UNIVERSITAT DE BARCELONA

# Identificación y caracterización de nuevos genes de predisposición al cáncer colorrectal familiar

Clara Esteban Jurado

UNIVERSITAT DE BARCELONA

# IDENTIFICACIÓN Y CARACTERIZACIÓN DE NUEVOS GENES DE PREDISPOSICIÓN AL CÁNCER COLORRECTAL FAMILIAR


Memoria presentada por

**Clara Esteban Jurado**

para optar al grado de

**Doctor en Medicina**

Directores:

**Sergi Castellví Bel**

**Juan José Lozano Salvatella**

**Sergi Castellví Bel** **Juan José Lozano Salvatella**



**Barcelona, 2016**

*Dedicada a mis padres y a Antonio*

**ÍNDICE**

## AGRADECIMIENTOS

Para empezar, quería agradecer esta tesis a mis dos directores, en especial a Sergi, que ha sido fundamental en el diseño experimental y ha estado pendiente en todo momento de que pudiera completarla. También quería agradecerle todo lo que he aprendido de él. También a todas mis compañeras y compañeros de trabajo por las que me he sentido muy acogida y he compartido muchos buenos momentos y risas (incluyendo a Míriam que me llevó de fiesta nada más conocerme y se ha convertido en una amiga). Además también quería agradecer a mis compañeras que me hayan enseñado con paciencia y me hayan ayudado cuando se me acumulaba la faena de laboratorio (especialmente a Jenny, Saray, Esther, Isa y Anna), así como en el tema de programación (o sea a Sebas, Maria y Marcos), pues sin su ayuda tampoco podría haber terminado esta tesis. También ha sido muy importante la colaboración con Rosa y David del laboratorio de bioquímica y biología molecular de la universidad, con quien hemos sufrido juntos las levaduras, pero al final lo habéis conseguido y gracias a ello pudimos enviar un artículo mucho más completo.

Y a todas las personas que me han hecho crecer en estos años, sobre todo a mis amigos que me han apoyado mucho, me han escuchado y cuidado. Y a Antonio que me hace muy feliz siempre y me anima, te quiero muchísimo! Por último agradecer a mis padres que me hayan cuidado y se preocupen por mí cada dia.

¡¡¡Gracias a todos!!!

# ABREVIATURAS

ADN ácido desoxirribonucleico

ADNc ácido desoxirribonucleico complementario

ATP adenosina trifosfato

ADP adenosina difosfato

ARN ácido ribonucleico

ARNm ácido ribonucléico mensajero

CCR cáncer colorrectal

CGH *comparative genomic hybridization*; hibridación genómica comparada

ChIP *chromatin immunoprecipitation,* immunoprecipitación de la cromatina

CIMP *CpG island methylator phenotype;* fenotipo metilador de islas CpG

CIN *chromosomal Instability*; inestabilidad cromosómica

CNV *copy number variation;* variación en el número de copia

CpG dinucleótido citosina-guanina

GWAS *genome-wide association study*; estudio de asociación de genoma completo

HR *hazard ratio*

IC intervalo de confianza

Kb kilobase

LOH *loss of heterozygosity*; pérdida de heterocigosidad

MAF *minor allele frequency*; frecuencia alélica menor

MLPA *multiplex ligation-dependent probe amplification*

MSI *microsatellite instability*; inestabilidad de microsatelites

NGS *next-generation sequencing*; secuenciación de nueva generación

OR *odds ratio*

PAF poliposis adenomatosa familiar

PAM poliposis asociada a *MUTYH*

PCR *polymerase chain reaction*; reacción en cadena de la polimerasa

*RR riesgo relativo*

WES *whole-exome sequencing*; secuenciación del exoma completo

WGS *whole-genome sequencing*; secuenciación del genoma completo

RT-qPCR *reverse transcription polymerase chain reaction*; reacción en cadena de la polimerasa con transcriptasa inversa

SNP *single nucleotide polymorphism;* polimorfismo de nucléotido único

# INTRODUCCIÓN

## 1- Epidemiologia del cáncer colorrectal

El cáncer colorrectal (CCR) representa una de las neoplasias más frecuentes a nivel mundial, variando su incidencia entre países en función del grado de desarrollo económico así como de la calidad de sus registros del cáncer (Torre 2015). Las tasas de incidencia más altas se encuentran en Australia y Nueva Zelanda, América del Norte y Europa, mientras que las más bajas están registradas en África y Sud Asia central (Ferlay 2008, GLOBOCAN 2012), (**Figura 1**).

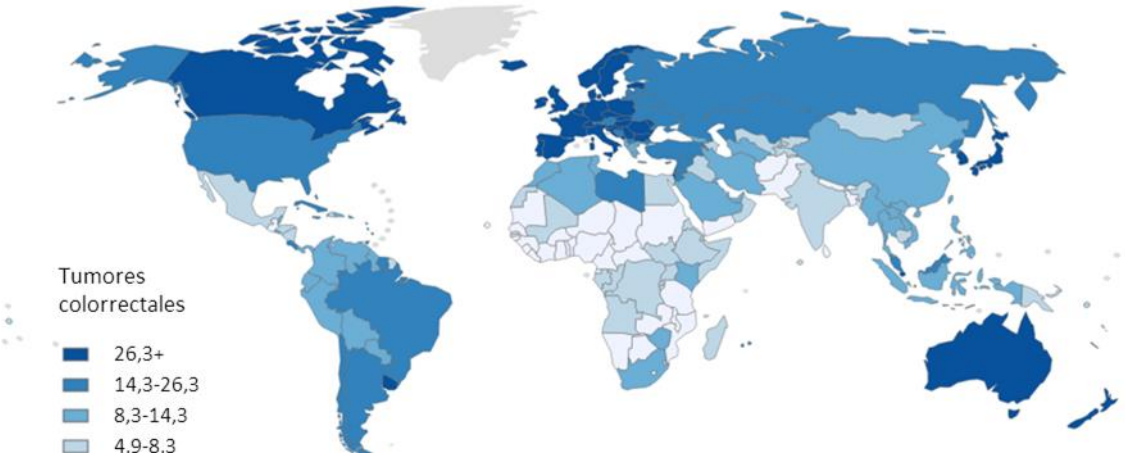



**Figura 1. Mapa mundial de la incidencia de cáncer colorrectal por país estandarizada por edad y calculada para ambos sexos en conjunto.** La layenda representa el número de nuevos casos diagnosticados por cada 100.000 habitantes durante el año 2012. Imagen extraída de Globocan 2012, *International Agency for Research on Cancer.*

En España representa el tumor más frecuente si se tienen en cuenta hombres y mujeres en conjunto (Ferlay 2013), siendo el tercero más frecuente en hombres (después de próstata y pulmón) y el segundo en mujeres (después de mama), (**Figura 2**). En el año 2012 se diagnosticaron en España 32.240 nuevos casos y 14.700 personas murieron de esta enfermedad (GLOBOCAN 2012). En cuanto a la supervivencia a 5 años, depende del estadio de la enfermedad en el momento del diagnóstico y suele variar desde el 90,1% para tumores localizados hasta un 11,7% para los canceres con metástasis distantes (Brenner 2013).

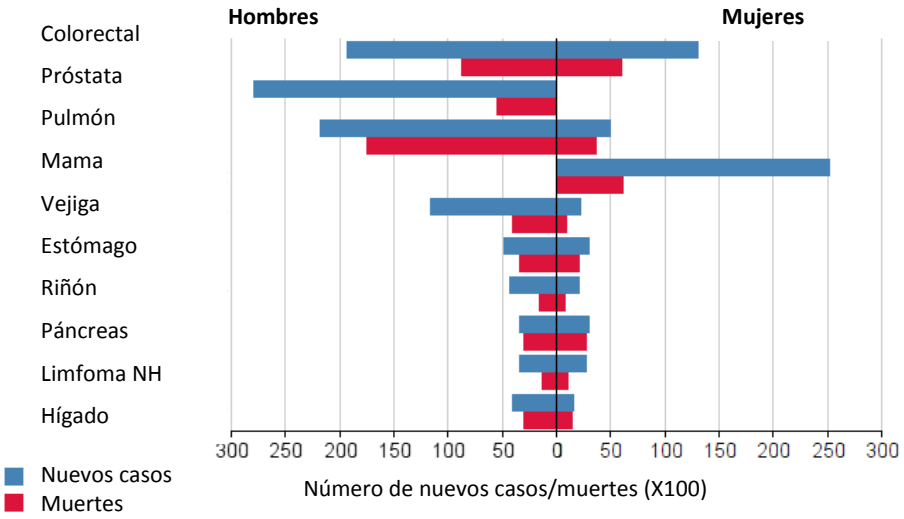



**Figura 2. Incidencia y mortalidad en España de los 10 tipos de cáncer más frecuentemente diagnosticados.** Abreviaturas: NH, no Hodgkin. Imagen extraída de Globocan 2012, *International Agency for Research on Cáncer.*

## 2- Lesiones precursoras

El CCR surge a partir de lesiones precursoras llamadas pólipos (masas celulares sobresalientes en la pared intestinal), y su detección y resección temprana es vital para la prevención del cáncer (Winawer 1993). Antes del desarrollo de la lesión conocida como pólipo ocurre una alteración a nivel microscópico que se conoce como cripta de foco aberrante, consistente en una acumulación de células epiteliales resistentes a la apoptosis, la cual acabará formando el pólipo (Di Gregorio 1997), (**Figura 3**). Las lesiones precursoras se dividen tradicionalmente en dos tipos de pólipos: adenomatosos y serrados.

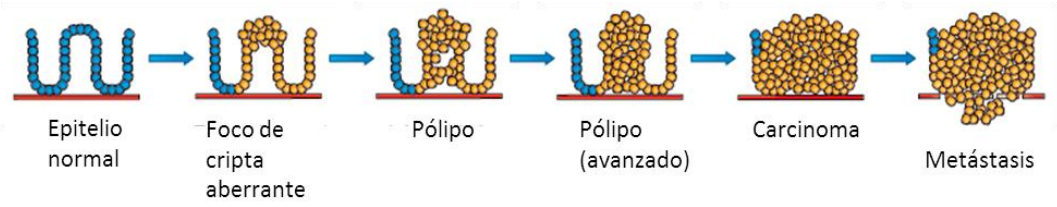



**Figura 3. Iniciación y progresión del CCR a partir del epitelio normal, pasando por la lesión precursora y desembocando en la formación de un tumor metastático.** Figura adaptada de Davies (2005).

### 2.1- Adenomas

Los pólipos adenomatosos o adenomas tienen una organización intracelular e intercelular anormal definida como displasia (las células que constituyen el adenoma presentan alteraciones en su morfología, disposición y tamaño). La incidencia de adenomas en la población general es del 40-60%, pero solo un porcentaje pequeño progresa a CCR, variando el

riesgo en función del tamaño y la composición del pólipo (Vogelstein 2002, Hofstad 2003). O'Brien introdujo el concepto de adenoma avanzado para definir adenomas con ciertas características que aumentan el riesgo de progresión a CCR de forma independiente, como un tamaño mayor o igual a 10 mm, la existencia de componente velloso (morfología en forma de crisantemo o de coliflor), o un alto grado de displasia (O'Brien 1990).

## 2.2- Pólipos serrados

Por otro lado, los pólipos serrados son un grupo heterogéneo caracterizado por una arquitectura epitelial con forma de dientes de sierra o estrellada y están subdivididos en pólipos hiperplásicos, pólipos mixtos, adenomas serrados tradicionales y adenomas serrados sésiles (Snover 2010). En los años 70 se creía que los pólipos adenomatosos eran los únicos con potencial maligno (Muto 1975), pero recientemente se ha descubierto que los pólipos serrados también pueden progresar a CCR (a excepción de los hiperplásicos de menos de 5mm), y por tanto también deben ser extraídos del colon (IJspeert 2015).

## 3- Etiología del cáncer colorrectal

A continuación se detallarán los factores que confieren riesgo al desarrollo de las diferentes formas de CCR. La mayoría de casos (70-80%) son debidos mayoritariamente a la edad y a los factores ambientales y aparecen en individuos sin antecedentes familiares. No obstante, hay una pequeña fracción de casos (~5%) que está causada principalmente por factores genéticos heredados con un patrón de herencia mendeliano (Jasperson 2010). Los factores genéticos heredados también pueden actuar en combinación con los factores ambientales en una proporción importante de casos, coincidiendo con un patrón de herencia compleja (**Figura 4**).

