.06	0.79		0.83	1.06	1.31	0.95
s.d.	0.22		0.23	-	0.38		0.12	0.32	-	0.25
Pyrosequencing										
rs334349	1.04	0.81	1.03		1.08		1.08	1.05		0.98
rs1590	1.01	NV	1.09		0.90		0.98	0.85		0.67
rs420549	1.00	NV	1.03	1.08	1.04			1.13	0.85	1.24
rs7850895						0.97				
ASE (median)	1.01	0.81	1.03	1.08	1.04	0.97	1.03	1.05	0.85	0.98
s.d.	0.02	-	0.03		0.09	-	0.07	0.14	-	0.29



Supplementary Figure 1. TGFBR1 ASE distribution in 156 CRC patients (70 MECC (black dots) and 86 BCCS (grey dots) CRC patients) and 75 controls. The median and cutoff points, defined as the median \pm 2*s.d. of controls, used to categorize ASE are indicated as discontinuous lines. Those samples with s.d. among pyrosequencing triplicates above 0.20 were excluded from the analysis. No differences were detected between MECC cases and MECC controls (median difference -0.003; 95%CI -0.032 to 0.025; $P=0.81$); between MECC cases and BCCS cases (median difference 0.016; 95%CI -0.015 to 0.047; $P=0.32$); and between all cases and controls (median difference 0.010; 95% CI -0.038 to 0.015; $P=0.41$). Also, no statistically significant differences were detected among all groups of cases and controls when ASE was categorized as a binary variable.

Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome

Fernando Bellido, Elisabet Guinó, Shantie Jagmohan-Changur, **Nuria Seguí**, Marta Pineda, Matilde Navarro, Conxi Lázaro, Ignacio Blanco, Hans FA Vasen, Víctor Moreno, Gabriel Capellá, Juul T Wijnen y Laura Valle.

European Journal of Human Genetics 21(5):511-6; 2013

Resumen:

El síndrome de Lynch es una forma hereditaria de predisposición al CCR causado por mutaciones en la línea germinal en los genes MMR. Se ha observado una gran variabilidad inter- e intra-familiar en el riesgo de cáncer, lo que sugiere la participación de factores modificadores que modulan el riesgo individual de los pacientes con SL. En este estudio hemos evaluado la participación de la variante rs2075786, situada en el intrón 10 del gen de la telomerasa (*hTERT*), como posible modificador del riesgo de cáncer en SL. Estudiamos un total de 147 individuos afectados y 108 no afectados portadores de mutación en los genes MMR, pertenecientes a 101 familias con SL españolas. Asimismo, también se analizaron 201 individuos afectados y 474 sanos portadores de mutación en los genes MMR de 127 familias con SL holandesas. El estudio de la serie española reveló que el alelo minoritario de la variante (A) confiere un riesgo incrementado a edades tempranas. Estos resultados fueron confirmados con el análisis de la serie holandesa, donde los portadores de la variante mostraron un incremento del riesgo de cáncer antes de los 45 años, especialmente en homocigosis (riesgo relativo_{LS<45AA} = 2.90; intervalo de confianza 95% = 1,02-8,26). Estudios *in silico* de función predicen que esta variante inactiva un sitio de unión de transcripción para un receptor de retinoides, de manera que podría estar provocando la activación temprana de la telomerasa, y consecuentemente la aceleración de la carcinogénesis. Asimismo, hemos observado que los afectados de SL portadores del genotipo AA tienen los telómeros más cortos que aquellos con genotipo GG. Por tanto, nuestros resultados sugieren que el polimorfismo rs2075786 actúa como modificador genético de la edad de desarrollo de los tumores en el SL.

ARTICLE

Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome

Fernando Bellido^{1,7}, Elisabet Guinó^{2,7}, Shantie Jagmohan-Changur³, Nuria Seguí¹, Marta Pineda¹, Matilde Navarro¹, Conxi Lázaro¹, Ignacio Blanco¹, Hans FA Vasen⁴, Victor Moreno^{2,5}, Gabriel Capellá¹, Juul T Wijnen^{3,6} and Laura Valle^{*,1}

Lynch syndrome (LS) is an inherited cancer-predisposing disorder caused by germline mutations in the mismatch repair (*MMR*) genes. The high variability in individual cancer risk observed among LS patients suggests the existence of modifying factors. Identifying genetic modifiers of risk could help implement personalized surveillance programs based on predicted cancer risks. Here we evaluate the role of the telomerase (*hTERT*) rs2075786 SNP as a cancer-risk modifier in LS, studying 255 and 675 *MMR* gene mutation carriers from Spain and the Netherlands, respectively. The study of the Spanish sample revealed that the minor allele (A) confers increased cancer risk at an early age. The analysis of the Dutch sample confirmed the association of the A allele, especially in homozygosity, with increased cancer risk in mutation carriers under the age of 45 (relative risk_{LSca<45_AA} = 2.90; 95% confidence interval = 1.02–8.26). Rs2075786 is associated with colorectal cancer (CRC) risk neither in the general population nor in non-Lynch CRC families. *In silico* studies predicted that the SNP causes the disruption of a transcription binding site for a retinoid receptor, retinoid X receptor alpha, probably causing early telomerase activation and therefore accelerated carcinogenesis. Notably, cancer-affected LS patients with the AA genotype have shorter telomeres than those with GG. In conclusion, *MMR* gene mutation carriers with *hTERT* rs2075786 are at high risk to develop a LS-related tumor at an early age. Cancer-preventive measures and stricter cancer surveillance at early ages might help prevent or early detect cancer in these mutation carriers.

European Journal of Human Genetics (2013) 21, 511–516; doi:10.1038/ejhg.2012.204; published online 5 September 2012

Keywords: Lynch syndrome; hereditary non-polyposis colorectal cancer; cancer-risk modifier; *hTERT*; telomerase; rs2075786

INTRODUCTION

Lynch syndrome (LS) (MIM 120435) is an autosomal-dominant inherited disorder caused by germline mutations in the DNA mismatch repair (*MMR*) genes. It is characterized by early-onset colorectal cancer (CRC) and an increased risk of upper gastrointestinal, urologic and gynecologic cancers.¹ Several studies have estimated the cancer risks associated with mutations in the *MMR* genes, but these estimates vary substantially across the studies (20–80% lifetime risk).^{2–9}

There is a considerable variation in LS expression and no obvious gene-specific genotype/phenotype correlations have been demonstrated nor does there appear to be any relationship between the location of a mutation and type of disease. In addition to environmental factors, there is evidence suggesting the existence of genetic factors that somehow explain the variability in individual cancer risk.¹⁰ Identifying the genetic modifiers of risk can lead to an efficient stratification of mutation carriers based on their predicted risk, implying thus, a more appropriate clinical management based on personalized surveillance programs.

Most of the attempts to identify cancer-risk modifiers in LS have been based on the study of candidate genes and most of them have

not been validated when tested in larger sample sizes.¹¹ In contrast, two genetic variants previously identified in CRC genome-wide association studies (GWAS), rs16892766 and rs3802842, might modify cancer risk in LS families.^{12–14}

Telomeres are located at the end of chromosomes and protect the chromosome ends from nucleolytic degradation, end-to-end fusions and irregular recombination, being thus critical for genome stability and integrity. Telomeres progressively shorten with each cell replication cycle. Telomere length anomaly appears to be one of the earliest and most prevalent genetic alterations in the multistep process of malignant transformation.^{15,16} Telomerase catalyzes the *de novo* addition of telomeric repeat sequences onto chromosome ends and it is usually inactive in normal somatic cells, whereas its expression has been linked to increased susceptibility to tumorigenesis.^{17–19} Genetic variants located in genes involved in telomere maintenance, and in particular in *hTERT* (MIM 187270), the gene encoding the catalytic subunit of telomerase, have been associated with increased risk to cancer.^{20–27} Among the better characterized cancer variants, the *hTERT* rs2075786 (c.2654 + 269G/A) SNP has been associated with lung cancer risk.^{28,29} However, no studies have reported any

¹Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, L'Hospitalet de Llobregat, Spain; ²Unit of Biomarkers and Susceptibility, Catalan Institute of Oncology, IDIBELL and CIBERESP, L'Hospitalet de Llobregat, Spain; ³Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; ⁴Department of Gastroenterology and Hepatology, Leiden University Medical Center, The Netherlands Foundation for the Detection of Hereditary Tumors, Leiden, The Netherlands; ⁵Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, L'Hospitalet de Llobregat, Spain; ⁶Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

⁷These authors contributed equally to this work.

*Correspondence: Dr L Valle, Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Av. Gran Via 199-203, L'Hospitalet de Llobregat, 08908 Barcelona, Spain. Tel: +34 93 2607145; Fax: +34 93 2607466; E-mail: lvalle@iconcologia.net

Received 7 June 2012; revised 2 August 2012; accepted 7 August 2012; published online 5 September 2012

Table 1 Characteristics of the *MMR* mutation carriers included in the study

		Spain			The Netherlands		
		Affected n (%)	Unaffected n (%)	Total n (Fam)	Affected n (%)	Unaffected n (%)	Total n (Fam)
No. of <i>MMR</i> carriers	LS cancer	147	108	255 (101)	201	474	675 (127)
	CRC	134	121		146	529	
<i>Gene</i>							
<i>MLH1</i>	LS cancer	82 (55.8)	69 (63.9)	151 (54)	73 (36.3)	175 (36.9)	248 (43)
	CRC	77 (57.5)	74 (61.2)		62 (42.5)	186 (35.2)	
<i>MSH2</i>	LS cancer	49 (33.3)	30 (27.8)	79 (34)	85 (42.3)	166 (35.0)	251 (57)
	CRC	43 (32.1)	36 (29.8)		61 (41.8)	190 (35.9)	
<i>MSH6</i>	LS cancer	11 (7.5)	9 (8.3)	20 (9)	42 (20.9)	133 (28.1)	175 (26)
	CRC	9 (6.7)	11 (9.1)		22 (15.1)	153 (28.9)	
<i>PMS2</i>	LS cancer	5 (3.4)	0 (0.0)	5 (4)	1 (0.5)	0 (0.0)	1 (1)
	CRC	5 (3.7)	0 (0.0)		1 (0.7)	0 (0.0)	
<i>Gender</i>							
Male	LS cancer	75 (50.3)	43 (41.7)	118	87 (43.3)	213 (44.9)	300
	CRC	73 (54.5)	45 (37.9)		79 (54.1)	221 (41.8)	
Female	LS cancer	73 (49.7)	64 (58.3)	137	114 (56.7)	261 (55.1)	375
	CRC	61 (45.5)	76 (62.1)		67 (45.9)	308 (58.2)	
Age ^a (average ± SD)	LS cancer	44.6 ± 12.0	36.8 ± 13.1		45.1 ± 10.8	46.2 ± 13.6	
	CRC	44.4 ± 12.3	39.1 ± 14.4		44.8 ± 11.5	47.6 ± 14.1	
Polypectomy	LS cancer	NA	NA		71 (42.8)	95 (57.2)	166
	CRC				58 (34.9)	108 (65.1)	101

Abbreviations: CRC, colorectal cancer; Fam, number of families; LS cancer, colon, rectum, endometrium, stomach, ovary, urinary tract and small intestine; *MMR*, mismatch repair; NA, not available information.

^aAge at diagnosis for cancer-affected subjects, or age at last follow-up or at death for unaffected subjects.

association between variants located at the *hTERT* locus and the risk of CRC.

Here we evaluate *hTERT* rs2075786 as a modifier of cancer in LS patients in 255 *MMR* gene mutation carriers from Spain. Also, 675 mutation carriers from the Netherlands were analyzed. We also assessed its role as cancer-risk factor in a population-based case-control series and in non-LS familial cases.

MATERIALS AND METHODS

Subjects

Spanish sample. A total of 255 individuals from 101 LS families whose mutation carrier statuses were known were included in the study. They were assessed through the Hereditary Cancer Program of the Catalan Institute of Oncology from 1998 to 2010. All families were of Caucasian origin. Informed consent was obtained from all individuals. *MMR* mutation analysis was performed on genomic DNA extracted from peripheral blood lymphocytes. Large genomic alterations were studied using multiplex ligation-dependent probe amplification (SALSA MLPA Kits, MRC-Holland, Amsterdam, The Netherlands). Mutation screening was performed by direct sequencing after PCR amplification (primers and conditions available upon request). In all, 147 (57.6%) *MMR* gene mutation carriers had been diagnosed with a LS-related tumor, of whom 134 with CRC. The clinical characteristics of the subjects are detailed in Table 1.

Also, DNA from 277 sporadic CRC cases and 280 controls were obtained from a hospital-based case-control study (Bellvitge Colorectal Cancer Study). Cases were consecutive patients with a first diagnosis of colorectal adenocarcinoma attending the Bellvitge University Hospital in Barcelona. For the group of patients, the male/female ratio is 1.2 and the mean age is 66.7 (range: 23–91), and for the controls, 1.1 and 65.5 (range: 24–92), respectively. Details about the study population were published elsewhere.³⁰

Dutch sample. A total of 675 *MMR* gene mutation carriers from 127 different families from the Dutch LS Registry were studied. Detailed description of the registry, individuals and DNA extraction method was published before.¹²

Information on the subjects has been recently updated and thus, follow-up data added. In all, 201 (29.8%) *MMR* gene mutation carriers had been diagnosed with a LS-related tumor, of whom 146 are with CRC. Table 1 summarizes the updated clinical characteristics of the subjects.

The Dutch case-control series consists of 324 CRC cases and 785 controls. Cases are probands assessed through a familial cancer clinic; therefore, they are suspected of genetic CRC susceptibility but with no germline mutations identified in the *MMR* genes. Regarding controls, 475 are healthy blood donors and 310 are healthy spouses of *MMR* gene mutation carriers.

MMR gene mutation carriers are derived from multiple-case families selected for genetic counseling and testing. In this situation, the disease status clearly affects the likelihood of testing, causing an overrepresentation of affected individuals. This observation is evident in the Spanish but not in the Dutch sample. The Dutch registry has been running longer (it was established in 1985) and follows large and complete LS families, whereas the Spanish series comes from a cancer hospital-based genetic counseling unit established in 1998, where a clear ascertainment bias towards the assessment of cancer-affected family members exists. Because of the differences in affected vs unaffected ratios, the studies were performed independently for the Spanish and Dutch samples.

Genotyping

The rs2075786 SNP was genotyped using the commercially available TaqMan assay C_15824034_10 (Applied Biosystems Inc., Foster City, CA, USA) following the manufacturer's instructions. Reactions were performed in duplicate in the LightCycler 480 real-time PCR detection system (Roche Diagnostics GmbH, Mannheim, Germany). Genotype calling was performed automatically by the LightCycler 480 II software (Roche Diagnostics GmbH, Mannheim, Germany). Genotyping failed in 78 of the 2596 (3%) samples included in the study.

In silico analyses

Public genome browsers (Ensembl, NCBI and USCS genome browsers) were checked out to investigate the existence of transcription of noncoding RNAs from the region where rs2075786 is located. The presence

of microRNA-binding sites was studied using the microRNA target-prediction algorithms provided by <http://www.microrna.org>. The strength of the pseudoexon splice sites was analyzed using three different prediction programs: www.fruitfly.org/seq_tools/splice.html, http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html and <http://ast.bioinfo.tau.ac.il/SpliceSiteFrame.html>, as previously described.³¹ Putative binding sites for transcription factors were identified using PROMO 3.0, a web-based program that employs the TRANSFAC database version 8.3 to construct specific binding-site weight matrices for prediction of transcription factor-binding sites.^{32,33}

Relative telomere length assessment

Telomere length quantification was performed using a monochrome multiplex quantitative PCR method, as described by Cawthon with slight modifications (available upon request).³⁴

Statistical analyses

Evaluation of the deviation of rs2075786 from the expected Hardy–Weinberg equilibrium was performed in controls from each population. The SNP was in Hardy–Weinberg equilibrium in controls from both populations ($P_{\text{Spanish}} = 0.48$; $P_{\text{Dutch}} = 0.93$).

The association between genotypes and risk of CRC or LS-related tumors was evaluated using conditional logistic regression adjusting for familial clustering. The end points considered were: age at diagnosis of the LS-related cancer or CRC for affected individuals and age of last follow-up for unaffected individuals. Likelihood ratio tests were applied to assess statistical significance. In general, relative risks (RRs), 95% confidence intervals (CIs) and associated *P*-values were estimated for different models of inheritance and the best model was selected using the Akaike information criterion. The interaction of genotypes with age stratified at 45 years was assessed to identify anticipation effects. For case–control studies, odds ratios (ORs) and 95% CI were calculated using logistic regression adjusted for gender. Kruskal–Wallis rank-based test was used to compare telomere length among rs2075786 genotype groups, adjusting for age and gender. Statistical analyses were performed using R.

RESULTS

Spanish study

No differences in rs2075786 genotype distributions were observed when comparing *MMR* gene mutation carriers affected with cancer, either LS-related or only CRC, and unaffected mutation carriers (Supplementary Table S1).

We further analyzed rs2075786 genotype distributions within two age intervals in the cancer group. For this aim, we established a cut off of 45 years of age based on the median age of all individuals included

in the study. *MMR* gene mutation carriers diagnosed with cancer before 45 years of age were compared with those diagnosed after 45 years of age and those who were cancer-free after 45 years of age. Under this condition, the A allele conferred an increased risk of cancer in LS carriers under the age of 45 compared with those diagnosed with cancer or cancer-free after the age of 45 (Table 2).

Although borderline and not fulfilling the required standards for declaring an association, the results obtained in the Spanish sample led us to analyze another series (Dutch study) in a more comprehensive manner, thanks to the larger sample size and the availability of follow-up and polypectomy information.

Dutch study

A total of 675 *MMR* gene mutation carriers from the Dutch hereditary CRC registry were included in the analysis. As previously observed for the Spanish *MMR* gene mutation carriers, no effect was detected on the general risk of cancer between cancer-affected and unaffected individuals (Supplementary Table S2).

To study the effect on cancer risk related with age intervals, cancer-affected vs unaffected *MMR* gene mutation carriers were compared within the groups under 45 years and within the groups ≥ 45 . For this purpose, the two age groups were created as: group age < 45 , *MMR* gene mutation carriers diagnosed with cancer under the age of 45 and unaffected *MMR* gene mutation carriers under age 45 and group age ≥ 45 , cancer-affected and unaffected *MMR* gene mutation carriers > 45 . Because of the larger sample size and the fact that unaffected carriers had a similar or even higher average age than cancer-affected carriers (Table 1), this type of distribution in age groups could be performed for the Dutch *MMR* gene mutation carriers.

In this series, the AA genotype confers an increased risk of LS-related cancers in individuals < 45 years of age (RR = 2.76; 95% CI = 1.02–7.50). When only CRC was considered, an RR of 2.46 (96% CI = 0.78–7.82) was found for the AA homozygotes, but this observation was not statistically significant, probably because of the effect on the risk of all LS-related tumors. No association of rs2075786 with cancer risk was observed in the ≥ 45 groups (Table 3).

MMR gene mutation carriers are usually under strict clinical surveillance and undergo colonoscopies on a routine basis, implying that whenever detected, polyps are removed. As polypectomy affects subsequent CRC risk, we included this information in the analysis. Within the group age < 45 , all unaffected subjects with a polypectomy were excluded. Moreover, within the group age ≥ 45 , all cancer-affected subjects with a polypectomy were performed under the age of 45 and all unaffected individuals with polypectomies were excluded from the analysis. A total of 100 individuals were removed. In this scenario, the increased cancer risk conferred by AA in the < 45 group is maintained or even increased (RR_{LSca < 45_AA} = 2.90; 95% CI = 1.02–8.26) (Table 4). Slight random differences from the previous analysis dependent on the genotype of the cases excluded were observed.

No effect of gender on cancer risks was observed (data not shown). Regarding the *MMR* gene affected, the analysis was performed considering *MLH1* and *MSH2* separately (Supplementary Table S3). Although differences did not reach significance and larger series should be studied to draw definitive conclusions, the results suggest that the effect of rs2075786 on cancer risk occurs for both *MLH1* and *MSH2* mutation carriers under the age of 45 (RR_{MLH1 < 45_AA} = 3.86; 95% CI = 0.70–21.33 and RR_{MSH2 < 45_AA} = 2.36; 95% CI = 0.42–13.39).

Table 2 Genotype distributions within age intervals in Spanish *MMR* mutation carriers

Genotype	LS ca. < 45 n (%)	LS ca. ≥ 45 n (%)	RR (95% CI) ^a	P
GG	31 (38.3)	43 (50.0)		
GA	43 (53.1)	29 (33.7)	4.69 (0.92–23.88)	0.062
AA	7 (8.6)	14 (16.3)	3.76 (0.41–34.31)	0.241
LRT (2 df)				0.126
GA + AA	50 (61.7)	43 (50.0)	4.66 (0.92–23.71)	0.064
LRT (1 df)				0.044
Per allele			1.98 (0.70–5.65)	0.199
LRT (1 df)				0.192

Abbreviations: CI, confidence interval; df, degrees of freedom; LRT, likelihood ratio test; LS ca., colon, rectum, endometrium, stomach, ovary, urinary tract or small intestine cancer; *MMR*, mismatch repair; RR, relative risk.

Young cancer-affected *MMR* mutation carriers (cancer diagnosed before 45 years of age) are compared with *MMR* mutation carriers who developed a LS-associated tumor after 45 years of age or who are cancer-free after 45 years of age.

^aRRs were estimated using conditional logistic regression adjusting for familial clustering.