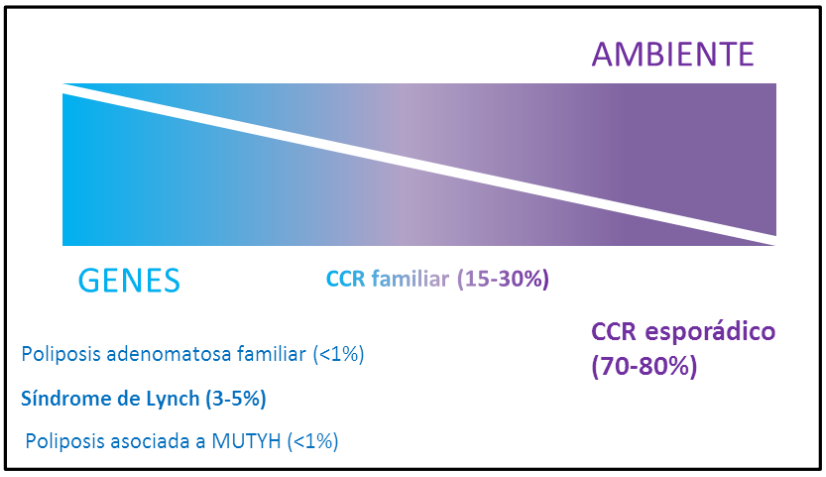



**Figura 4. El cáncer de colon como enfermedad compleja.**

## 3.1- La edad como factor de riesgo

La edad es un factor de riesgo determinante y hace aumentar la probabilidad de desarrollar CCR de forma progresiva debido a mutaciones provocadas por la acumulación de radicales libres de oxígeno derivados del metabolismo, a la secreción de factores oncogénicos por las células senescencentes, así como a cambios en la metilación del genoma (Toyota 1999, Hamilton ML 2001, Martincorena 2015). Tal es el peso de la edad que el 90% de los casos se dan en pacientes de más de 50 años (Parker 1996) (**Figura 5**)

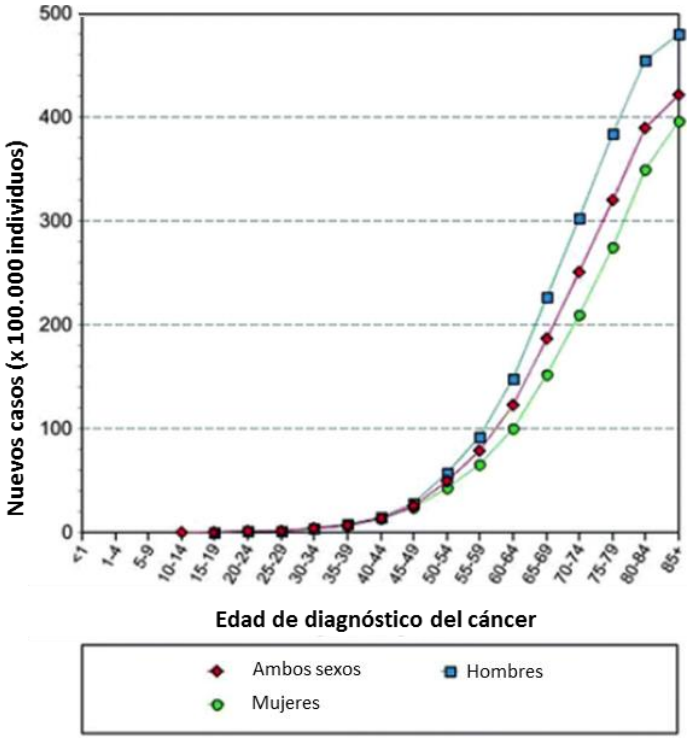



**Figura 5. Incidencia de CCR (número de casos por 100.000 habitantes) clasificada por edad y sexo en los Estados Unidos entre 1992 y 2006.** Datos recogidos por el programa SEER (Surveillance, Epidemiology, and End Results) del National Cáncer Institute.

## 3.2- Factores de riesgo ambientales

Por otro lado, los factores ambientales son condiciones externas al organismo y aumentan la probabilidad de desarrollar CCR debido a la exposición continua si se trata de un agente químico cancerígeno (como el tabaco) o bien debido a hábitos que crean unas condiciones metabólicas favorables para la tumorigenesis (como la inflamación e hiperinsulinemia causada por la mala alimentación y el sedentarismo) (Chan y Giovanucci 2010). Algunos de estos factores están compartidos en los países desarrollados y son responsables de la mayor incidencia en estos, incluyendo un estilo de vida sedentario, la obesidad y una dieta rica en carne roja o altamente procesada (Larsson y Wolk 2006, Moghaddam 2007, Chan y Giovanucci

2010, Boyle 2012). Varios estudios sugieren que las frutas y verduras con alto contenido en fitoquímicos anticancerígenos tienen un efecto protector, así como el consumo de un elevado contenido de fibra en la dieta (Bingham et al 2003, Michels et al 2006). No obstante, otros estudios reportan que el consumo de estos alimentos no tendría efecto preventivo (Michels et al 2001, Koushik et al 2007). Así mismo, se ha postulado un efecto protector en la ingesta moderada de medicamentos como la aspirina y/o antiinflamatorios no esteroideos (AINE) (Cook 2013, Friis 2015). Por otro lado, el consumo elevado de alcohol y el tabaquismo aumentan considerablemente el riesgo de CCR y adenomas (Giovannucci 2001) (**Tabla 1**).

**Tabla 1. Factores de riesgo ambientales y el riesgo asociado** Para valorar el riesgo de ser afecto de una condición dentro de un grupo (expuestos a un factor protector o de riesgo) frente a otro grupo (no expuestos) se utilizan distintas medidas:

RR (Riesgo Relativo)= Probabilidad que tiene un sujeto expuesto a un factor (protector o de riesgo) de desarrollar la enfermedad durante un periodo de tiempo determinado en comparación con la probabilidad de desarrollar la enfermedad en los no expuestos.

OR (*Odds ratio*)= Hipotetizando que hubiera el mismo número de expuestos y de no expuestos, el valor de la Odds ratio definiría cuantos individuos afectados expuestos habría por cada individuo afectado no expuesto (Bland 2000).

HR (*Hazard Ratio*)= Ritmo (frecuencia por unidad de tiempo) en que los individuos expuestos a un factor desarrollan la enfermedad en proporción al ritmo en que los no expuestos al factor desarrollan la enfermedad.

IC (Intervalo de Confianza)= rango de valores (calculado en una muestra) en el cual se encuentra el verdadero valor del parámetro, con una probabilidad determinada.

| | CCR/ Adenomas | Factor protector/ de riesgo | RR (95% IC)/ *Odds ratio* OR (95% IC)/*Hazard ratio* HR (95% IC) | Estudio |
|---|---|---|---|---|
| **Consumo elevado de verduras y frutas** | Adenomas y CCR | No efecto | RR 1,02 (0,98-1,05) para frutas; 1,03 (0,97-1,09) para verduras | Michels y col 2001 |
| | | Factor protector | OR 0,60 (0,44-0,81) para frutas; 0,82 (0,65-1,05) para verduras | Michels y col 2006 |
| **Consumo elevado de fibra** | CCR | Factor protector | RR 0,58 (0,41–0,85) | Bingham y col 2003 |
| | | No efecto | RR 0,91 (0,82-1,01) | Koushik y col 2007 |
| **Actividad física elevada** | CCR | Factor protector | RR 0,76 (0,72-0,81) | Wolin 2009, Boyle y col 2012 |
| **Ingesta continua de aspirina (bajas dosis)** | CCR | Factor protector | HR 0,80 (0,67–0,97) | Cook y col 2013, Friis S y col 2015 |
| **Ingesta continua de AINE (bajas dosis)** | CCR | Factor protector | OR 0,57 (0,44-0,74) | Friis S y col 2015 |
| **Carne roja** | CCR | Factor de riesgo | RR 1,28 (1,15-1,42) | Larsson y Wolk 2006 |
| **Carne procesada** | CCR | Factor de riesgo | RR 1,20 (1,11-1,31) | Larsson y Wolk 2006 |
| **Obesidad** | CCR | Factor de riesgo | RR 1,19 (1,11-1,29) | Moghaddam y col 2007 |
| **Tabaquismo** | Adenomas y CCR | Factor de riesgo | RR 2.96 (1.47–5.98) | Giovannucci 1994, Giovannucci 2001 |
| **Consumo elevado de alcohol** | CCR | Factor de riesgo | RR 1,56 (1,42-1,70) | Huxley y col 2009 |

### 3.3- Factores de riesgo genéticos heredados

Por último, el tener un familiar de primer grado afectado se ha demostrado que aumenta considerablemente el riesgo de CCR (OR 2,25, 95% CI 2,00–2,53). Tener más de un familiar de primer grado afectado aumenta la OR hasta 4,25 (95% CI 3,01–6,08) (Johns 2001). Estudios en gemelos permitieron calcular la heredabilidad, término que indica la proporción en la que los factores genéticos heredados contribuyen a la aparición de la enfermedad. Así, se encontró que en el CCR el componente genético heredado era responsable de un 35% de los casos (Lichtenstein 2000). Los factores genéticos heredados corresponden a mutaciones germinales que se transmiten a la descendencia y que pueden provocar agregación para la enfermedad dentro de una familia, que es un número de casos mayor que el que esperaríamos ver en la población general. Estas variantes pueden dividirse entre alta y baja penetrancia. La penetrancia puede definirse como la proporción de individuos con un genotipo determinado (como una variante en heterocigosis) que presenta un fenotipo asociado a este genotipo (como el desarrollo de CCR).

#### 3.3.1- Variantes de baja penetrancia

Las variantes de baja penetrancia son comunes en la población general (típicamente con frecuencias superiores al 5%) y contribuyen de forma modesta al riesgo de padecer una enfermedad (**Figura 6**). No obstante, su efecto es aditivo y la combinación de varias de estas mutaciones con los factores de riesgo ambientales puede desencadenar la patología. Estos cambios de baja penetrancia pueden contribuir al desarrollo del CCR con o sin historia familiar. En este último caso, puede ocurrir que haya pocas variantes de baja penetrancia en cada uno de los padres sanos pero que al transmitirse a la descendencia en conjunto, este número de variantes aumente y pueda provocar la aparición de la enfermedad junto con los factores ambientales. Hasta el momento se han descubierto 62 variantes comunes de baja penetrancia a CCR en 56 loci. Cada una tiene un RR<1,5 pero todas juntas podrían explicar hasta un 11% de la susceptibilidad genética estimada (Peters 2015).

#### 3.3.2- Variantes de alta penetrancia

En cuanto a las variantes de alta penetrancia, son raras en la población general (típicamente con frecuencias inferiores al 0,5%) y confieren un riesgo alto a contraer la enfermedad, la mayor parte de los portadores de estas variantes desarrollan la enfermedad a lo largo de su vida (**Figura 6**). Se transmiten con un patrón de herencia mendeliana, y por tanto provocan una fuerte agregación familiar para la enfermedad. Más adelante se describirán los genes en

que se han encontrado variantes germinales de alta penetrancia y los síndromes asociados a estas.