Table 3 Risk of cancer associated with *hTERT* rs2075786 in Dutch MMR mutation carriers within age intervals (<45 and ≥45 years)

Genotype	LS ca. n (%)	LS ca.-free n (%)	RR (95% CI) ^a	P	CRC n (%)	CRC-free n (%)	RR (95% CI) ^a	P
<i>Group age < 45</i>								
GG	32 (36.8)	112 (47.1)	1		27 (40.3)	113 (46.7)	1	
GA	41 (47.1)	109 (45.8)	1.12 (0.59–2.12)	0.735	30 (44.8)	112 (46.3)	0.92 (0.45–1.91)	0.829
AA	14 (16.1)	17 (7.1)	2.76 (1.02–7.50)	0.046	10 (14.9)	17 (7%)	2.46 (0.78–7.82)	0.126
<i>Group age ≥ 45</i>								
GG	53 (48.6)	94 (42.3)	1		37 (48.7)	114 (42.1)	1	
GA	41 (37.6)	102 (45.9)	0.86 (0.46–1.61)	0.644	30 (39.5)	121 (44.6)	0.93 (0.47–1.83)	0.830
AA	15 (13.8)	26 (11.7)	0.67 (0.27–1.69)	0.398	9 (11.8)	36 (13.3)	0.50 (0.17–1.45)	0.201
LRT (5 df)				0.259				0.397
Interaction with age (χ^2 test)				0.112				0.088
<i>Per allele</i>								
Age < 45			1.46 (0.92–2.31)	0.109			1.31 (0.76–2.23)	0.328
Age ≥ 45			0.84 (0.54–1.28)	0.413			0.76 (0.47–1.22)	0.260
LRT (3 df)				0.163				0.445
Interaction with age (χ^2 test)				0.070				0.120

Abbreviations: CI, confidence interval; df, degrees of freedom; LRT, likelihood ratio test; LS ca., colon, rectum, endometrium, stomach, ovary, urinary tract or small intestine cancer; MMR, mismatch repair; RR: relative risk.

^aRRs were estimated using a conditional logistic regression adjusting for familial clustering.

Table 4 Risk of cancer associated with *hTERT* rs2075786 in Dutch MMR mutation carriers within age intervals (<45 and ≥45 years) considering polypectomies

Genotype	LS ca. n (%)	LS ca.-free n (%)	RR (95% CI) ^a	P	CRC n (%)	CRC-free n (%)	RR (95% CI) ^a	P
<i>Group age < 45</i>								
GG	32 (36.8)	104 (48.6)	1		27 (40.3)	105 (48.4)	1	
GA	41 (47.1)	94 (43.9)	1.19 (0.61–2.34)	0.604	30 (44.8)	96 (44.2)	1.03 (0.48–2.20)	0.938
AA	14 (16.1)	16 (7.5)	2.90 (1.02–8.26)	0.047	10 (14.9)	16 (7.4)	2.28 (0.70–7.40)	0.170
<i>Group age ≥ 45</i>								
GG	50 (48.1)	64 (41.3)	1		35 (48.6)	79 (41.1)	1	
GA	40 (38.5)	74 (47.7)	0.96 (0.45–1.91)	0.898	29 (40.3)	89 (46.4)	1.05 (0.49–2.25)	0.898
AA	14 (13.5)	17 (11.0)	0.52 (0.18–1.54)	0.239	8 (11.1)	24 (12.5)	0.33 (0.10–1.10)	0.070
LRT (5 df)				0.038				0.098
Interaction with age (χ^2 test)				0.082				0.040
<i>Per allele</i>								
Age < 45			1.51 (0.93–2.44)	0.096			1.34 (0.77–2.30)	0.298
Age ≥ 45			0.79 (0.48–1.30)	0.354			0.68 (0.40–1.16)	0.155
LRT (3 df)				0.017				0.118
Interaction with age (χ^2 test)				0.064				0.071

Abbreviations: CI, confidence interval; df, degrees of freedom; LRT, likelihood ratio test; LS ca., colon, rectum, endometrium, stomach, ovary, urinary tract or small intestine cancer; MMR, mismatch repair; RR: relative risk.

^aRRs were estimated using a conditional logistic regression adjusting for familial clustering.

Rs2075786 and CRC risk in the general population and in non-Lynch CRC families

To evaluate whether rs2075786 is associated with CRC risk in the general population and in non-Lynch familial CRC cases, we performed an association study in a case–control series from Spain and in non-Lynch CRC familial cases from the Netherlands. Rs2075786 was in Hardy–Weinberg equilibrium in both the Spanish ($P=0.59$) and Dutch ($P=0.88$) controls. No statistically significant association between rs2074786 and CRC was detected in any of the two situations, with ORs close to 1 (Supplementary Table S4).

Also, when separating by age groups (<45 and ≥45) in the non-Lynch familial CRC study (The Netherlands), no statistically

significant differences were identified in either group (Supplementary Table S5). The age stratification could not be performed for the sporadic case–control series (Spain) because of the scarcity of sporadic CRC patients with <45 years of age present.

In silico studies to unravel rs2075786 functionality

We performed an *in silico* search of potential effects of the variant on gene splicing, transcription factor or microRNA binding, as well as of noncoding RNAs being transcribed from that genomic region. Public databases showed inexistence of noncoding RNAs, and microRNA target-prediction algorithms found no microRNA-binding sites when either allele (A or G) was present. Likewise, gene splicing was not

affected. *In silico* search for putative binding sites of transcription factors revealed that when the rare allele (A) of rs2075786 is present, the elimination of a binding site for retinoid X receptor alpha (RXR- α) is predicted.

Rs2075786 and telomere length

To test the hypothesis that telomere length is compromised by the presence of the A allele of rs2075786, we measured the relative telomere length in blood DNA of the cancer-affected LS patients from Spain. Individuals with the AA genotype have shorter telomeres than individuals with the GG ($P=0.011$) and GA ($P=0.058$) genotypes (Figure 1). No association of relative telomere length with rs2075786 genotypes was found in sporadic CRC patients (Supplementary Figure S1).

DISCUSSION

There remains a considerable variability in LS disease expression that cannot be readily explained by genetic variance occurring solely within the *MMR* genes. Several studies have tried to determine whether there are any genetic modifying factors that could be associated with an increased likelihood of developing cancer. In this study, we show that the AA genotype of the variant rs2075786 of the telomerase gene (*hTERT*) is associated with an increased risk of LS-related cancers in *MMR* gene mutation carriers under 45 years of age, indicating anticipation in the age of onset of cancer. Moreover, when information about polypectomies, a therapeutic intervention that affects CRC risk and might lead to false negatives, was incorporated to the analyses, the results obtained improved, strengthening the observation ($RR_{<45_AA}=2.90$; 95% CI = 1.02–8.26).

The field of cancer-risk modifiers of LS is very active and has provided in the recent years controversial results. First, candidate genes that have some plausible biological role in the disease were evaluated. Except for rare exceptions, the initial findings identified in small-sized samples were not validated in larger series.¹¹ More recently, some polymorphisms identified through GWAS of unselected CRC patients and controls have been shown to modify cancer risk in LS families. Wijnen *et al*¹² first studied six GWAS CRC susceptibility SNPs in 675 *MMR* gene mutation carriers from 127 LS families. They found that two of them, at 8q23.3 and 11q23.1, were associated with CRC risk in mutation carriers, particularly in females.¹² These results were partly confirmed in a cohort of Australian and Polish *MMR* mutation carriers, but were not replicated in a French cohort.^{13,14}

We selected the *hTERT* variant rs2075786 by somehow combining candidate gene and GWAS-based approaches. First, *hTERT* has an important role at very early stages of carcinogenesis, its expression has been linked to increased susceptibility to tumorigenesis, and rs2075786 has been associated with cancer risk.^{15,16,18,19,28,29} And second, through GWAS, genetic variants located in genes involved in telomere maintenance, and in particular in *hTERT*, were also associated with increased risk to cancer.^{20–27} In contrast to previous studies examining the role of modifier genes,¹¹ our study includes two independent series of *MMR* gene mutation carriers, of 255 and 675 individuals, the latest being among the largest series used to study genetic modifiers of LS. In view of our results with rs2075786, it may be relevant to study the role of other *hTERT* polymorphisms as cancer-risk modifiers in LS, in particular those variants with functional implications and/or with strong levels of evidence of being associated with cancer risk.³⁵

In order to unravel the functional effect of rs2075786, *in silico* analysis predicted that when the minor allele (A) is present, a binding

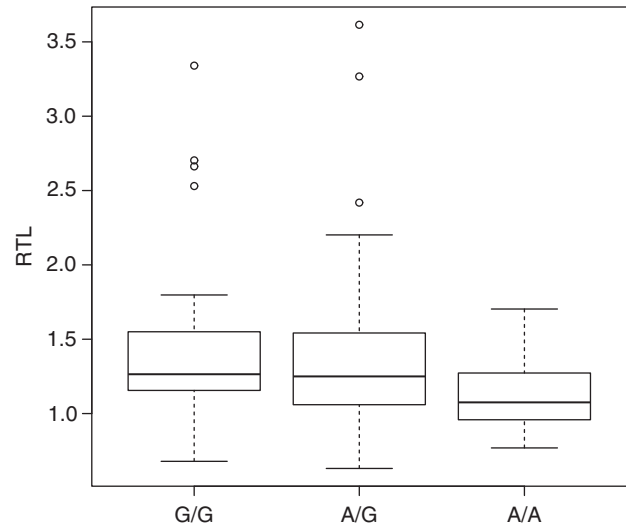


Figure 1 Blood relative telomere length (RTL) in 146 cancer-affected LS patients according to rs2075786 genotype distribution ($n_{GG}=66$, $n_{AG}=64$ and $n_{AA}=16$). The boxes represent the interquartile range of distributions (25th–75th percentile); the horizontal lines within the boxes, the medians; and the vertical lines, the 5th and 95th percentiles.

site for RXR- α is eliminated. Already existent evidence indicates that retinoids, through retinoid receptors, inhibit telomerase activity and downregulate *hTERT* expression.^{36,37} These observations suggest that in the presence of the rs2075786 minor allele A, where a binding site for retinoid receptors is absent, natural retinoids cannot efficiently restrain *hTERT* expression, causing accelerated tumor cell growth.

It is nevertheless intriguing that rs2075786 does not influence CRC risk in sporadic or in other non-Lynch familial cases. It is well known that the carcinogenic pathway of *MMR*-deficient tumors clearly differs from that involved in most sporadic and non-LS familial CRC. It has been observed that *MMR*-deficient cell lines and colon tumors show high mutation frequencies at telomere ends, leading to accelerated telomere shortening.^{38–41} The effect of an increased telomere-shortening rate is likely to require early activation of telomerase in such tumors. As hypothesized above, the presence of the A allele of rs2075786 might imply even earlier activation of telomerase. Although the effect might be very subtle to have an impact in *MMR*-proficient cells, it could trigger important consequences in a *MMR*-deficient context, such as in LS tumors. In this line of research, it may be interesting to evaluate the role of rs2075786 as a risk allele for sporadic CRC with microsatellite instability.

A possible explanation for the rs2075786 acting as modifier of cancer risk at early ages but not later in life might be that the effect the variant has on telomere shortening (Figure 1) is not efficient at late ages where telomeres are already physiologically shortened due to the aging process.⁴² Nevertheless, further functional assays are required to provide experimental evidence about this matter.

When assessing small variations in cancer risk, the issue of personalizing follow-up immediately arises. A potential applicability of rs2075786 genotyping in genetic counseling for clinical management of LS patients can be envisioned. Based on our results, young *MMR* gene mutation carriers with the rs2075786 AA genotype might benefit from an even more intensive clinical surveillance of CRC and other LS-related tumors than the one suggested by the standard LS surveillance protocols. Validation of our observation in additional series is warranted prior to translation to routine clinical practice.