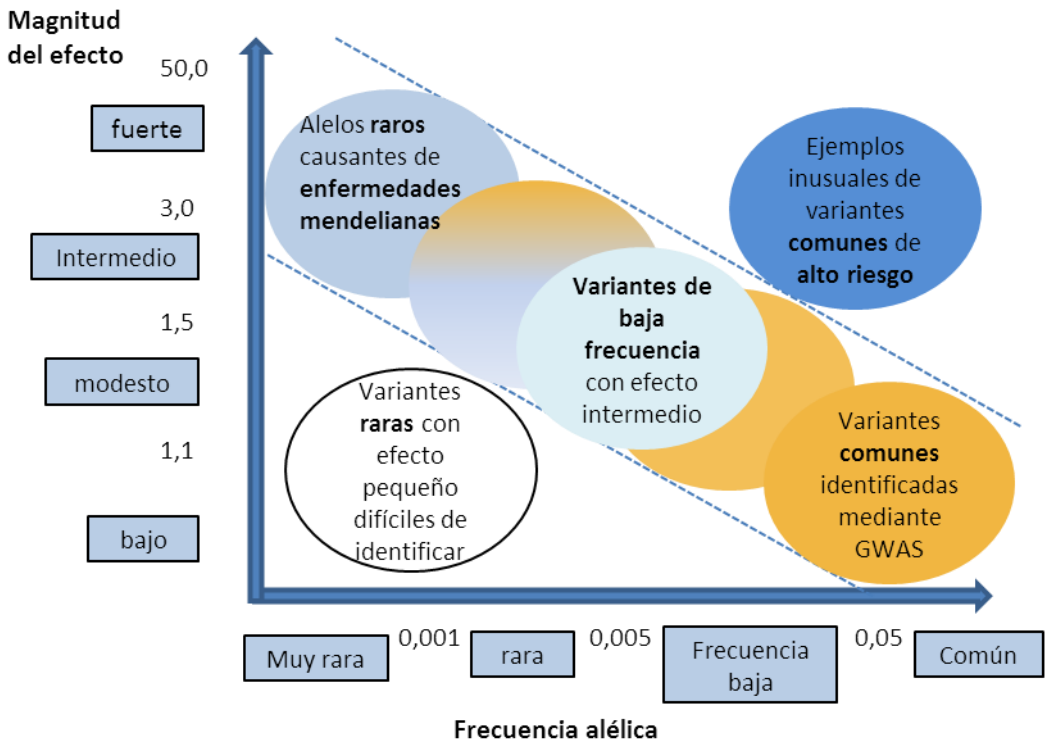



**Figura 6. Clasificación de la variación genética en base a la frecuencia y la magnitud del efecto (odds ratio).** Las variantes comunes con un efecto fuerte son poco usuales, ya que el efecto deletéreo provoca una disminución en la viabilidad de los portadores. Las variantes con baja frecuencia y un efecto bajo son muy difíciles de identificar. Las variantes raras de alta penetrancia son típicas de los síndromes hereditarios, mientras que las variantes comunes de efecto modesto, identificadas mediante estudios de asociación también contribuyen al desarrollo de la enfermedad. Figura adaptada de Manolio (2009).

## 4- Tipos de cáncer colorrectal

La agregación familiar para una enfermedad determinada se define como la aparición de esta enfermedad con mayor frecuencia dentro de una familia de lo que se esperaría por azar en la población general. El CCR se puede clasificar en base a la agregación familiar en formas esporádicas, que no presentan agregación, CCR hereditario que corresponde a las formas sindrómicas causadas por mutaciones en genes conocidos con un patrón de herencia mendeliana, y por último CCR familiar, donde se encuentra cierta agregación familiar y/o una edad de aparición temprana en comparación con los casos esporádicos (Jasperson 2010), (**Figura 7**). A continuación se detallarán los diferentes tipos de CCR.

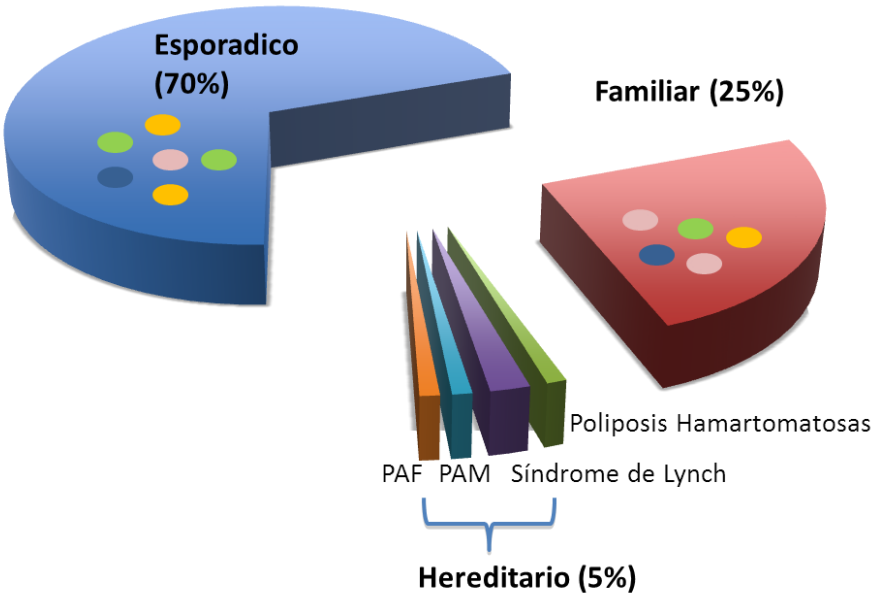



**Figura 7. Clasificación y frecuencia de los casos de CCR según la agregación familiar y/o pertenencia a una forma sindrómica.** Los círculos coloreados representan variantes de susceptibilidad en genes aún por identificar. Abreviaturas: PAF, poliposis adenomatosa familiar, PAM, poliposis asociada a *MUTYH*.

## 4.1-Cáncer colorrectal esporádico

El CCR llamado esporádico corresponde a la mayoría de los casos y no presenta agregación familiar. Es causado por mutaciones somáticas generadas durante la vida del individuo debido a la edad o a factores ambientales. Éstas afectan genes cuya alteración favorece la acumulación posterior de mutaciones en oncogenes y genes supresores de tumores, provocando la formación de lesiones precursoras y su progresión a CCR (Vogelstein 1988). Existen tres tipos de CCR en base a la serie de eventos mutacionales acumulados: tumores con inestabilidad cromosómica, tumores con inestabilidad de microsatélites, y tumores con el fenotipo metilador de las islas CpG (**Figura 8**).

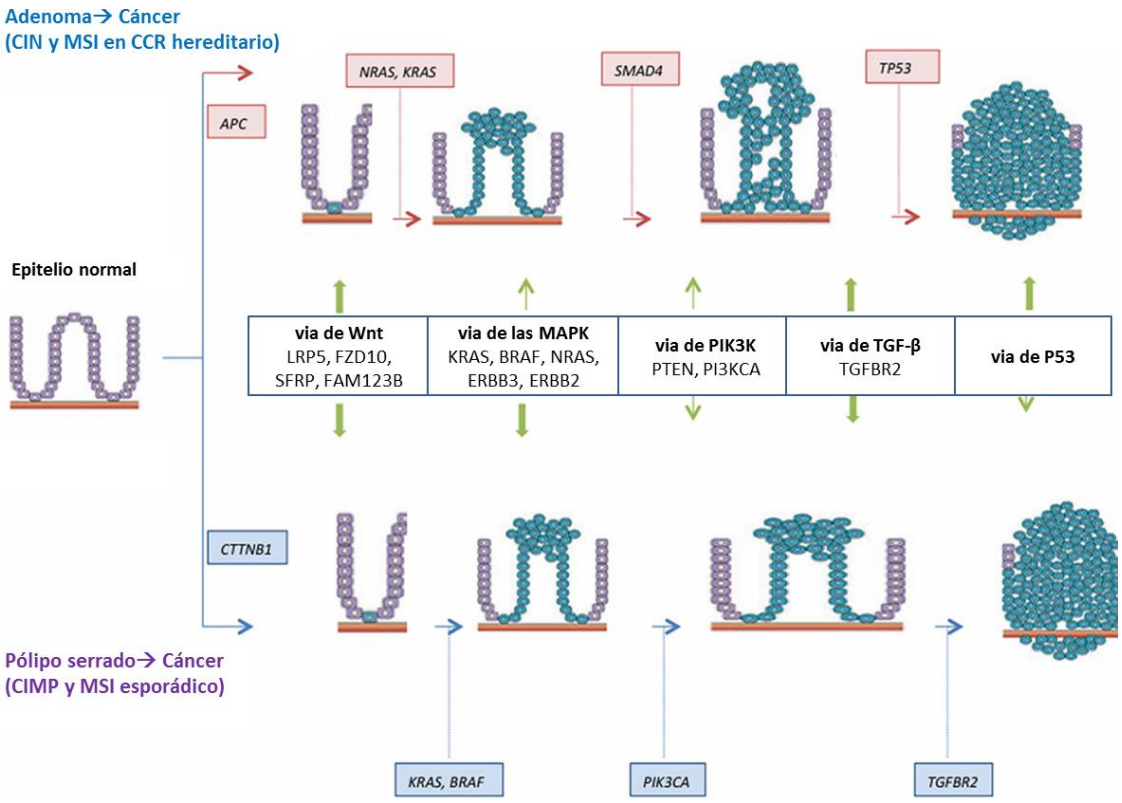



**Figura 8. Secuencia pólipo-carcinoma en el cáncer colorrectal.** La secuencia adenoma-carcinoma clásica conlleva generalmente un fenotipo de inestabilidad cromosómica o de inestabilidad de microsatélites que provoca la acumulación de mutaciones, conllevando la formación de pólipos adenomatosos que progresan a CCR. Por otro lado, la formación de pólipos serrados que progresan a carcinoma está caracterizada por una metilación aberrante de las islas CpG. Los genes mutados o alterados epigenéticamente están indicados para cada secuencia. Las vías de señalización desreguladas también están marcadas con flechas, variando el grosor de éstas según la importancia de la vía de señalización en la progresión de la secuencia. Abreviaturas: CIN=Inestabilidad cromosómica, MSI= Inestabilidad de microsatélites. CIMP=Fenotipo metilador de las islas CpG. Figura adaptada de Dickinson (2015).

### 4.1.1- Tumores con inestabilidad cromosómica

Esta es la vía mayoritaria en el cáncer de colon esporádico y está presente en el 65% de los tumores (Pino 2010). La inestabilidad cromosómica (CIN, *chromosomal instability*) es un proceso que conlleva deleciones, duplicaciones y reordenamientos cromosómicos así como un espectro de mutaciones característico en oncogenes (ej. *KRAS*) y genes supresores de tumores (ej. *APC*, *TP53*) específicos que activan vías de señalización necesarias para la tumorigenesis colorrectal. En los CCR con fenotipo CIN se observa típicamente pérdida de heterocigosidad (LOH, de *loss of heterozygosity*), con una pérdida media del 25-30% de los alelos. El fenotipo CIN puede originarse debido a errores en los mecanismos que establecen una correcta segregación cromosómica, como el punto de control del ensamblaje de huso mitótico (Boveri

1914), la duplicación correcta de los centrosomas durante la mitosis (Ganem 2009) y la longitud de los telómeros (O'Hagan 2002). Otro mecanismo de CIN es el causado por mutaciones en genes de reparación de daño en el ADN como *TP53* (Pino 2010).

### 4.1.2- Tumores con inestabilidad de microsatélites

Los microsatélites son secuencias de ADN de entre dos y cinco pares de bases que se repiten contiguamente un número determinado de veces. Estas secuencias están repartidas por todo el genoma y son propensas a variar en número de repeticiones debido a errores de apareamiento de bases durante la replicación del ADN. Las proteínas encargadas de corregir estos errores constituyen el sistema reparador de las bases desapareadas y están codificadas por los genes *MSH2*, *MLH1, MSH6* y *PMS2*. La inactivación bialélica somática de estos genes reparadores provoca una variación descontrolada en el número de repeticiones de estas secuencias a lo largo del genoma, incluyendo regiones codificantes de ciertos oncogenes y genes supresores de tumores (ej. *TGF-B, TGFBR2, BAX*), alterando así su expresión y desencadenando la tumorigenesis (Markowitz 1995, Rampino 1997). El 15% del CCR esporádico tiene este fenotipo de inestabilidad de microsatélites (MSI, *microsatellite instability*), debido a la inactivación de los genes reparadores mediante dos eventos mutacionales somáticos a lo largo de la vida del individuo. En una fracción importante de los casos esporádicos esta inactivación es debida a una hipermetilación del promotor de *MLH1* asociada con una mutación somática en el gen *BRAF* (p.V600E) (Kane 1997). Además, mutaciones germinales en estos genes son causantes del síndrome de Lynch, una forma de CCR hereditario que se describirá más adelante (Lynch 2007). Los tumores con fenotipo MSI tienen una baja frecuencia de LOH, y un menor número de mutaciones en *KRAS* y *TP53* en comparación con los tumores con fenotipo CIN.

### 4.1.3- Tumores con el fenotipo metilador de las islas CpG

En la zona promotora de la mayoría de genes existen regiones ricas en el dinucleotido citosina-guanina, llamadas islas CpG. En la célula normal, la metilación de las citosinas en estas islas regula la expresión génica de forma reversible. No obstante, un 15% de los CCR esporádicos tienen un fenotipo que se caracteriza por la metilación aberrante en los promotores de un conjunto de genes específicos (ej. *THBS1*, *p16* y *MLH1*) por lo que se han denominado tumores con fenotipo metilador de las islas CpG (CIMP del inglés *CpG island methylator phenotype*). (Toyota 1999). Debido a la hipermetilación aberrante de *MLH1*, este fenotipo puede solaparse con el de MSI en los tumores. Este fenotipo de metilación aberrante es característico, por ejemplo, de los pólipos serrados que progresan a CCR. (Dickinson 2015).

## 4.2- Formas clásicas del cáncer colorrectal hereditario

Una pequeña fracción del total de casos de CCR (hasta un 10%) está clasificada como CCR hereditario debido a que presenta una fuerte agregación familiar con un patrón de herencia mendeliano (Tomlinson 2015). Además, la edad de presentación es más temprana que en las formas esporádicas debido a que está causado por mutaciones germinales de alta penetrancia, por lo que solo se necesita un *hit* o evento mutacional somático adicional para desarrollar la enfermedad. En el caso de genes supresores de tumores, el hecho de nacer con un alelo mutado en todas las células del organismo hace que sea mucho más fácil la inactivación bialélica del gen en las formas hereditarias, pues solo sería necesaria una mutación somática en el alelo normal a lo largo de la vida del individuo (Knudson 1971) (**Figura 9**). Ésta sería la causa por la cual el riesgo es mayor y las edades de aparición son más tempranas en el caso del CCR hereditario. En el caso de oncogenes o genes supresores de tumores con haploinsuficiencia, no sería necesario el segundo evento somático para provocar la tumorigenesis.

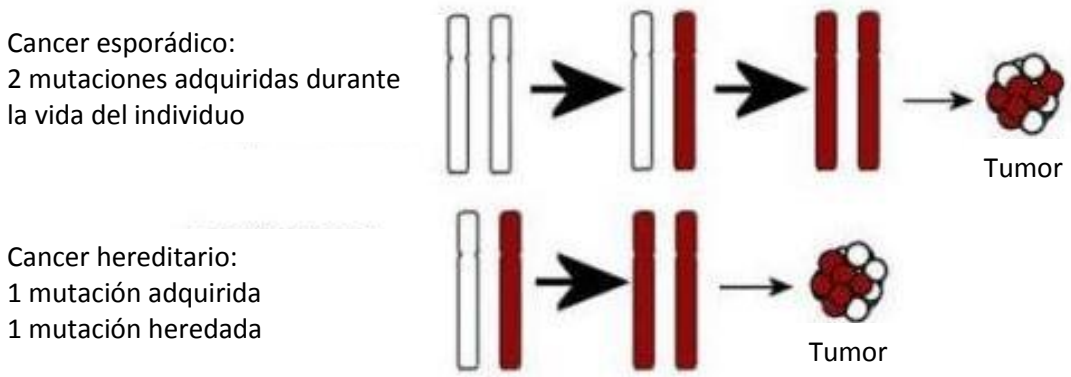



**Figura 9. Hipótesis de los 2 hits (o eventos inactivadores) de Knudson para genes supresores de tumores**.

Existe un abanico relativamente amplio de síndromes hereditarios que predisponen a poliposis colónica, dependiendo el tipo de lesión precursora mayoritaria y del gen afectado. Sin embargo la forma hereditaria más frecuente es el síndrome de Lynch, el cual predispone mayoritariamente a CCR en ausencia de pólipos. Además de las formas hereditarias clásicas (**Figura 10**) se han descubierto nuevos síndromes causados por otros genes de predisposición a CCR hereditario como *GREM1, BUB1, BUB3, POLE, POLD1, RPS20, SEMA4A, NTHL1, FAN1 y BLM* (Jaeger 2012, de Voer 2013, Palles 2013, Nieminen 2014, Schulz 2014, Weren 2015b, Seguí 2015, de Voer 2015) pero en la mayoría de hospitales aún no están incluídos en el

programa de diagnóstico molecular de esta enfermedad debido a su relativa novedad, la falta de una clara caracterización del fenotipo clínico y su baja frecuencia.

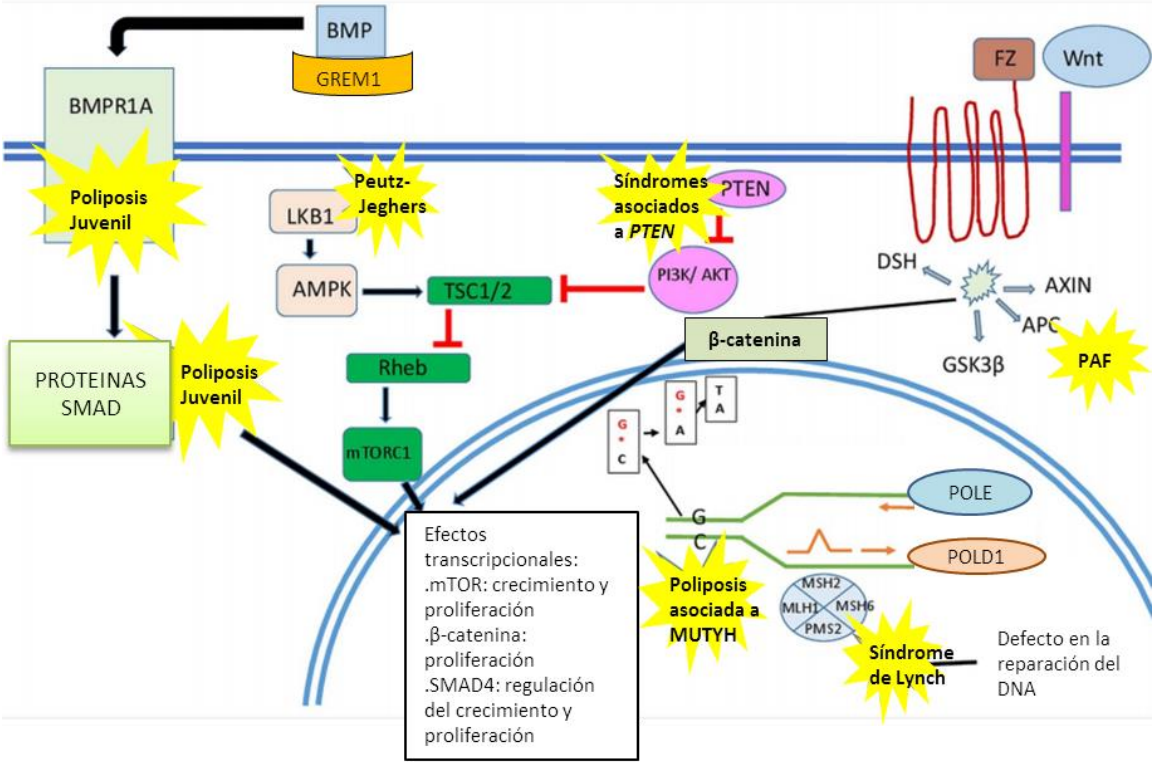



**Figura 10. Diagrama ilustrando las relaciones entre genes y vías de señalización implicadas en los síndromes de CCR hereditarios.** Abreviaturas: PAF =Poliposis Adenomatosa Familiar. Adaptado de Short (2015).

### 4.2.1- Poliposis adenomatosas

Las poliposis adenomatosas se definen como la presentación de un número anormalmente alto de adenomas en el colon y el recto. Esta categoría engloba dos formas clásicas de CCR hereditario, véase la poliposis adenomatosa familiar y la poliposis asociada al gen *MUTYH*.

#### 4.2.1.1- Poliposis adenomatosa familiar (PAF)

La poliposis adenomatosa familiar (PAF) es el síndrome polipósico más común, y está caracterizado por el desarrollo de cientos a miles de pólipos en el colon y el recto. En la mayoría de los pacientes los pólipos empiezan a desarrollarse durante la segunda década de la vida y casi el 100% de los pacientes sin tratar desarrollan tumores entre los 40 y los 50 años (Half 2009). Los pacientes con PAF también tienen predisposición a desarrollar una serie de manifestaciones extracolónicas, incluyendo fibromas, lipomas, quistes sebáceos y epidermioides, osteomas faciales, hipertrofia congénita del epitelio pigmentario de la retina, tumores desmoides y cánceres extracolónicos (tiroides, hígado, tracto biliar y sistema nervioso

central) (Half 2009). El cáncer de duodeno es la segunda neoplasia más común, con un riesgo acumulado del 4% al 12%. Los pólipos adenomatosos se encuentran también en el estómago y el duodeno, especialmente en el área periampular y tienen potencial malignizante (Bülow 2004) Algunas lesiones como los osteomas de cráneo y mandíbula, anormalidades dentales y fibromas son indicativos del síndrome de Gardner, una variante clínica de la FAP donde las características extracolónicas son prominentes. La PAF está causada por mutaciones germinales en el gen *APC*, que juega un papel importante en la vía de Wnt (**Figura 11**). Esta vía de señalización celular activa la proliferación responsable de la renovación constante de las células del colon (Groden 1991). La inactivación bialélica del gen *APC* debido a una mutación somática en el otro alelo o a una metilación aberrante desregulará esta vía activando una proliferación celular descontrolada. Esto provocará la formación de adenomas, que progresarán hacia CCR a medida que se acumulen otras mutaciones en oncogenes y genes supresores de tumores. El patrón de herencia para la PAF es autosómico dominante, por tanto en las familias portadoras existe una fuerte agregación, aunque aproximadamente un 25% de los casos son causados por mutaciones de novo (Jasperson 2010). Algunas mutaciones situadas en los extremos 5' y 3' o en el exón 9 de *APC* provocan una poliposis más atenuada, con un número de pólipos inferior a 100 y una edad de presentación más tardía, así como un riesgo de desarrollar CCR del 70% (van der Luijt 1996, Pineda 2009).

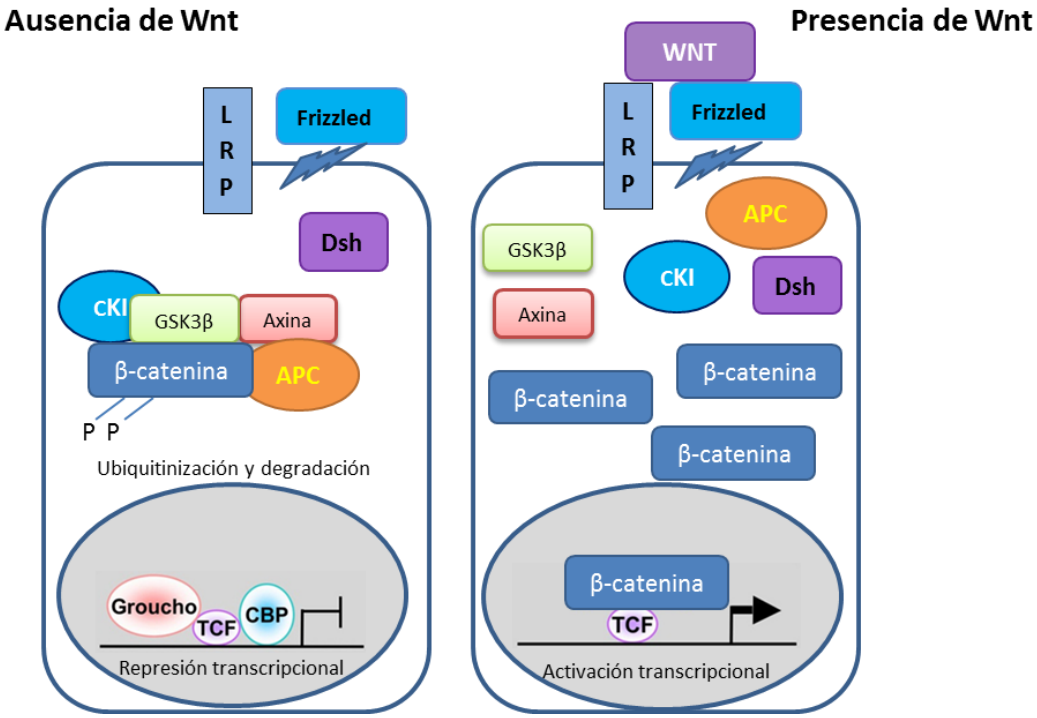



**Figura 11. Representación de los estados inactivado (izquierda) y activado (derecha) de la vía canónica de Wnt.** Adaptado de Korkaya y Wicha 2010.