Nevertheless, both *MMR* gene mutation carriers and clinicians should be conscious of the higher risk that rs2075786 carriers have at early ages, observing or even increasing the screening frequency for colorectal, endometrial and other LS-related tumors. Also, lifestyle changes that affect environmental modifier risk factors might help reduce the risk of developing cancer. Likewise, the recent promising results of aspirin-based chemoprevention obtained in LS patients suggest that *MMR* gene mutation carriers with the rs2075786 AA genotype might benefit from this chemopreventive treatment starting at early age.⁴²

In conclusion, we observe an association that might help identify *MMR* gene mutation carriers with higher risk to prematurely develop a LS-associated tumor. Rs2075786 adds up to the already known LS modifiers, increasing thus the opportunity to incorporate the information of modifier factors into clinical practice.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all the people responsible for genetic counseling and genetic testing in hereditary cancer at both Catalan Institute of Oncology and Leiden University Medical Center, and Gemma Aiza for technical support. This work was partly funded by the Spanish Ministry of Science and Innovation (grant BFU2009-10281 and Ramón y Cajal contract, both to LV; and fellowship to FB), the Scientific Foundation of Asociación Española Contra el Cáncer, the Carlos III Health Institute (ISCIIIRETIC: RD06/0020/1051 and RD06/0020/1050; FIS PI08/1635, FIS PI08/1359 and FIS PS09-01037; and fellowship to NS), the Catalan Health Institute and the Autonomous Government of Catalonia (2009SGR290, 2009SGR1489), and CIBERESP (CB07/02/2005).

- Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR: Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medico-legal ramifications. *Clin Genet* 2009; **76**: 1–18.
- Dunlop MG, Farrington SM, Carothers AD *et al*: Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum Mol Genet* 1997; **6**: 105–110.
- Aaltonen LA, Salovaara R, Kristo P *et al*: Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998; **338**: 1481–1487.
- DeFrancisco J, Grady WM: Diagnosis and management of hereditary non-polyposis colon cancer. *Gastrointest Endosc* 2003; **58**: 390–408.
- Quehenberger F, Vasen HF, van Houtwelingen HC: Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. *J Med Genet* 2005; **42**: 491–496.
- Jenkins MA, Baglietto L, Dowty JG *et al*: Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study. *Clin Gastroenterol Hepatol* 2006; **4**: 489–498.
- Alarcon F, Lasset C, Carayol J *et al*: Estimating cancer risk in HNPCC by the GRL method. *Eur J Hum Genet* 2007; **15**: 831–836.
- Choi YH, Cotterchio M, McKeown-Eyssen G *et al*: Penetrance of colorectal cancer among MLH1/MSH2 carriers participating in the colorectal cancer familial registry in Ontario. *Hered Cancer Clin Pract* 2009; **7**: 14.
- Bonadona V, Bonaiti B, Olschwang S *et al*: Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA* 2011; **305**: 2304–2310.
- Antoniou AC, Chenevix-Trench G: Common genetic variants and cancer risk in Mendelian cancer syndromes. *Curr Opin Genet Dev* 2010; **20**: 299–307.
- Scott RJ, Lubinski J: Genetic epidemiology studies in hereditary non-polyposis colorectal cancer. *Methods Mol Biol* 2009; **472**: 89–102.
- Wijnen JT, Brohet RM, van Eijk R *et al*: Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. *Gastroenterology* 2009; **136**: 131–137.
- Talseth-Palmer BA, Brenne IS, Ashton KA *et al*: Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in Lynch syndrome. *J Med Genet* 2010; **48**: 279–284.
- Houille S, Charbonnier F, Houivet E *et al*: Evaluation of Lynch syndrome modifier genes in 748 *MMR* mutation carriers. *Eur J Hum Genet* 2011; **19**: 887–892.
- Londono-Vallejo JA: Telomere length heterogeneity and chromosome instability. *Cancer Lett* 2004; **212**: 135–144.
- Meeker AK, Hicks JL, Iacobuzio-Donahue CA *et al*: Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. *Clin Cancer Res* 2004; **10**: 3317–3326.
- Collins K, Mitchell JR: Telomerase in the human organism. *Oncogene* 2002; **21**: 564–579.
- Blasco MA, Lee HW, Hande MP *et al*: Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* 1997; **91**: 25–34.
- Calado RT, Chen J: Telomerase: not just for the elongation of telomeres. *Bioessays* 2006; **28**: 109–112.
- McKay JD, Hung RJ, Gaborieau V *et al*: Lung cancer susceptibility locus at 5p15.33. *Nat Genet* 2008; **40**: 1404–1406.
- Shete S, Hosking FJ, Robertson LB *et al*: Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* 2009; **41**: 899–904.
- Wrensch M, Jenkins RB, Chang JS *et al*: Variants in the *CDKN2B* and *RTEL1* regions are associated with high-grade glioma susceptibility. *Nat Genet* 2009; **41**: 905–908.
- Jin G, Xu L, Shu Y *et al*: Common genetic variants on 5p15.33 contribute to risk of lung adenocarcinoma in a Chinese population. *Carcinogenesis* 2009; **30**: 987–990.
- Landi MT, Chatterjee N, Yu K *et al*: A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009; **85**: 679–691.
- Rafnar T, Sulem P, Stacey SN *et al*: Sequence variants at the *TERT-CLPTM1L* locus associate with many cancer types. *Nat Genet* 2009; **41**: 221–227.
- Petersen GM, Amundadottir L, Fuchs CS *et al*: A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet* 2010; **42**: 224–228.
- Hofer P, Baierl A, Feik E *et al*: *MNS16A* tandem repeats minisatellite of human telomerase gene: a risk factor for colorectal cancer. *Carcinogenesis* 2011; **32**: 866–871.
- Hosgood 3rd HD, Cawthon R, He X, Chanock S, Lan Q: Genetic variation in telomere maintenance genes, telomere length, and lung cancer susceptibility. *Lung Cancer* 2009; **66**: 157–161.
- Van Dyke AL, Cote ML, Wenzlaff AS *et al*: Chromosome 5p Region SNPs Are Associated with Risk of NSCLC among Women. *J Cancer Epidemiol* 2009; **2009**: 242151.
- 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**: 2101–2108.
- Pros E, Gomez C, Martin T, Fabregas P, Serra E, Lazaro C: Nature and mRNA effect of 282 different *NF1* point mutations: focus on splicing alterations. *Hum Mutat* 2008; **29**: E173–E193.
- Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM: PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 2002; **18**: 333–334.
- Farre D, Roset R, Huerta M *et al*: Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res* 2003; **31**: 3651–3653.
- Cawthon RM: Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 2009; **37**: e21.
- Mocellin S, Verdi D, Pooley KA *et al*: Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. *J Natl Cancer Inst* 2012; **104**: 840–854.
- Ding Z, Green AG, Yang X, Chernenko G, Tang SC, Pater A: Retinoic acid inhibits telomerase activity and downregulates expression but does not affect splicing of *hTERT*: correlation with cell growth rate inhibition in an *in vitro* cervical carcinogenesis/multidrug-resistance model. *Exp Cell Res* 2002; **272**: 185–191.
- Pendino F, Dudognon C, Delhommeau F *et al*: Retinoic acid receptor alpha and retinoid-X receptor-specific agonists synergistically target telomerase expression and induce tumor cell death. *Oncogene* 2003; **22**: 9142–9150.
- Pickett HA, Baird DM, Hoff-Olsen P *et al*: Telomere instability detected in sporadic colon cancers, some showing mutations in a mismatch repair gene. *Oncogene* 2004; **23**: 3434–3443.
- Mendez-Bermudez A, Hills M, Pickett HA *et al*: Human telomeres that contain (CTAGGG)*n* repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res* 2009; **37**: 6225–6238.
- Rampazzo E, Bertorelle R, Serra L *et al*: Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers. *Br J Cancer* 2010; **102**: 1300–1305.
- Mendez-Bermudez A, Royle NJ: Deficiency in DNA mismatch repair increases the rate of telomere shortening in normal human cells. *Hum Mutat* 2011; **32**: 939–946.
- Djojicubroto MW, Choi YS, Lee HW, Rudolph KL: Telomeres and telomerase in aging, regeneration and cancer. *Mol Cells* 2003; **15**: 164–175.
- Burn J, Gerdes AM, Macrae F *et al*: Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet* 2011; **378**: 2081–2087.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)

New insights into *POLE* and *POLD1* germline mutations in familial colorectal cancer and poliposis

Laura Valle, Eva Hernández-Illán, Fernando Bellido, Gemma Aiza, Adela Castillejo, María-Isabel Castillejo, Matilde Navarro, **Nuria Seguí**, Gardenia Vargas, Carla Guarinos, Miriam Juárez, Xavier Sanjuán, Silvia Iglesias, Cristina Alenda, Cecilia Egoavil, Ángel Segura, María-José Juan, María Rodríguez-Soler, Joan Brunet, Sara González, Rodrigo Jover, Conxi Lázaro, Gabriel Capellá, Marta Pineda, José Luís Soto y Ignacio Blanco.

Human Molecular Genetics, 2014

Resumen:

Recientemente se han identificado mutaciones en la línea germinal en los genes que codifican las ADN polimerasas ϵ (*POLE*) y δ (*POLD1*) en familias con múltiples adenomas colorrectales y CCR. Todos los casos descritos son portadores de la mutación *POLE* c.1270C>G (p.Leu424Val) o *POLD1* c.1433G>A (p.Ser478Asn), localizadas en la secuencia que codifica el dominio exonucleasa de la proteína. Debido al reducido número de casos descritos hasta el momento, el fenotipo asociado a estas mutaciones aún no está bien definido. Con el fin de evaluar con mayor precisión la prevalencia de estas mutaciones en casos de CCR familiar, con CCR a edad temprana y con poliposis sin causa genética subyacente determinada, estudiamos mediante genotipado con sondas KASPar o secuenciación Sanger un total de 858 muestras correspondientes a 581 casos de CCR familiar/jóvenes sin deficiencias en el sistema de reparación de bases desapareadas, a 86 casos con defectos en esta vía de reparación y a 191 casos de poliposis. Identificamos la variante *POLE* p.L424V en un individuo de 28 años de edad afecto de poliposis y de CCR. El estudio de la mutación en los progenitores reveló que se trata de una mutación *de novo*. En el caso de *POLD1* identificamos una nueva mutación, c.1421T>C (p.Leu474Pro), en una familia Ámsterdam positiva sin defectos en los genes MMR. El estudio de segregación de la variante con el cáncer en la familia, el análisis *in silico* de función y evidencias previas en ensayos funcionales con *Saccharomyces cerevisiae* apoyan la patogenicidad de la mutación. Este estudio aporta información adicional sobre el fenotipo clínico asociado, la prevalencia y las características genéticas de las mutaciones que afectan al dominio exonucleasa de *POLE* y de *POLD1*. En base a los resultados obtenidos proponemos la secuenciación del dominio exonucleasa de *POLE* y *POLD1* en los protocolos de diagnóstico genético, considerando tanto casos de poliposis como de CCR y teniendo en cuenta la posibilidad de mutaciones *de novo*.