#### 4.2.1.2- Poliposis asociada a *MUTYH* (PAM)

La poliposis asociada a *MUTYH* (PAM) se caracteriza por una poliposis atenuada (10-100 pólipos adenomatosos) con una edad de presentación alrededor de los 50 años (Al-Tassan 2002, Sampson 2003, Farrington 2005) y una probabilidad de desarrollar CCR en los pacientes sin tratar del 48% al 100% (Sampson 2003, Gismondi 2004, Farrington 2005). Una fracción de pacientes tienen CCR en ausencia de pólipos (Farrington 2005, Cleary 2009). La PAM está provocada por mutaciones en el gen *MUTYH,* que codifica una glicosilasa que mediante la escisión de bases mal aparejadas repara el daño en el ADN provocado por el estrés oxidativo (Al-Tassan 2002) (**Figura 12**).

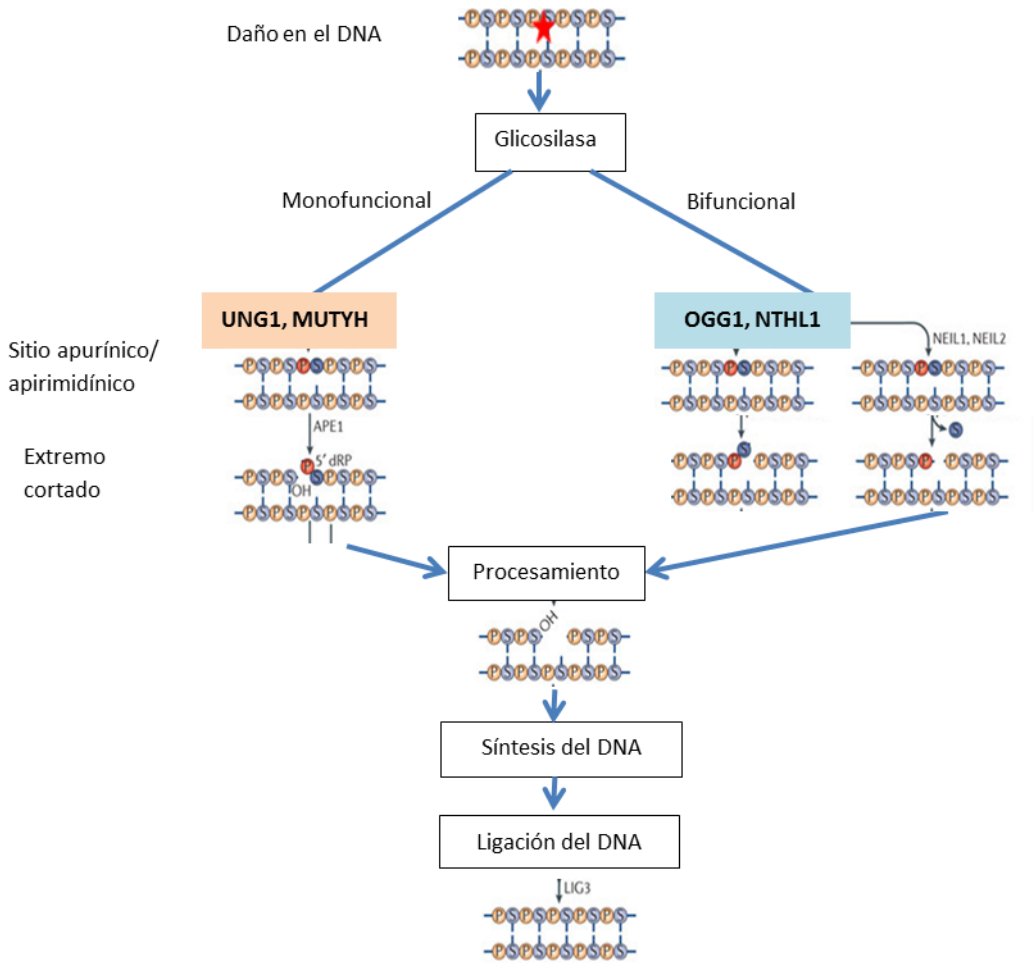



**Figura 12. Mecanismo de reparación del ADN por escisión de bases (*base excision repair*, BER).** Las lesiones en las bases son reconocidas y eliminadas por ADN glicosilasas. Las glicosilasas monofuncionales (UNG, MUTYH) eliminan las bases dañadas, y el ADN es cortado por la endonucleasa APE1. Las glicosilasas bifuncionales (OGG1, NTHL1) van un paso más allá y cortan la base dañada o bien eliminan también la ribosa, requiriendo las proteínas APE1 o PKNP para procesar el extremo 3' del ADN. A continuación la ADN polimerasa 1 sintetiza el ADN necesario en el extremo 5' para que la ligasa3 (LIG3) pueda unir los extremos del ADN. Figura adaptada de Kazak (2012).

Los procesos metabólicos celulares liberan radicales libres de oxígeno que se incorporan a la guanina (8-oxo-G) y esto facilita que se apareje erróneamente con la adenina. La ausencia de la enzima MUTYH conlleva un aumento de transversiones G:C a T:A en el tumor y este aumento de la tasa de mutaciones puede afectar a genes implicados en cáncer, como por ejemplo la transversión c.34G>T (p.G12C) en el gen *KRAS* (Jones 2004). También son comunes este tipo de transversiones en el gen *APC* (Al-Tassan 2002, Jones 2002). El patrón de herencia para este síndrome es autosómico recesivo.

### 4.2.2- Síndrome de Lynch

El síndrome de Lynch es la forma de CCR hereditaria más frecuente representando del 1-5% de total de casos de CCR (Lynch 2009). Los individuos con este síndrome presentan un riesgo del 60-80% de desarrollar CCR a lo largo de su vida, con edad de aparición habitualmente antes de los 45 años. Además también tienen predisposición a padecer tumores extracolónicos, principalmente cáncer de endometrio, con un riesgo estimado a lo largo de la vida del 40-60%. El espectro del síndrome de Lynch incluye además otras neoplasias como las de ovario, estómago, vía biliar, intestino delgado, páncreas, uréter y pelvis renal, así como tumores sebáceos conocidos como síndrome de Muir-Torre, y tumores del sistema nervioso central (gliomas y astrocitromas, en la variante conocida como síndrome de Turcot). El síndrome de Lynch está causado por mutaciones germinales en los genes de reparación de las bases mal aparejadas del ADN (*mismatch repair genes*) *MSH2*, *MLH1*, *MSH6* y *PMS2* (Fishel 1993, Papadopoulos 1994, Nicolaides NC 1994, Miyaki 1997). Alteraciones en estos genes provocan la generación de tumores con MSI debido a que no es posible corregir los errores en estas secuencias repetitivas durante la replicación del ADN (Umar 2004) (**Figura 13**).

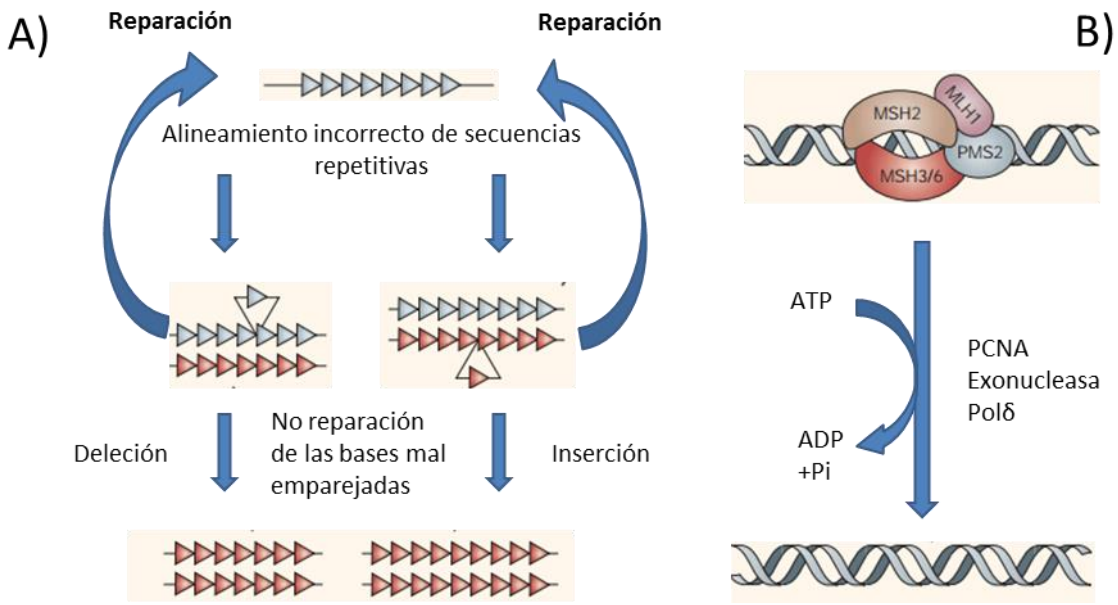



**Figura 13. Papel del sistema de reparación de las bases mal aparejadas del ADN en el mantenimiento de la longitud de los microsatélites.** A) Las secuencias repetitivas, representadas por "puntas de flechas" pueden quedar mal aliniadas en la cadena molde o bien en la de nueva síntesis durante la replicación del ADN. Estos deslizamientos son reparados por el sistema de reparación de las bases mal aparejadas del ADN (mismatch repair). Las consecuencias de la pérdida de este sistema se observan en el síndrome de Lynch (y en un número pequeño de cánceres esporádicos con inestabilidad de microsatélites). B) El sistema de reparación de las bases mal apareadas incluye *MSH2*, *MSH3/6*, *MLH1* y *PMS2*, y es capaz de reconocer los errores en el ADN y repararlos mediante la desfosforilación de ATP. Esta reparación necesita además proteínas adicionales como PCNA, una exonucleasa y la ADN polimerasa δ (Pol δ). Figura adaptada de Umar (2004).

Estas alteraciones genéticas se transmiten con un patrón de herencia autosómico dominante. El diagnóstico de esta enfermedad se basa en criterios de agregación familiar conocidos como los criterios de Ámsterdam, los cuales fueron sometidos a posteriores revisiones para incluir los cánceres extracolónicos dentro del concepto de la agregación familiar. La sensibilidad de estos criterios para detectar pacientes portadores de las mutaciones es inferior al 60% (Umar 2004). Como consecuencia, los criterios de Betheshda fueron creados para recuperar poder diagnóstico e identificar los afectos de CCR que no cumplían los criterios de Ámsterdam pero que debían ser igualmente sometidos al diagnóstico molecular (**Tabla 2**).

**Tabla 2. Criterios de Ámsterdam II y de Bethesda revisados para el diagnóstico clínico del síndrome de Lynch.**

**Criterios de Ámsterdam II:**

- Tres o más individuos afectados de CCR o neoplasia relacionada (endometrio, intestino delgado, uréter o pelvis renal), uno de ellos familiar de primer grado de los otros dos, y
- Afectación de 2 generaciones consecutivas y
- Como mínimo un caso diagnosticado antes de los 50 años, y
- Exclusión del diagnóstico de poliposis adenomatosa familiar

**Criterios de Bethesda revisados:**

- CCR diagnosticado antes de los 50 años
- Presencia de CCR sincrónico o metacrónico u otra neoplasia relacionada (endometrio, estómago, ovario, páncreas, urinario, cerebro, intestino delgado), con independencia de la edad
- CCR con infiltración linfocitaria, células en anillo de sello o crecimiento medular diagnosticado antes de los 60 años
- Paciente con CCR y uno o más familiares de 1º grado con CCR o neoplasia relacionada (endometrio, estómago, ovario, páncreas, urinario, cerebro, intestino delgado) diagnosticada antes de los 50 años
- Paciente con CCR y dos o más familiares de 1º o 2º grado con CCR o neoplasia relacionada (endometrio, estómago, ovario, páncreas, urinario, cerebro, intestino delgado), con independencia de la edad

Si estos criterios se cumplen, se procede a realizar el diagnóstico molecular que consiste en analizar la MSI en el tejido tumoral, así como las pruebas de inmunohistoquímica de las proteínas reparadoras (Suraweera 2002, Baudhuin 2005). Si existe MSI y pérdida de alguna de las proteínas reparadoras, se procede a la parte final del diagnóstico el síndrome de Lynch que consiste en secuenciar de forma completa y dirigida los genes que codifican para las proteínas perdidas en la inmunohistoquímica hasta detectar la mutación germinal causante de la predisposición.

### 4.2.3- Poliposis hamartomatosas

Los pólipos hamartomatosos están compuestos de elementos celulares normales del tracto gastrointestinal, pero tienen una arquitectura característicamente distorsionada y pueden progresar a CCR. Las poliposis hamartomatosas incluyen tres síndromes caracterizados por una predisposición a desarrollar estos pólipos junto con otras características clínicas específicas de cada síndrome. Todos juntos constituyen menos del 1% del total de casos de CCR

#### 4.2.3.1- Peutz-Jeghers

Está caracterizado por la presencia de pólipos hamartomatosos por todo el tracto gastrointestinal (principalmente en el intestino delgado aunque también en el estómago, colon y recto). Estos pólipos tienen una apariencia microscópica característica con una gran proliferación del músculo liso y un patrón de formación de pólipos elongado y arborescente (Estrada 1983). Otra característica distintiva de este síndrome es la presencia de una pigmentación azul oscura o marrón en los labios, mucosa oral, manos y pies. Los pacientes afectos tienen un alto riesgo de desarrollar diferentes tipos de cáncer a lo largo de su vida (incluyendo intestino delgado, estómago, páncreas, colon, esófago, mama, pulmón y cérvix) (Giardiello 2000, Lim 2004). En el caso del CCR el riesgo acumulado al largo de la vida es del 39% (Van Lier 2010). Para el 70-80% de los casos, la base genética de esta enfermedad reside en la inactivación inicial germinal del gen *STK11* (Hemminki 1998), cuyo producto inhibe la señalización de la vía de mTOR por AMPK, inhibiendo así el crecimiento celular, la proliferación y la angiogénesis (Corradetti 2004) **(Figura 14)**.

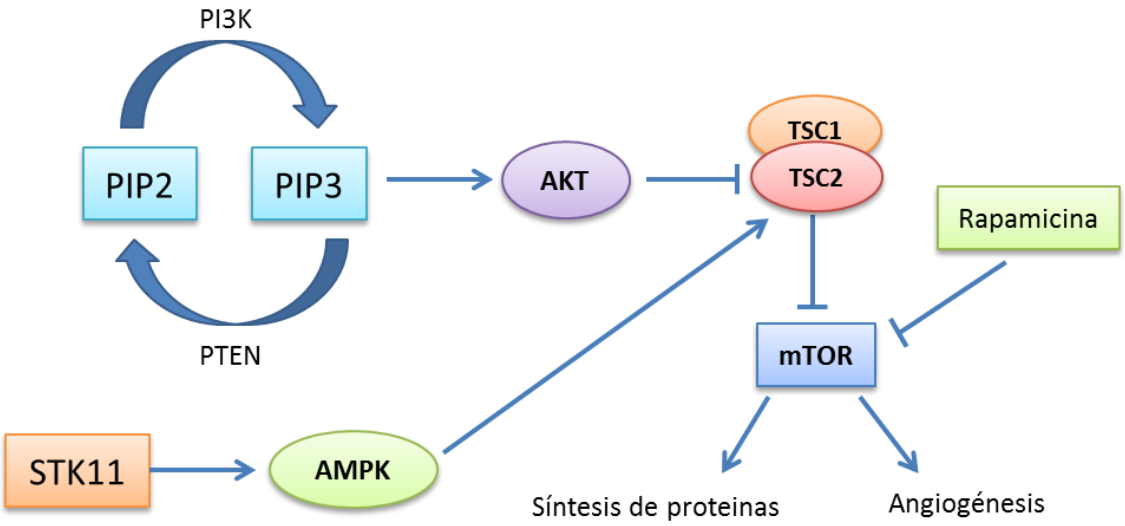



**Figura 14. Vías de señalización desteguladas en pacientes con el síndrome de Peutz–Jeghers o los síndromes hamartomatosos asociados a *PTEN*.** Tanto PTEN como STK11 inhiben la vía de mTOR, una vía de señalización promotora del crecimiento, que cuando se desregula contribuye a la patogénesis de estas enfermedades. Tanto PTEN como STK11 actúan de forma indirecta en mTOR a través de la modulación del complejo TSC1-TSC2 por encima de mTOR. Figura adaptada de Zbuk y Eng (2007).

#### 4.2.3.2- Síndromes hamartomatosos asociados a *PTEN*

Constituyen un grupo heterogéneo de hamartomatosis causadas por alteraciones en el gen *PTEN*, incluyendo la enfermedad de Cowden, el síndrome de Bannayan-Riley-Ruvalcaba (BRRS) y el síndrome de Proteus. La enfermedad de Cowden confiere un mayor riesgo de cáncer de

mama, tiroides y endometrio, así como el desarrollo de pólipos gastrointestinales, macrocefalia y lesiones mucocutáneas características (Nelen 1997). La relación entre esta enfermedad y un mayor riesgo de padecer CCR no está bien documentada, aunque se ha descrito un paciente con una mutación germinal en *PTEN* con dos tumores colónicos malignos surgidos a partir de dos pólipos hamartomatosos. Además, estos tumores tenían un segundo hit somático en *PTEN*, cumpliendo con la hipótesis de Knudson de la inactivación bialélica de los genes supresores de tumores (Bosserhof 2006). El BRRS está caracterizado por la presencia de lipomas múltiples, pólipos hamartomatosos gastrointestinales, macrocefalia, hemangiomas, retraso en el desarrollo y manchas pigmentadas en los genitales masculinos (Gorlin 1992). El síndrome de Proteus es un desorden complejo multisistémico caracterizado por malformaciones congénitas, hemihipertrofia, tumores hamartomatosos, nevos epidérmicos e hiperostosis (Wiedemann 1983). Se han encontrado mutaciones germinales en *PTEN* en el 85% de los pacientes con enfermedad de Cowden y en más del 60% de los pacientes con BRRS. En los pacientes afectos de Cowden estos cambios se han encontrado en la región promotora de *PTEN*, en cambio en los pacientes con BRRS se han detectado grandes deleciones que afectan todo el gen o parte de él. *PTEN* se expresa de forma ubicua y es una fosfatasa dual específica tanto de lípidos como proteínas. Su principal substrato es el fosfoinositol trifosfato. Al desfosforilar su substrato, PTEN inhibe la vía de señalización de mTOR por AKT (Suzuki 1998, Waite 2002). La vía de mTOR está desregulada también en Peutz-Jeghers.

#### 4.2.3.3- Poliposis juvenil

Este síndrome se define por la presencia de pólipos juveniles en el tracto gastrointestinal, que empiezan a desarrollarse a edades tempranas. Los pólipos juveniles se caracterizan por tener forma esférica, son lisos y brillantes en apariencia y a menudo son pedunculados. A nivel microscópico, no hay proliferación del músculo liso (a diferencia de los pólipos de Peutz-Jeghers) y muestran dilatación glandular (Jass 1988). A menudo se desarrollan decenas a cientos de pólipos en el tracto gastrointestinal, siendo el colon y el recto los lugares más afectados, seguidos por el estómago y el intestino delgado. Las personas afectas tienen además una mayor predisposición a padecer tumores gastrointestinales, especialmente CCR, con una edad de diagnóstico alrededor de los 43 años (Brosens 2007). Se han descrito mutaciones germinales en los genes SMAD4, BMPR1A y ENG como causantes de predisposición a la poliposis juvenil (Howe 1998, Howe 2001, Ngeow 2013). Estos tres genes están implicados en la vía del TGF-B, que regula la proliferación, diferenciación y adhesión celular (Waite 2003) (**Figura 15**).

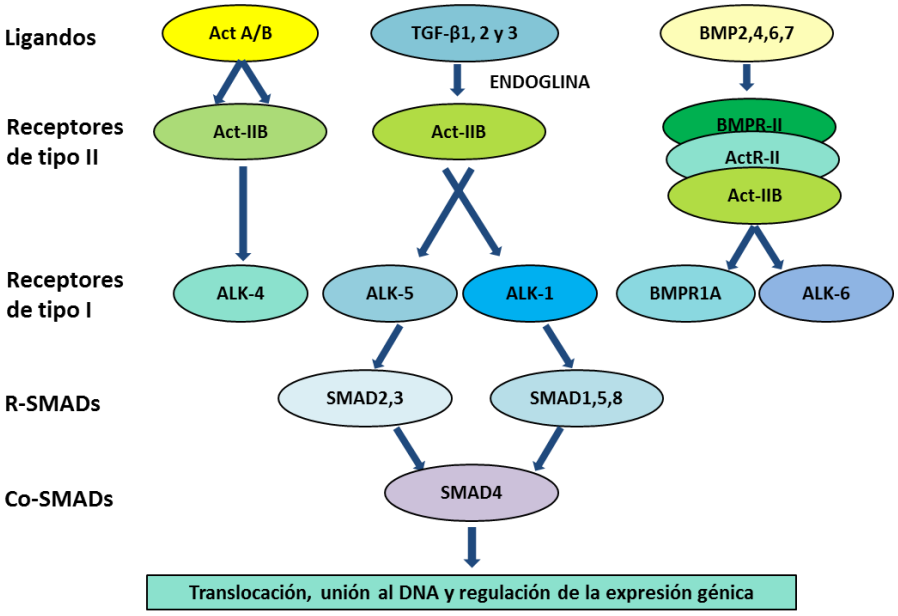



**Figura 15. Complejidad de la vía de señalización del TGF-beta.** Los ligandos de TGF-beta se unen a un rango de receptores de tipo II para formar complejos que interactuan con receptores de tipo I. Los receptores entonces forman heterotetrámeros que resultan en la fosforilación y activación de los receptores regulados por SMADs (R-SMADs), los cuales a su vez forman complejos con SMAD4. Este complejo se transloca al núcleo, donde regula la transcri`pción génica, directa o indirectamente. La endoglina (ENG) es un correceptor tanto para TGF-beta1 como para TGF-beta3

## 4.3- Cáncer colorrectal con agregación familiar

Aproximadamente un 30% de los casos de CCR presentan algún tipo de historia familiar para la enfermedad pero no entran en la categoría anterior de CCR hereditario y son denominados CCR familiar (Jasperson 2010). En un estudio epidemiológico de la población española esta categoría abarcaba cerca del 30% de los casos (Piñol 2005).