New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis

Laura Valle^{1,†,*}, Eva Hernández-Illán^{3,†}, Fernando Bellido¹, Gemma Aiza², Adela Castillejo⁶, María-Isabel Castillejo⁶, Matilde Navarro¹, Nuria Seguí¹, Gardenia Vargas¹, Carla Guarinos³, Miriam Juárez³, Xavier Sanjuán⁷, Silvia Iglesias¹, Cristina Alenda⁴, Cecilia Egoavil⁴, Ángel Segura⁸, María-José Juan⁹, María Rodríguez-Soler⁵, Joan Brunet¹⁰, Sara González¹, Rodrigo Jover⁵, Conxi Lázaro¹, Gabriel Capellá¹, Marta Pineda¹, José Luís Soto⁶ and Ignacio Blanco¹

¹Hereditary Cancer Program and ²Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Spain, ³Research Laboratory, ⁴Department of Pathology and ⁵Department of Gastroenterology, Alicante University Hospital, Alicante, Spain, ⁶Molecular Genetics Laboratory, Elche University Hospital, Elche, Spain, ⁷Department of Pathology, Bellvitge University Hospital, IDIBELL, Hospitalet de Llobregat, Spain, ⁸Hereditary Cancer Unit, La Fe University Hospital, Valencia, Spain, ⁹Hereditary Cancer Unit, Valencian Institute of Oncology, Valencia, Spain and ¹⁰Hereditary Cancer Program, Catalan Institute of Oncology, IDIBGi, Girona, Spain

Received January 8, 2014; Revised and Accepted January 31, 2014

Germline mutations in DNA polymerase ϵ (*POLE*) and δ (*POLD1*) have been recently identified in families with multiple colorectal adenomas and colorectal cancer (CRC). All reported cases carried *POLE* c.1270C>G (p.Leu424Val) or *POLD1* c.1433G>A (p.Ser478Asn) mutations. Due to the scarcity of cases reported so far, an accurate clinical phenotype has not been defined. We aimed to assess the prevalence of these recurrent mutations in unexplained familial and early-onset CRC and polyposis, and to add additional information to define the clinical characteristics of mutated cases. A total of 858 familial/early onset CRC and polyposis patients were studied: 581 familial and early-onset CRC cases without mismatch repair (MMR) deficiency, 86 cases with MMR deficiency and 191 polyposis cases. Mutation screening was performed by KASPar genotyping assays and/or Sanger sequencing of the involved exons. *POLE* p.L424V was identified in a 28-year-old polyposis and CRC patient, as a *de novo* mutation. None of the 858 cases studied carried *POLD1* p.S478N. A new mutation, *POLD1* c.1421T>C (p.Leu474Pro), was identified in a mismatch repair proficient Amsterdam II family. Its pathogenicity was supported by cosegregation in the family, *in silico* predictions, and previously published yeast assays. *POLE* and *POLD1* mutations explain a fraction of familial CRC and polyposis. Sequencing the proofreading domains of *POLE* and *POLD1* should be considered in routine genetic diagnostics. Until additional evidence is gathered, *POLE* and *POLD1* genetic testing should not be restricted to polyposis cases, and the presence of *de novo* mutations, considered.

INTRODUCTION

Estimates indicate that familial colorectal cancer (CRC) defined by the presence of two or more first-degree relatives affected with CRC involves over 20% of all cases (1–3). Nevertheless,

CRC syndromes caused by known high-penetrance CRC genes collectively account for only 2–6% of all CRC cases. Germline mutations and epimutations in the DNA mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6* and *PMS2* cause Lynch syndrome, explaining a proportion of hereditary non-polyposis CRC cases;

*To whom correspondence should be addressed at: Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Av. Gran Vía 199-203, 08908 Hospitalet de Llobregat Barcelona, Spain. Tel: +34 932607145; Fax: +34 932607466; Email: lvalle@iconcologia.net

†These authors contributed equally to this work.

mutations in *APC* and *MUTYH* primarily predispose to multiple colonic adenomas, a benign precursor of CRC; a 40 kb upstream duplication in *GREM1* cause hereditary mixed polyposis; and several types of hamartomatous polyposis are explained by mutations in *SMAD4*, *BMPRIA*, *STK11* and *PTEN* (4–9). Nevertheless, there are still a number of CRC families suggestive of carrying a mutation in a high-penetrance predisposition gene, but without mutations in the known genes. Among these, a number of familial adenomatous polyposis cases are not explained by germline mutations in *APC* or *MUTYH*.

Recently, using a combination of whole-exome sequencing and linkage analysis in probands with > 10 adenomas by age 60 but no germline mutations in *APC*, *MUTYH* or the *MMR* genes, Palles *et al.* identified DNA polymerase ϵ (*POLE*; MIM #174762) and δ (*POLD1*; MIM #174761) mutations in individuals/families with multiple colorectal adenomas and CRC (10). In all, two pathogenic variants, *POLE* c.1270C>G (p.Leu424Val) (NM_006231) and *POLD1* c.1433G>A (p.Ser478Asn) (NM_002691), and an additional variant whose pathogenicity has not yet been determined, *POLD1* c.981C>G (p.Pro327Leu), were identified. All three genetic changes affect the proofreading (exonuclease) domain of the respective polymerase, suggesting deficient proofreading repair during DNA replication (10–13).

After a comprehensive screening of the identified pathogenic mutations in over 3800 CRC patients of European ancestry enriched for a familial CRC history, multiple adenomas and early-onset disease, a total of 13 families with *POLE* p.L424V and 3 with *POLD1* p.S478N were identified (10). To date, no additional *POLE/POLD1* mutated families have been reported in the literature. Clinical data from the reported families indicate that the two pathogenic mutations show dominant inheritance and confer high risk to multiple colorectal adenomas, large adenomas, early-onset CRC or multiple CRCs. *POLD1* p.S478N also confers increased risk to endometrial cancer in female carriers. Nevertheless, the phenotype varies among carriers, and until additional cases are identified, an accurate description of the clinical characteristics of this syndrome cannot be provided (13).

In this study, we aimed to assess the prevalence of *POLE* p.L424V and *POLD1* p.S478N in polyposis and non-polyposis familial and early-onset CRC cases, and to add additional information to help define the phenotypic/clinical characteristics of mutated cases.

RESULTS

Neither *POLE* p.L424V nor *POLD1* p.S478N was identified in genetically uncharacterized familial non-polyposis CRC cases, including 581 MMR-proficient and 86 MMR-deficient cases. Likewise, *POLD1* p.S478N was not detected in 191 polyposis cases.

POLE p.L424V was identified in a polyposis family (Series no. 1) (Fig. 1A). The index case was a female patient diagnosed with CRC (pT2pN0pM0) and > 35 colonic polyps at age 28. From a total of 33 polyps analyzed, 31 were adenomas, 1 a hyperplastic polyp and 1 a mixed polyp. At 30 years old, 2 years after the surgery, she had developed 8 additional adenomas. No genetic alterations in *APC* and absence of the common *MUTYH* variants had been identified. No loss of heterozygosity (LOH) of the *POLE* chromosomal region, analyzed with two informative

microsatellites 1.13 Mb apart, and studying the allelic abundance of the mutated and wild-type alleles by SNaPshot, was detected in tumor DNA extracted from the colon tumor developed by the mutation carrier (Supplementary Material, Fig. S1).

Based on the clinical findings of the proband, her father was subjected to a colonoscopy at age 56, which revealed a pT2pN0pM0 tumor (adenocarcinoma arising from a tubulovillous adenoma) at the proximal colon and one hyperplastic polyp. Her mother was diagnosed with breast cancer at age 36 and died at 41. No information on gastrointestinal clinical findings in the mother had been reported, and no colon cancer surveillance measures were followed in the maternal family branch. The study of cosegregation revealed that the father did not carry the p.L424V mutation, suggesting a different etiology for his CRC. Paternity was confirmed by microsatellite analysis (data not shown). Likewise, p.L424V was not identified in the DNA extracted from an archived cytology sample obtained from an affected node (metastasis) of the mother's breast cancer. Therefore, these findings indicated that *POLE* p.L424V occurred as a *de novo* germline mutation in the index case.

On account of the mutation-screening method used in the Series no.2, consisting of sequencing exons 13 and 11 of *POLE* and *POLD1*, respectively, a novel genetic change, *POLD1* c.1421T>C (p.Leu474Pro), was detected in an Amsterdam II MMR-proficient family. The index case was a female patient diagnosed with a well-differentiated left colon cancer (pT2pN0pM0) and a synchronous gastrointestinal stromal tumor (GIST) in the large bowel at age 36. No polyps were ever found during surgical removal or follow-up. Her mother was diagnosed with endometrial cancer at age 52. A maternal aunt was diagnosed with metachronous CRC (pT3pN0pM0) and endometrial cancer (Stage IB) at ages 33 and 56, respectively, and no polyps were found in the intestinal tract during surgical intervention and follow-up. A maternal uncle was diagnosed with a gastric cancer at age 72, and his daughter died of a brain tumor at age 42. The maternal grandmother died from a bladder cancer at age 51 (Fig. 1B). Cosegregation analysis performed in the maternal aunt, diagnosed with CRC (33 years) and endometrial cancer (56 years), confirmed her status of heterozygous carrier. Therefore, the mother of the index case was an obligate mutation carrier.

The variant *POLD1* p.L474P is localized in a highly conserved residue located within the proofreading domain of DNA polymerase δ . *In silico* analysis using SNPs3D, PolyPhen-2, Condel and SIFT algorithms predicted relevant functional effects with scores of -3.36 (deleterious), 1 (probably damaging), 1 (deleterious) and 0 (damaging), respectively. Human *POLD1* p.L474 is the homologous residue of p.L479 in *Saccharomyces cerevisiae*. The mutation p.L479S Pol3 in this organism has been shown to cause a mutator phenotype (14). Moreover, human *POLD1* p.L474 is the paralogous residue of the human *POLE* p.L424, the residue where the recurrent *POLE* p.L424V mutation occurs (10). In summary, evidence from cosegregation, *in silico* predictions of the variant's functionality and yeast functional assays strongly suggests a pathogenic nature for *POLD1* p.L474P.

Mutation screening of the driver genes *KRAS* (codons 12 and 13, and exons 3 and 4), *NRAS* (exons 2–4), and *BRAF* p.V600E in the colorectal tumor developed by the index case and in the endometrial tumors developed by her maternal aunt revealed no somatic mutations.