### 4.3.1- Cáncer colorrectal familiar de tipo X

De entre las familias que cumplen los criterios de Ámsterdam I, entre un 35 y un 50% no tienen mutaciones en los genes *MSH1, MSH2, MSH6 o* PMS2 y presentan tumores sin MSI. En base a la fuerte agregación para CCR existente, la hipótesis es que en estas familias se estarían transmitiendo variantes de alta penetrancia situadas en genes aún por descubrir en la mayoría de familias. Debido a ese desconocimiento se estableció el término de CCR familiar de tipo X (Lindor 2005). Varios estudios se han llevado a cabo con el fin de comparar las características clínicas de esta entidad con las del síndrome de Lynch y se ha observado que la edad media de presentación del CCR en los pacientes con CCR familiar de tipo X es unos 10 años mayor. Por otro lado, los individuos pertenecientes a las familias con CCR de tipo X presentan un riesgo menor de presentar CCR que los individuos pertenecientes a familias con síndrome de Lynch

(OR 2,3 en CCR familiar X frente a 6,1 en Lynch). Además los tumores colorrectales tienen una distribución más distal y es menos frecuente el desarrollo de CCR sincrónico o metacrónico, así como la formación de tumores extracolónicos (Schiemann 2004, Lindor 2005, Llor 2005, Mueller–Koch y col, 2005, Dove-Edwin 2006, Valle 2007, Chen 2008, Koh 2011, Klarskov 2012, Francisco 2011, Abdel-Rahman 2005, Woods 2005). Un estudio reciente comparó las características epidemiológicas e histopatológicas de los pacientes con CCR esporádico y CCR familiar tipo X, y no observó diferencias significativas en la edad de aparición del CCR ni en su localización, variando únicamente el grado de diferenciación de los tumores, el cual era significativamente menor en los pacientes con CCR familiar tipo X. De hecho se observaron más diferencias significativas al comparar las características de los pacientes con CCR familiar de tipo X y los pacientes con Síndrome de Lynch que entre los pacientes con CCR familiar de tipo X y los pacientes con CCR esporádico (Shiovitz 2014). En cuanto al perfil tumoral somático, se ha observado que es frecuente el fenotipo de CIN, el cual también es mayoritario en los tumores esporádicos. No obstante, se observa mayor frecuencia de ganancia del brazo cromosómico 20q y una frecuencia mayor de LOH de número de copia neutro por tumor que en los tumores esporádicos, habiendo en comparación un número menor de pérdidas cromosómicas (Laiho 2003, Middeldorp y col 2013, Chen 2013).

#### 4.3.2- cáncer colorrectal de aparición temprana

El CCR de aparición temprana (antes de los 40-50 años) comprende entre el 2% y el 8% de los casos de CCR (Silla 2014). Una edad temprana de aparición es sugestiva de un componente genético heredado. De hecho, tener un pariente de primer grado con CCR diagnosticado antes de los 45 aumenta el riesgo de padecer CCR de 3 a 6 veces en comparación con la población general y eso hace que exista cierta agregación familiar (Johns 2001). El CCR de aparición temprana con agregación familiar puede ser debido a formas sindrómicas como el síndrome de Lynch o bien puede tener causas genéticas heredadas desconocidas, como por ejemplo en el caso del CCR familiar de tipo X (Perea 2011, Hawkins 2001). No obstante, existe una proporción importante de CCR de aparición temprana que no presenta agregación familiar, llamado CCR de aparición temprana esporádico. En cuanto a la localización anatómica, los tumores suelen ser distales en los casos de CCR de aparición temprana esporádicos, mientras que CCR de aparición temprana con agregación familiar se ha encontrado una localización más proximal (Mahdavinia 2005). La histología de los tumores colorrectales de aparición temprana en general tiende a ser mucinosa (9%-49% frente al 17% de todos los CCR), tienden a estar poco diferenciados (12-98% frente al 15% de todos los CCR) y suelen diagnosticarse en estadios más avanzados de la tumorigenesis (Stigliano 2014). Molecularmente, se ha

observado que estos tumores tienen las regiones LINE-1 hipometiladas en comparación con los CCR de aparición más tardía (Antelo 2012). Este subgrupo de CCR de aparición temprana se puede subdividir en tumores con MSI correspondiendo a los casos con síndrome de Lynch, así como tumores con el sistema reparador intacto, pero con CIN. También se ha observado un subgrupo de tumores con el sistema reparador intacto pero sin estabilidad cromosómica, los cuales presentan una longitud telomérica anormalmente grande independiente de la activación de la telomerasa (Boardman 2013)

#### 4.3.3- Poliposis adenomatosa atenuada no filiada

Existe una variante de la poliposis adenomatosa familiar llamada poliposis atenuada que se caracteriza por un número menor de pólipos (menos de 100) y una presentación de estos a edades más avanzadas (siendo el promedio los 44 años). El riesgo de padecer CCR a lo largo de la vida sería del 70%, con una edad de aparición 10-15 años después de desarrollar los pólipos. Un 30% de los casos de poliposis atenuada se explica por mutaciones en el gen *APC* que están mayoritariamente localizadas en los extremos 5' y 3', así como en el exón 9 (Pineda 2009, van der Luijt 1996). Un 40% de los casos restantes es debido a mutaciones bialélicas en *MUTYH* (Sieber 2003, Sampson 2003). Otro gen con mutaciones patogénicas encontradas en casos con poliposis adenomatosa atenuada es *BMPR1A*, previamente implicado en poliposis juvenil (Lipton 2003). El descubrimiento de los genes *POLE* y *POLD1* también ha caracterizado molecularmente parte de los casos de poliposis atenuada (Palles 2013). Sin embargo aún existe una fracción de casos con poliposis adenomatosa atenuada sin causa genética conocida, siendo también el caso de la poliposis mixta y la poliposis serrada no sindrómica.

#### 4.3.4.- Poliposis hereditaria mixta

La poliposis hereditaria mixta constituye una forma sindrómica de penetrancia variable en la que se incluye el desarrollo de pólipos de diferente naturaleza: pólipos hiperplásicos, de tipo Peutz-Jeghers, serrados, juveniles y también adenomas. El CCR ocurre además en una alta proporción de familias afectas pero no hay predisposición aumentada a tumores extracolónicos (Whitelaw SC 1997). Se han encontrado mutaciones en *BMPR1A* en diferentes familias con este síndrome (Cao 2006, Cheah 2009, O'Riordan 2010). Posteriormente, se encontró en una familia judía Askenazí afectada una duplicación de 40 kb situada upstream del gen *GREM1* como causante de la enfermedad (Jaeger 2012). El cribado en casos y controles permitió encontrar la duplicación en una familia independiente también de origen judío Askenazí.

### 4.3.5- Síndrome de la poliposis serrada

Este síndrome está caracterizado por la presentación de pólipos serrados distribuidos por todo el colon y conlleva un riesgo aumentado de desarrollar CCR y cáncer de páncreas a lo largo de la vida (Boparai 2009, Win 2012). Estos pólipos tienen ciertas características distintivas: pueden ser grandes (mayores de 1 cm de diámetro), son generalmente numerosos y suelen estar localizados en el colon proximal. No obstante, también pueden ser pequeños y tener una distribución más dispersa. La Organización Mundial de la Salud (OMS) estableció unos criterios de diagnóstico de este síndrome (Snover 2010).

- Presencia de al menos 5 pólipos serrados proximales a sigma, siendo 2 de ellos iguales o mayores a 10mm.
- Cualquier número de pólipos serrados proximales al colon sigmoide en un individuo con antecedentes familiares de primer grado (padres, hermanos, hijos) de síndrome de poliposis serrada.
- Más de 20 pólipos serrados de cualquier tamaño distribuidos a lo largo del colon.

Estos criterios son una herramienta provisional para la inclusión de pacientes en estudios de investigación, y sin duda se refinará a medida que las bases moleculares se conozcan mejor. Se ha observado que en pacientes afectos de este síndrome los pólipos serrados presentan a menudo el fenotipo metilador de las islas CpG, e incluso se ha encontrado que la mucosa normal de estos pacientes presenta a veces este fenotipo metilador (Chan 2002, Minoo 2006). El gen *BRAF* se encuentra a menudo mutado somáticamente los pólipos serrados de los pacientes con este síndrome (Beach 2005). Actualmente se desconocen las causas genéticas germinales de predisposición, a excepción de una minoría de casos portadores de mutaciones en *MUTYH* que además del síndrome presentan adenomas tubulares (Boparai 2008). Un estudio reciente en individuos con múltiples adenomas serrados sésiles identificó 5 genes candidatos portadores de mutaciones germinales implicados en la vía de senescencia celular: *ATM*, *PIF1*, *TELO2*, *XAF1* y *RBL1*. Además se identificó una mutación germinal (p.R113X) en dos familias independientes en el gen *RNF43*, que está implicado en la vía de Wnt. Estudios funcionales demostraron que en las células con el gen *RNF43* truncado la respuesta al daño en el ADN estaba comprometida (Gala 2013). Un estudio posterior reportó otra mutación germinal en el gen *RNF43,* que segregaba con patrón autosómico dominante en los afectos de una familia con síndrome de poliposis serrada, afianzando el papel de este gen como candidato de predisposición a la enfermedad (Taupin 2015).

## 5- Técnicas moleculares para descubrir genes de predisposición al cáncer colorrectal

Existen diferentes estrategias a nivel molecular que permiten identificar nuevos genes de predisposición al CCR familiar.

### 5.1- Estudios de desequilibrio de ligamiento

La técnica del análisis del ligamiento estudia marcadores informativos como microsatélites o SNPs repartidos por todo el genoma y su segregación dentro de una familia con agregación para una enfermedad. Se seleccionan las regiones del genoma que se transmiten en bloque (sin ninguna recombinación de los cromosomas) entre los miembros afectados y se calcula la probabilidad de que esa transmisión en bloque sea por azar. Cuando el número de veces que una región se hereda en bloque entre los individuos afectos supera el número de veces que se esperaría por azar se dice que esa región está en desequilibrio de ligamiento con la enfermedad dentro de esa familia. Estos estudios son útiles para detectar variantes de alta penetrancia que se transmiten con herencia Mendeliana. Mediante estos estudios se han detectado la mayoría de genes responsables de los síndromes hereditarios clásicos. Además se han hecho estudios posteriores en familias con CCR familiar de tipo X, detectándose múltiples regiones en desequilibrio de ligamiento, pero en la mayoría de casos no se han encontrado genes candidatos definitivos dentro de estas regiones (**Tabla 3**).

**Tabla 3. Regiones descubiertas en estudios de desequilibrio de ligamiento en CCR familiar de tipo X.**

| Autores | Regiones identificadas |
|---|---|
| Wiesner y col Proc Natl Acad Sci USA 2003 | 9q22.2–31.2 |
| Djureinovic y col Gut 2006 | 11q<br>14q<br>22q |
| Kemp y col Hum Mol Genet 2006 | 3q21–q24 |
| Skoglund y col J Med Genet 2006 | 9q22.2–31.2 |
| Neklason y col Cancer Res 2008 | 7q31<br>3q21–q24 |
| Papaemmanuil y col Eur J Hum Genet 2008 | 3q21 |
| Picelli y col BMC Cancer 2008 | 3q21.1–q26.2 |
| Middeldorp y col Genes Chromosom Cancer 2010 | 3q |
| Gray-McGuire y col Cancer Res 2010 | 9q22 |
| Daley y col Am J Hum Genet 2008 | 1p31.1<br>15q14-q22<br>17p13.3 |
| Nieminen y col Gastroenterology 2011 | 10q23 *(BMPR1A)* |
| Cicek y col PLoS One 2012 | 4q21<br>8q13<br>12q24<br>15q22 |
| Sánchez-Tomé y col J Gastroenterol 2014 | 2p24.3<br>4q13.1<br>4q31.21<br>12q21.2–q21.31 |
| Rudkjøbing y col Fam Cancer 2015 | 11q24 |

La excepción sería el gen *BMPR1A* situado en la región 10q23, que se encontró mutado en una familia con CCR familiar de tipo X mediante el estudio de ligamiento combinado con secuenciación dirigida (Nieminen 2011)

**5.2- Estudios de asociación del genoma completo**

Los estudios de asociación del genoma completo (*Genome-wide association studies*, GWAS) se basan en el genotipado en casos y controles de miles de variantes comunes distribuidas por todo el genoma llamadas SNPs (*single nucleotide polymorphisms*). Las variantes que estén significativamente enriquecidas en los casos en comparación con los controles son seleccionadas como variantes de baja penetrancia que confieren un riesgo bajo pero acumulativo de padecer la enfermedad. Como se ha mencionado anteriormente, hasta el momento se han descrito 62 SNPs asociados a susceptibilidad a CCR en 56 loci (**Tabla 4**). Estos estudios han descubierto varias regiones asociadas que afectan genes pertenecientes a vías de señalización implicadas en cáncer como la vía del TGF-B, que regula el crecimiento celular, así como la diferenciación celular y la apoptosis (ej. *BMP2, BMP4, SMAD7, CCND2, GREM1*). También se han encontrado distintas variantes de baja penetrancia que afectan genes de la vía de señalización de las proteína-quinasas activadas por mitogénos, proceso inductor de la división celular, supervivencia y apoptosis (ej. *DUSP10, DUSP4, MYOB, MYC, CCND2, SH2B*). Otro descubrimiento que ha surgido a partir de la identificación de estas variantes de baja penetrancia es que un mismo gen puede estar implicado en predisposición a CCR como portador de variantes de alta y baja penetrancia, teniendo distintas variantes un efecto más o menos fuerte en el riesgo de presentar la enfermedad. Un ejemplo sería el gen *GREM1* para el cual recientemente se han encontrado mutaciones que conferirían un alto riesgo de padecer CCR (Venkatachalam 2011, Jaeger 2012, Rohlin 2016) y para el cual existen también variantes de baja penetrancia que conferirían un riesgo modesto de padecer la enfermedad (Tomlinson 2011).