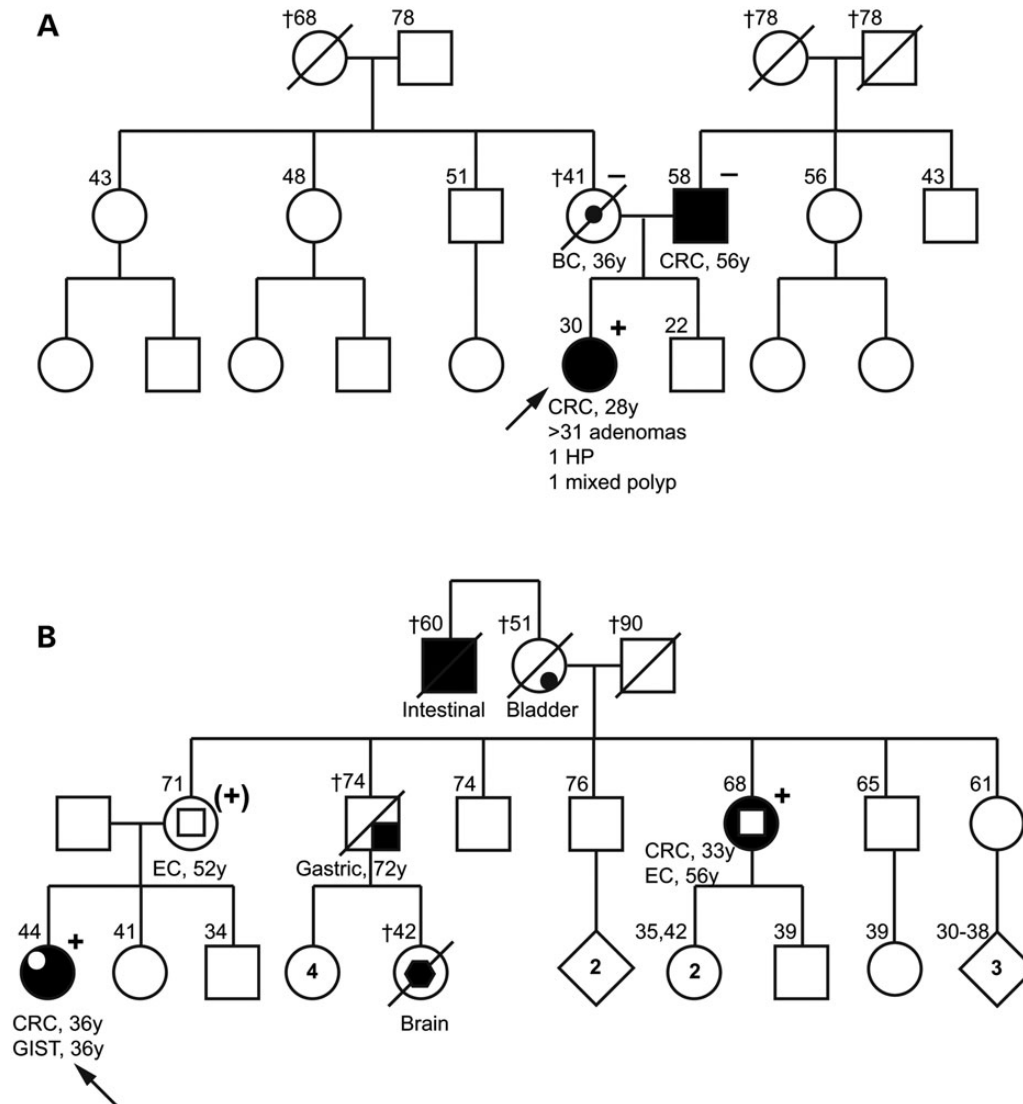


Figure 1. Pedigrees of the families with *POLE* p.L424V (**A**) and *POLD1* p.L474P (**B**) mutations. Filled symbol, CRC; centered filled circle, breast cancer; bottom-right filled circle, bladder cancer; centered unfilled square, endometrial cancer; top-left unfilled circle, GIST; bottom-right filled square, gastric cancer; filled hexagon, tumor of the central nervous system; +, mutation carrier; (+) obliged mutation carrier; -, wild-type; arrow, index case. Ages at information gathering or at death, when available, are indicated on the top-right corner of each individual's symbol. CRC, colorectal cancer; BC, breast cancer; EC, endometrial cancer; GIST, gastrointestinal stromal tumor; HP, hyperplastic polyp; y, years.

DISCUSSION

POLE p.L424V and *POLD1* p.S478N mutation screening in 858 Caucasian (Spanish) patients with CRC and/or colonic polyposis, enriched for a family history of colorectal tumors, multiple colonic polyps and/or early-onset disease, identified one carrier of *POLE* p.L424V. This accounts for 0.12% (1/858) of the total, 0.52% (1/191) of the polyposis cases, and 0.86% (1/116) of the adenomatous polyposes studied. Despite its infrequency and based on the simplicity of the test, our findings provide further evidence to advise that at least *POLE* p.L424V, as a recurrent mutation, should be tested in adenomatous polyposis cases without mutations in *APC* and *MUTYH*.

Together with the family identified in our series, a total of 14 families carrying the *POLE* p.L424V mutation have been

reported and described in the literature (10). Eleven of them were CRC-only and/or polyposis families. Of the other two families previously described, an astrocytoma and tumors of the ureter, ovary and breast were reported in mutation carriers, or probable mutation carriers, who had also been diagnosed with at least two additional colorectal tumors (10). Recently, an additional carrier of a *POLE* mutation, the deletion c.5621_5622delGT, has been identified in a patient diagnosed with CRC at 26 years of age, with no further information reported about family history of cancer or polyposis (15). In our family, the p.L424V mutation occurred *de novo* and caused early-onset CRC (28 years) and adenomatous polyposis. To date, this is the first *de novo* case reported for *POLE/POLD1* germline mutations. Nevertheless, as occurs in 20% *APC* mutation carriers

with an apparent *de novo* mutation (16), it could also be a consequence of a somatic mosaicism in one parent. Being *POLE* p.L424V a recurrently found mutation, it can be hypothesized that its recurrent nature may be the result of a founder effect. Even though this could still be true for certain instances, the existence of *de novo* p.L424 V mutations supports the idea of a mutation hotspot, which may also originate non-founder recurrent mutated cases.

In this study, we also report a novel mutation, *POLD1* p.L474P, identified in an Amsterdam II family without defects in the MMR system. This finding suggests that polymerase proofreading mutations in *POLD1* explain a proportion of the uncharacterized hereditary non-polyposis CRC cases. This finding supports the screening of the genes, at least of *POLD1*, in non-polyposis CRC cases. In this regard, the term 'polymerase proofreading-associated polyposis' may be misleading and should be carefully used, at least until more *POLE/POLD1* families are described and the full phenotypic spectrum of this syndrome is refined.

In addition to the family herein identified, carrying *POLD1* p.L474P, three additional families with germline *POLD1* mutations have been previously described, all of them carrier of p.S478N (10). In three of the four *POLD1* families reported, including the one described here, two or more endometrial tumors have been diagnosed, indicating the importance of cancer surveillance of this type of tumor in *POLD1* mutation carriers (10,17).

Tumors developed in the context of polymerase proofreading mutations, both germline and somatic, show an ultramutated, apparently microsatellite-stable phenotype, sometimes leading to over a million base substitutions per tumor. In these tumors, the mutation spectrum is changed, with a particular increase in the proportion of G : C to T : A and A : T to C : G transversions. Therefore, it would be expected to find numerous mutations in common CRC genes. However, except for rare mutations in *APC*, the frequency of mutations in other driver genes is low (13). Here, the screening of common *KRAS*, *NRAS* and *BRAF* mutations in two tumors (1 colorectal and 1 endometrial cancer) from two *POLD1* p.L474P carriers, which also showed MMR proficiency, revealed no mutations, even when some of the most common mutations found in those genes in colorectal cancer, such as *KRAS* c.34G>T (p.G12C) or *BRAF* c.1799T>A (p.V600E), are transversions. This agrees with the results obtained by Palles *et al.* (10), where no mutations in driver genes were identified in 4 of 6 tumors from five *POLD1* mutation carriers, and in 5 of 10 tumors from three *POLE* mutation carriers. Additional studies analyzing the mutation burden and the presence of mutations in additional known driver genes in the tumors developed by *POLE* and *POLD1* mutation carriers, both CRC and endometrial, will provide a clearer picture of the somatic molecular landscape of this syndrome.

Our results, together with the information gathered so far, support the recommendation of sequencing the exons encoding the proofreading domains of *POLE* and *POLD1* in all familial CRC and polyposis cases without mutations in the known predisposing genes. Based on our findings, *POLE* and *POLD1* mutation screening should not be restricted to polyposis cases, and the presence of *de novo* mutations should be taken into account. Also, whenever endometrial cancer cases are reported in the family, genetic testing of *POLD1* should be prioritized.

MATERIALS AND METHODS

Patients

A total of 858 familial/early-onset CRC and polyposis patients from 840 families were included in the analysis. Written informed consent was obtained from all subjects and the study received the approval of the Ethics Committees of the involved institutions.

Series no.1

A total of 612 familial and/or early-onset colorectal cancer and/or polyposis patients from 594 families without identified mutations in the known CRC or polyposis genes were included in the study. All had been referred to the Genetic Counseling Units of the Catalan Institute of Oncology in the Spanish region of Catalonia between 1999 and 2012. Referral was based on familial history of colorectal cancer or polyps, presence of early-onset colorectal cancer and/or personal history of polyposis at early age.

All non-polyposis cases ($n = 524$) had been previously tested for MMR deficiency, either by immunohistochemistry of the MMR proteins *MLH1*, *MSH2*, *MSH6* and *PMS2*, and/or by PCR-based microsatellite instability analysis. MMR-deficient tumors ($n = 86$) did not present somatic *MLH1* promoter methylation and/or the *BRAF* V600E mutation, thus suggesting a hereditary component. Nevertheless, no germline mutations were identified in the candidate MMR genes, according to the MMR protein expression pattern in the tumor. Clinical features of non-polyposis cases are shown in Table 1.

All polyposis patients ($n = 88$) had undergone *MUTYH* genetic testing of the three most recurrent genetic variants in Spanish population, i.e. p.Tyr179Cys, p.Gly396Asp and p.Glu410Glyfs*43 (NM 001128425.1) (18). If one of these three was detected, all the coding regions of *MUTYH* were subsequently sequenced. When the number of adenomas was >20 , the *APC* gene, including exons and flanking regions, was also sequenced. Clinical features of polyposis cases are shown in Table 2.

Series no.2

Series no.2 consisted of 246 uncharacterized hereditary CRC and/or polyposis index patients. The 143 non-polyposis CRC patients included in the study were recruited through the Cancer Genetic Counseling Units of the Spanish region of Valencia between 2005 and 2013. Of them, 63 cases fulfilled the Amsterdam criteria (I or II) and were MMR-proficient. The remaining 80 cases were diagnosed with MMR-proficient non-polyposis CRC and had two or more first or second-degree relatives diagnosed with a Lynch syndrome-related tumor, regardless of age (Bethesda criterion no. 5) (Table 1). The status of MMR deficiency was assessed either by immunohistochemistry of the MMR proteins and/or by PCR-based microsatellite instability analysis. Biological samples and clinicopathological information were obtained from the Valencian Biobank Network and from the Hereditary Cancer Program of the Valencia Region, both in Spain.

A total of 103 polyposis cases were recruited through the EPI-POLIP project, which comprises a multicentric Spanish series (19). All cases were diagnosed with attenuated polyposis with >10 polyps and at least one first-degree relative affected with

Table 1. Characteristics of the non-polyposis CRC cases analyzed

	<i>N</i> (fam.)	Criteria <i>n</i> (%)				n.a.	Age at cancer diagnosis Mean (± SD)
		Ams. I	Ams. II	Beth.			
Series no.1 (506 families) ^a							
MMR-proficient ^b	438 (423)	31 (7.1%)	11 (2.5%)	390 (89.0%)	6 (1.4%)	49.0 (± 12.6)	
MMR-deficient ^c	86 (86)	1 (1.2%)	4 (4.7%)	63 (73.3%)	18 (20.9%) ^d	51.4 (± 13.3)	
Series no.2 (143 families)							
MMR-proficient ^b	143 (143)	17 (11.9%)	46 (32.2%)	80 (55.9%)	0	49.4 (± 11.6)	
Total	667 (649)	49 (7.4%)	61 (9.1%)	533 (79.9%)	24 (3.6%)	–	

MMR, mismatch repair; *N*, number of individuals; fam., number of families; Ams., Amsterdam criteria (I or II); Beth., Bethesda criteria; n.a., not available data; SD, standard deviation.

^aThree families shared MMR-proficient and MMR-deficient cases.

^bNon-polyposis cases whose tumors showed microsatellite stability and intact expression of the MMR proteins MLH1, MSH2, MSH6 and PMS2.

^cNon-polyposis cases whose tumors showed microsatellite instability and/or loss of expression of at least one MMR protein.

^dCases referred from the Department of Pathology (CSUB, IDIBELL) to the Hereditary Cancer Program (ICO, IDIBELL) based on tumor histopathological features suggestive of MMR deficiency, which was subsequently confirmed. Somatic promoter *MLH1* methylation was discarded and/or the presence of *BRAF* V600E confirmed. No information on familial cancer history was available.

Table 2. Characteristics of the polyposis cases analyzed

	Adenomatous polyposis ^a	Attenuated adenomatous polyposis ^b	Non-adenomatous polyposis ^c	n.a. ^d
Series no.1 (n = 88)				
<i>N</i> (%)	15 (17.0%)	42 (47.7%)	14 (15.9%)	17 (19.3)
Mean age at polyposis diagnosis (± SD)	43.0 (± 11.1)	53.3 (± 13.3)	51.8 (± 8.5)	–
CRC; <i>n</i> (%)	9 (60.0%)	31 (73.8%)	8 (57.1%)	–
Polyposis family history; <i>n</i> (%)	4 (26.7%)	14 (33.3%)	3 (21.4%)	–
CRC family history; <i>n</i> (%)	2 (13.3%)	11 (26.2%)	4 (28.6%)	–
Series no.2 (n = 103)				
<i>N</i> (%)	0	58 (56.3%)	42 (40.8%)	3 (2.9%)
Mean age at polyposis diagnosis (± SD)	0	60.6 (± 10.5)	50.9 (± 9.7)	–
CRC <i>n</i> (%)	0	21 (36.2%)	9 (21.4%)	–
Polyposis family history; <i>n</i> (%)	0	16 (27.6%)	40 (95.2%)	–
CRC family history; <i>n</i> (%)	0	54 (93.1%)	13 (31.0%)	–
TOTAL (n = 191)	15 (7.8%)	100 (52.4%)	56 (29.3%)	20 (10.5%)

N, number of individuals; SD, standard deviation; CRC, colorectal cancer; n.a., not available data.

^aAdenomatous polyposis: >100 adenomatous polyps.

^bAttenuated adenomatous polyposis: 10–100 adenomatous polyps.

^cNon-adenomatous polyposis includes hyperplastic, serrated and mixed polyposis.

^dReferred to the corresponding unit of genetic diagnosis as ‘polyposis’ but with no specific clinical information available.

CRC. All cases had previously undergone genetic testing of the three most frequent *MUTYH* genetic variants in Spanish population, by sequencing exons 7 and 13 of the gene. If one of these three was detected, all the coding regions of *MUTYH* were subsequently sequenced. The APC gene was analyzed by Sanger sequencing in all individuals with >10 adenomas.

***POLE* p.L424V and *POLD1* p.S478N screening**

In Series no.1, KASPar assays (KASP-By-Design genotyping assays, LGC group, Teddington, UK) were used to genotype the two mutations. Reactions were carried out in the LightCycler 480 real-time PCR detection system (Roche Diagnostics GmbH, Germany), including a corresponding positive control in each run. Positive controls for *POLE* p.L424V and *POLD1* p.S478N were kindly provided by Professor Ian Tomlinson (The Wellcome Trust Center for Human Genetics, Oxford, UK). Genotype calling was performed automatically by the LightCycler 480 II software. Validation of genotyping results deviated from the wild-type cluster, analysis of samples that had failed (no amplification) the genotyping experiment, and

cosegregation studies, were carried out by direct automated sequencing. Primers and PCR conditions are shown in Supplementary Material, Table S1. Sequencing was performed on an ABI Sequencer 3730 and data analyzed using Mutation Surveyor v.3.10.

In Series no.2, Sanger sequencing was used to screen for mutations in exon 13 of *POLE*, where *POLE* p.L424V is located, and in exon 11 of *POLD1*, where *POLD1* p.S478N is located. Primers and PCR conditions are shown in Supplementary Material, Table S1. Sequencing was performed on a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), and data were analyzed using Sequencing Analysis v.5.1 and Variant Reporter v.1.1 (Applied Biosystems, Foster City, CA, USA).

Loss of heterozygosity

Microsatellites mapping close to *POLE* and expanding 1.356 Mb, D12S1723, D12S1628, D12S357 and D12S1638, were analyzed to assess LOH in DNA extracted from formalin-fixed paraffin-embedded tissue (10). Also, SNaPshot targeting the

mutation p.L424V was used to assess LOH and to discriminate wild-type and mutated alleles. Primers and conditions are shown in Supplementary Material, Table S1. LOH was scored if the intensity of any allele was reduced by $\geq 50\%$ relative to the other allele after taking account of the relative allelic intensities in paired constitutional DNA.

KRAS, NRAS and BRAF mutation screening

Analysis of *KRAS* mutations at codons 12 and 13 was performed using *KRAS* StripAssay (VienaLab Diagnostics GmbH, Vienna, Austria), following manufacturer's instructions. Exons 3 and 4 of *KRAS*, exons 2, 3 and 4 of *NRAS*, and *BRAF* V600E were assessed by direct automated (Sanger) sequencing. Primers, and PCR and sequencing conditions are available upon request.

In silico prediction analysis

In silico studies to assess the impact of amino acid substitutions (missense variants) on protein structure, function and evolutionary conservation were performed with SNPs3D, PolyPhen-2, SIFT and CONDEL algorithms (20–23).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank all the people responsible for genetic counseling and genetic testing in hereditary cancer from all involved institutions; the Spanish Epidemiological Polyposis Study, EPIPOLIP; Isabel Català, director of the Cytology Unit at Bellvitge University Hospital; and Amelia Rodríguez and Angel Carracedo from the Institute of Legal Medicine at University of Santiago de Compostela.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Spanish Ministry of Economy and Competitiveness (State Secretariat for Research, Development and Innovation) (SAF2012-38885 to L.V.); the Spanish Ministry of Health and the Carlos III Health Institute (FIS PI08/0726 to R.J.); L'Oréal-UNESCO 'For Women in Science'; the Scientific Foundation Asociación Española Contra el Cáncer; and the Government of Catalonia (2009SGR290). L.V. is a recipient of a Ramón y Cajal contract and F.B. of a fellowship both from the Spanish Ministry of Economy and Competitiveness. E.H.-I. and N.S. hold fellowships from the Carlos III Health Institute and C.G. from the Conselleria d'Educació of the Valencian Autonomous Community.

REFERENCES

- Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000)

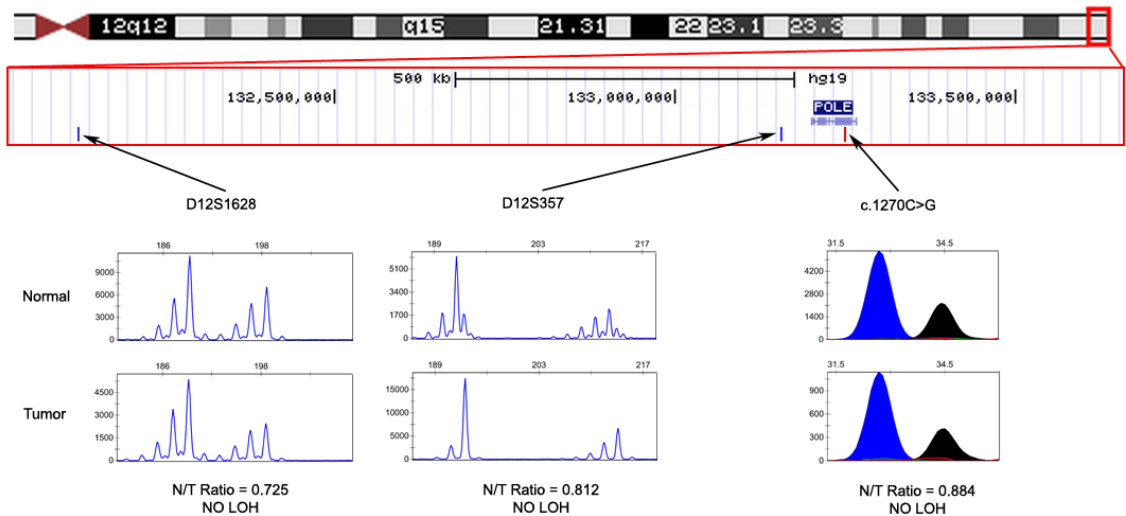
- Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*, **343**, 78–85.
- Aaltonen, L., Johns, L., Jarvinen, H., Mecklin, J.P. and Houlston, R. (2007) Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin. Cancer Res.*, **13**, 356–361.
- Abdel-Rahman, W.M. and Peltomaki, P. (2008) Lynch syndrome and related familial colorectal cancers. *Crit. Rev. Oncog.*, **14**, 1–22; discussion 23–31.
- Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F. and Boland, C.R. (2009) Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.*, **76**, 1–18.
- Sieber, O.M., Lipton, L., Crabtree, M., Heinemann, K., Fidalgo, P., Phillips, R.K., Bisgaard, M.L., Orntoft, T.F., Aaltonen, L.A., Hodgson, S.V. *et al.* (2003) Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N. Engl. J. Med.*, **348**, 791–799.
- Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., Becker, M., Cuadrado, P.R., Davis, H., Kaur, K., Heinemann, K., Howarth, K. *et al.* (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. *Nat. Genet.*, **44**, 699–703.
- Howe, J.R., Sayed, M.G., Ahmed, A.F., Ringold, J., Larsen-Haidle, J., Merg, A., Mitros, F.A., Vaccaro, C.A., Petersen, G.M., Giardiello, F.M. *et al.* (2004) The prevalence of MADH4 and BMPRIA mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations. *J. Med. Genet.*, **41**, 484–491.
- Beggs, A.D., Latchford, A.R., Vasen, H.F., Moslein, G., Alonso, A., Aretz, S., Bertario, L., Blanco, I., Bulow, S., Burn, J. *et al.* (2010) Peutz-Jeghers syndrome: a systematic review and recommendations for management. *Gut*, **59**, 975–986.
- Mester, J. and Eng, C. (2013) When overgrowth bumps into cancer: the PTEN-opathies. *Am. J. Med. Genet. C Semin. Med. Genet.*, **163**, 114–121.
- Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I. *et al.* (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.*, **45**, 136–144.
- Donehower, L.A., Creighton, C.J., Schultz, N., Shinbrot, E., Chang, K., Gunaratne, P.H., Muzny, D., Sander, C., Hamilton, S.R., Gibbs, R.A. *et al.* (2013) MLH1-silenced and non-silenced subgroups of hypermutated colorectal carcinomas have distinct mutational landscapes. *J. Pathol.*, **229**, 99–110.
- Church, D.N., Briggs, S.E., Palles, C., Domingo, E., Kearsey, S.J., Grimes, J.M., Gorman, M., Martin, L., Howarth, K.M., Hodgson, S.V. *et al.* (2013) DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum. Mol. Genet.*, **22**, 2820–2828.
- Briggs, S. and Tomlinson, I. (2013) Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. *J. Pathol.*, **230**, 148–153.
- Murphy, K., Darmawan, H., Schultz, A., Fidalgo da Silva, E. and Reha-Krantz, L.J. (2006) A method to select for mutator DNA polymerase deltas in *Saccharomyces cerevisiae*. *Genome*, **49**, 403–410.
- Smith, C.G., Naven, M., Harris, R., Colley, J., West, H., Li, N., Liu, Y., Adams, R., Maughan, T.S., Nichols, L. *et al.* (2013) Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum. Mutat.*, **34**, 1026–1034.
- Hes, F.J., Nielsen, M., Bik, E.C., Konvalinka, D., Wijnen, J.T., Bakker, E., Vasen, H.F., Breuning, M.H. and Tops, C.M. (2008) Somatic APC mosaicism: an underestimated cause of polyposis coli. *Gut*, **57**, 71–76.
- Kirk, R. (2013) Genetics: new classification for endometrial cancer puts genes in POLE position. *Nat. Rev. Clin. Oncol.*, **10**, 304.
- Gomez-Fernandez, N., Castellvi-Bel, S., Fernandez-Rozadilla, C., Balaguer, F., Munoz, J., Madrigal, I., Mila, M., Grana, B., Vega, A., Castells, A. *et al.* (2009) Molecular analysis of the APC and MUTYH genes in Galician and Catalanian FAP families: a different spectrum of mutations? *BMC Med. Genet.*, **10**, 57.
- Guarinos, C., Sanchez-Fortun, C., Rodriguez-Soler, M., Perez-Carbonell, L., Egoavil, C., Juarez, M., Serradesanferm, A., Bujanda, L., Fernandez-Banares, F., Cubiella, J. *et al.* (2013) Clinical subtypes and molecular characteristics of serrated polyposis syndrome. *Clin. Gastroenterol. Hepatol.*, **11**, 705–711; quiz e746.

20. Yue, P. and Moulton, J. (2006) Identification and analysis of deleterious human SNPs. *J. Mol. Biol.*, **356**, 1263–1274.
21. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
22. Ng, P.C. and Henikoff, S. (2002) Accounting for human polymorphisms predicted to affect protein function. *Genome Res.*, **12**, 436–446.
23. Gonzalez-Perez, A. and Lopez-Bigas, N. (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.*, **88**, 440–449.

SUPPLEMENTARY MATERIAL

Suppl. Table 1. Primers and PCR conditions.

	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon size	Annealing temperature	Series
<i>POLE</i> _exon 13	CATCCTGGCTTCTGTTCTCA	GTGGCCATCTGGATGTGTG	223	60°C	No.2
<i>POLD1</i> _exon 11	GTGTGTCCCTGTCCCTTGAA	GTCAGAGGTTGGGGTGAGAG	217	60°C	No.2
<i>POLE</i> _L424V	GGTGCTGTTAGGAACTTGC	CCGCACACACAGTAAGGAGA	449	57°C	No.1
<i>POLD1</i> _S478N	GGAGTACAAGCTCCGCTCCT	GAAAAAGTGGGCGTCAGGTA	250	57°C	No.1
SNaPshot					
<i>POLE</i> _L424V_LOH	TTACCTTCTGTGGGCAGTC	TAGCTCCACGGGATCATAGC	73	54°C	
SNaPshot extension	TTCCTGTGGGCAGTCATAAT	-	-	-	



Suppl. Figure 1. Absence of LOH at *POLE* in the colon tumor developed by the *POLE* L424V mutation carrier. LOH results using two informative microsatellites, D12S1628 and D12S357, and the mutation, assessed by SNaPshot, are shown.

ANEXO II: Informes del director

INFORME DEL DIRECTOR SOBRE EL FACTOR DE IMPACTO DE LOS ARTÍCULOS PUBLICADOS Y GRADO DE PARTICIPACIÓN

Como directora de la tesis doctoral de Nuria Seguí Gracia titulada “Estudio de la longitud telomérica e identificación de nuevos genes causales en el cáncer colorrectal hereditario no polipósico”, hago constar que la doctoranda ha participado activamente en el diseño y realización experimental de los trabajos incluidos en esta tesis, análisis de resultados, discusión de conclusiones y preparación de los manuscritos finales. A continuación se especifica su contribución específica a cada artículo junto con el factor de impacto de la revista.

ARTÍCULO 1

Telomere length and genetic anticipation in Lynch syndrome

Nuria Seguí, Marta Pineda, Elisabet Guinó, Ester Borràs, Matilde Navarro, Fernando Bellido, Victor Moreno, Conxi Lázaro, Ignacio Blanco, Gabriel Capellá, Laura Valle.

PLoS One 23;8(4):e61286; 2013. [doi: 10.1371/journal.pone.0061286]

Factor de impacto (2012): 3,73

Contribución de la doctoranda: Puesta a punto de la técnica para la cuantificación de la longitud telomérica mediante PCR cuantitativa monocroma multiplex, selección y preparación de las muestras, realización de los experimentos y participación en el análisis e interpretación de los resultados.

ARTÍCULO 2

Longer Telomeres Are Associated with Cancer Risk in MMR-Proficient Hereditary Non-Polyposis Colorectal Cancer

Nuria Seguí, Elisabet Guinó, Marta Pineda, Matilde Navarro, Fernando Bellido, Conxi Lázaro, Ignacio Blanco, Victor Moreno, Gabriel Capellá, Laura Valle.

PLoS One 3;9(2):e86063; 2014. [doi: 10.1371/journal.pone.0086063]

Factor de impacto (2012): 3,73

Contribución de la doctoranda: Selección y preparación de las muestras, cuantificación de la longitud telomérica mediante PCR cuantitativa monocroma multiplex, realización de los experimentos y participación en el análisis e interpretación de los resultados.

ARTÍCULO 3

***GALNT12* is not a major contributor of familial colorectal cancer type X**

Nuria Seguí, Marta Pineda, Matilde Navarro, Conxi Lázaro, Joan Brunet, Mar Infante, Mercedes Durán, José Luís Soto, Ignacio Blanco, Gabriel Capellá, Laura Valle.

Human Mutation 35(1):50-2; 2014. [doi: 10.1002/humu.22454]

Factor de impacto (2012): 5,21

Contribución de la doctoranda: Procesado de las muestras de ADN analizadas, puesta a punto del cribado mediante CSCE y secuenciación de las regiones codificantes y secuencias colindantes mediante. Estudio de la predicción funcional *in silico* de las variantes identificadas y estudios de co-segregación. Participación en la escritura del manuscrito.

ARTÍCULO 4

Exome sequencing identifies *MUTYH* mutations in a family with an atypical phenotype

Nuria Seguí, Matilde Navarro, Marta Pineda, Nicole Köger, Fernando Bellido, Sara González, Olga Campos, Silvia Iglesias, Rafael Valdés-Mas, Adriana López-Doriga, Marta Gut, Ignacio Blanco, Conxi Lázaro, Gabriel Capellá, Xose S. Puente, Guido Plotz, Laura Valle.

Gut. Publicado online: 1 Abril 2014. [doi: 10.1136/gutjnl-2014-307084]

Factor de impacto (2012): 10,73

Contribución de la doctoranda: Selección y procesado de muestras de ADN, secuenciación de los fragmentos correspondientes para el análisis de co-segregación, estudio *in silico* de función a nivel de ARN y proteína. Cultivo de linfocitos, extracción de ARN y síntesis de cDNA.

ARTÍCULO 5: Manuscrito en preparación

Germline mutations in *GEN1X* cause hereditary colorectal cancer through impaired DNA repair

Nuria Seguí et al.

Contribución de la doctoranda: Preparación y procesado de las muestras analizadas. Extracción de ADN de sangre periférica y tejido, tanto fresco como parafinado. Secuenciación Sanger de los casos de las series de validación e identificación y de los familiares para el estudio de segregación. Cribado de la mutación inicial mediante CSCE, análisis de predicción *in silico* de función de las diferentes variantes identificadas, inmortalización de linfocitos. Cultivo celular y silenciamiento estable del gen la línea celular Caco-2. Ensayo de viabilidad celular en respuesta a MMC. Análisis de la expresión proteica mediante *western blot* y estudio de la pérdida de heterocigosidad utilizando la técnica de SNaPshot.

Otros artículos publicados (Anexo)

ARTÍCULO 6

No association between germline allele-specific expression of *TGFBR1* and colorectal cancer risk in Caucasian and Ashkenazi populations

Nuria Seguí, Kristen N. Stevens, Elisabet Guinó, Laura S. Rozek, Victor Moreno, Gabriel Capellá, Stephen B. Gruber y Laura Valle.

British Journal of Cancer 104(4):735-40; 2011. [doi: 10.1038/sj.bjc.6606079]

Factor de impacto (2012): 5,08

Contribución de la doctoranda: Genotipado de los SNP estudiados, estudio de la expresión alelo-específica utilizando la técnica de pirosecuenciación y análisis de los resultados. Trabajo realizado a lo largo de una estancia en la Universidad de Michigan, Estados Unidos.

ARTÍCULO 7

Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome

Fernando Bellido, Elisabet Guinó, Shantie Jagmohan-Changur, **Nuria Seguí**, Marta Pineda, Matilde Navarro, Conxi Lázaro, Ignacio Blanco, Hans F.A. Vasen, Victor Moreno, Gabriel Capellá, Juul T. Wijnen y Laura Valle.

European Journal of Human Genetics 21(5):511-6; 2013. [doi: 10.1038/ejhg.2012.204]

Factor de impacto (2012): 4,32

Contribución de la doctoranda: Cuantificación de la longitud de los telómeros mediante PCR cuantitativa monocroma multiplex.

ARTÍCULO 8

New insights into *POLE* and *POLD1* germline mutations in familial colorectal cancer and poliposis

Laura Valle, Eva Hernández-Illán, Fernando Bellido, Gemma Aiza, Adela Castillejo, María-Isabel Castillejo, Matilde Navarro, **Nuria Seguí**, Gardenia Vargas, Carla Guarinos, Miriam Juárez, Xavier Sanjuán, Silvia Iglesias, Cristina Alenda, Cecilia Egoavil, Ángel Segura,

María-José Juan, María Rodríguez-Soler, Joan Brunet, Sara González, Rodrigo Jover, Conxi Lázaro, Gabriel Capellá, Marta Pineda, José Luís Soto, Ignacio Blanco.

Human Molecular Genetics. Publicado online: Febrero 2014.

Factor de impacto (2012): 7,69

Contribución de la doctoranda: Preparación de muestras de ADN y validaciones mediante secuenciación Sanger.

Atentamente,

Laura Valle Velasco

Investigadora Ramón y Cajal

Programa de Cáncer Hereditario

Instituto Catalán de Oncología, IDIBELL