**Tabla 4. SNPs detectados en estudios de asociación del genoma completo.** Hasta ahora se han encontrado 62 SNPs en 56 loci. Crom, región cromosómica.

| SNP | Crom | Gen | (casos/ controles) | OR (95%CI) | Referencias |
|---|---|---|---|---|---|
| rs6983267 | 8q24.21 | *MYC* | 8.264/6.206 | 1,21(1,15-1,27) | Tomlinson I, y col. Nat Genet 2007 |
| rs719725 | 9p24 | *TPD52L3-UHRF2* | 6223/ 6443 | 1,14(1,05- 1,15) | Zanke BW y col. Nat Genet 2007 |
| rs4939827 | 18q21.1 | *SMAD7* | 18.831/18.540 | 1,18 (1,12-1,23) | Tomlinson IP, y col. Nat Genet 2008 |
| rs7229639 | 18q21 | *SMAD7* | 5385/6165 | 1,22(1,15- 1,29) | Zhang B, y col. Int J Cancer 2014a |
| rs16892766 | 8q23.3 | *EIF3H* | 18.831/18.540 | 1,25 (1,19-1,32) | Tomlinson IP, y col. Nat Genet 2008 |
| rs3802842 | 11q23.1 | *COLCA2/ COLCA1* | 14.500/13.294 | 1,12 (1,07-1,17) | Tenesa A, y col. Nat Genet 2008 |
| rs4779584 | 15q13.3 | *GREM1* | 18.831/18.540 | 1,26 (1,19-1,34) | Tomlinson IP, y col. Nat Genet 2008 |
| rs10795668 | 10p14 | *ARN5SP299/ GATA3* | 18.831/18.540 | 1,12 (1,10-1,16) | Tomlinson IP, y col. Nat Genet 2008 |
| rs4444235 | 14q22.2 | *BMP4* | 20.288/20.971 | 1,11 (1,08-1,15) | Houlston y col. Nat Genet 2008 |
| rs9929218 | 16q22.1 | *CDH1* | 20.288/20.971 | 1,10 (1,06-1,12) | Houlston y col. Nat Genet 2008 |
| rs10411210 | 19q13 | *RHNP2* | 20.288/20.971 | 1,15 (1,10-1,20) | Houlston y col. Nat Genet 2008 |
| rs961253 | 20p12.3 | *BMP2* | 20.288/20.971 | 1,12 (1,08-1,16) | Houlston y col. Nat Genet 2008 |
| rs6691170 | 1q41 | *DUSP10* | 18.185/20.197 | 1,09(1,06–1,12) | Houlston y col. Nat Genet 2010 |
| rs7136702 | 12q13.13 | *LARP4/ DIP2B* | 18.185/20.197 | 1,06(1,04- 1,08) | Houlston y col. Nat Genet 2010 |
| rs10936599 | 3q26.2 | *TERC* | 18.185/20.197 | 0,93 (0,91-0,96) | Houlston y col. Nat Genet 2010 |
| rs11169552 | 12q13.3 | *DIP2B/ ATF1* | 18.185/20.197 | 0,92 (0,90-0,95) | Houlston y col. Nat Genet 2010 |
| rs4925386 | 20q13.33 | *LAMA5* | 18.185/20.197 | 0,93 (0,91-0,95) | Houlston y col. Nat Genet 2010 |
| rs1957636 | 14q22.2 | *BMP4* | 24.910/26.275 | 1,08 (1,05-1,11) | Tomlinson y col. PLoS Genet 2011 |
| rs4813802 | 20p12.3 | *BMP2* | 24.910/26.275 | 1,09 (1,06-1,12) | Tomlinson y col. PLoS Genet 2011 |
| rs2736100 | 5p15.33 | *TERT* | 16.039/16.430 | 1,07 (1,04-1,10) | Kinnersley y col. Br J Cancer 2012 |
| rs1321311 | 6p21 | *CDKN1A* | 21.096/19.555 | 1,10 (1,07-1,13) | Dunlop y col. Nat Genet 2012 |
| rs3824999 | 11q13.4 | *POLD3* | 21.096/19.555 | 1,10 (1,07-1,13) | Dunlop y col. Nat Genet 2012 |
| rs5934683 | Xp22.2 | *SHROOM2* | 21.096/19.555 | 1,07 (1,04-1,10) | Dunlop y col. Nat Genet 2012 |
| rs12080929 | 1p33 | *SLC5A9* | 2.317/2.447 | 0,86 (0,78-0,95) | Fernandez-Rozadilla y col. BMC Genomics 2013a |
| rs11987193 | 8p12 | *DUSP4* | 2.317/2.447 | 0,78 (0,70-0,87) | Fernandez-Rozadilla y col. BMC Genomics 2013a |
| rs10774214 | 12p13.32 | *CCND2* | 11.870/14.190 | 1,04 (1,00-1,09) | Jia WH y col. Nat Genet. 2013 |
| rs647161 | 5q31.1 | *PITX1* | 11.870/14.190 | 1,07 (1,02-1,11) | Jia WH y col. Nat Genet. 2013 |
| rs2423279 | 20p12.3 | *HAQ1* | 11.870/14.190 | 1,07 (1,03-1,12) | Jia WH y col. Nat Genet. 2013 |
| rs11903757 | 2q32.3 | *NABP1* | 15.752/21.771 | 1,16 (1,10-1,22) | Peters y col. Gastroenterology 2013 |

| SNP | Crom | Gen | (casos/ controles) | OR (95%CI) | Referencias |
|---|---|---|---|---|---|
| rs10911251 | 1q25.3 | *LAMC1* | 15.752/21.771 | 1,09 (1,06-1,13) | Peters y col. Gastroenterology 2013, Whiffin y col Hum Mol Genet 2014 |
| rs3217810 | 12p13.32 | *CCND2* | 13.654/16.022 | 1,20 (1,12-1,28) | Peters y col. Gastroenterology 2013, Whiffin y col Hum Mol Genet 2014 |
| rs3217901 | 12p13.32 | *CCND2* | 15.752/21.771 | 1,10 (1,06-1,14) | Peters y col. Gastroenterology 2013 |
| rs59336 | 12q24.21 | *TBX3* | 15.752/21.771 | 1,09 (1,06-1,13) | Peters y col. Gastroenterology 2013 |
| rs704017 | 10q22.3 | *ZMIZ1-AS1* | 14.963/31.945 | 1,10 (1,06-1,13) | Zhang y col. Nat Genet. 2014b |
| rs11196172 | 10q25.2 | *TCF7L2* | 14.963/31.945 | 1,14(1,10–1,18) | Zhang y col. Nat Genet 2014b |
| rs174537 | 11q12.2 | *MYRF* | 14.963/31.945 | 1,16(1,12–1,19) | Zhang y col. Nat Genet 2014b |
| rs4246215 | 11q12.2 | *FEN1* | 14.963/31.945 | 1,15(1,12–1,19) | Zhang y col. Nat Genet 2014b |
| rs174550 | 11q12.2 | *FADS1* | 14.963/31.945 | 1,15(1,12–1,19) | Zhang y col. Nat Genet 2014b |
| rs1535 | 11q12.2 | *FADS2* | 14.963/31.945 | 1,15(1,12–1,19) | Zhang y col. Nat Genet 2014b |
| rs10849432 | 12p13.31 | *CD9* | 14.963/31.945 | 1,14(1,09–1,18) | Zhang y col. Nat Genet 2014b |
| rs12603526 | 17p13.3 | *NXN* | 14.963/31.945 | 1,10(1,06–1,14) | Zhang y col. Nat Genet 2014b |
| rs1800469 | 19q13.2 | *TGFB1* | 14.963/31.945 | 1,09(1,06–1,12) | Zhang y col. Nat Genet 2014b |
| rs2241714 | 19q13.2 | *B9D2* | 14.963/31.945 | 1,09(1,06–1,12) | Zhang y col. Nat Genet 2014b |
| rs812481 | 3p14.1 | *LRIG1* | 23.024/29.625 | 1,09(1,05-1,11) | Schumacher y col. Nat Commun 2014 |
| rs35360328 | 3p22.1 | *CTNNB1* | 23.024/29.625 | 1,14(1,09-1,19) | Schumacher y col. Nat Commun 2014 |
| rs3184504 | 12q24.12 | *SH2B3* | 23.024/29.625 | 1,09(1,06-1,12) | Schumacher y col. Nat Commun 2014, Cheng TH y col Scientific Reports 2015 |
| rs73208120 | 12q24.22 | *NOS1* | 23.024/29.625 | 1,16(1,11- 1,23) | Schumacher y col. Nat Commun 2014 |
| rs6066825 | 20q13.13 | *PREX1* | 23.024/29.625 | 1,09(1,06-1,12) | Schumacher y col. Nat Commun 2014 |
| rs11190164 | 10q24.2 | *?* | 23.024/29.625 | 1,09(1,06-1,12) | Schumacher y col. Nat Commun 2014 |
| rs1035209 | 10q24.2 | *ABCC2/MRP2* | 19.663/23.754 | 1,12(1,08-1,16) | Whiffin y col Hum Mol Genet 2014 |
| rs3987 | 4q26 | *NDST3* | 923/2183 | 1,36 | Real LM y col Plos One 2014 |
| rs72647484 | 1p36.2 | *CDC42/WNT4* | 7577/9979 | 1,24(1,15-1,33) | Al-Tassan y col. Scientific Reports 2015 |
| rs16941835 | 16q24.1 | *RP11-58A18.1/ FOXL1* | 7577/9979 | 1,24(1,15-1,33) | Al-Tassan y col. Scientific Reports 2015 |
| rs10904849 | 10p13 | *CUBN* | 7577/9979 | 1,13(1,08-1,19) | Al-Tassan y col. Scientific Reports 2015 |
| rs17094983 | 14q23.1 | *RTN1* | 16.517/14.487 | 0,87(0,83-0,91) | Lemire y col Hum Genet. 2015 |
| rs12241008 | 10q25.2 | *VTI1A* | 21.344/26.711 | 1,19(1,12-1,26) | Wang H y col. Nat Commun. 2014 |
| rs12970291 | 18q22.3 | *TSHZ1* | 13.365/40.245 | 1,27(1,16-1,38) | Cheng TH y col Scientific Reports 2015 |
| rs4711689 | 6p21.1 | *TFEB* | 19.071/34.624 | 1,12(1,06-1,18) | Zeng y col, Gastroenterology 2016 |
| rs6469656 | 8q23.3 | *EIF3H* | 19.071/34.624 | 1,12(1,07-1,16) | Zeng y col, Gastroenterology 2016 |
| rs11064437 | 12p13.3 | *SPSB2* | 19.071/34.624 | 1,14(1,09-1,20) | Zeng y col, Gastroenterology 2016 |
| rs6061231 | 20q13.3 | *RPS21* | 19.071/34.624 | 1,19(1,12-1,27) | Zeng y col, Gastroenterology 2016 |
| rs992157 | 2q35 | *PNKD/ TMBIM1* | 13.656/21.667 | 1,10(1,06-1,13) | Orlando G y col HMG 2016 |

## 5.3- Detección de variantes de número de copia

Las variantes de número de copia o CNVs (*copy number variants*) son un tipo de variación genética estructural que consiste en duplicaciones o deleciones de fragmentos de ADN mayores de 1 kilobase. Pueden encontrarse de forma germinal como variación común en la población general o bien tratarse de variantes patogénicas causantes de enfermedades. En el caso del cáncer también se pueden encontrar como alteraciones somáticas (Scherer 2007) (**Figura 16**).

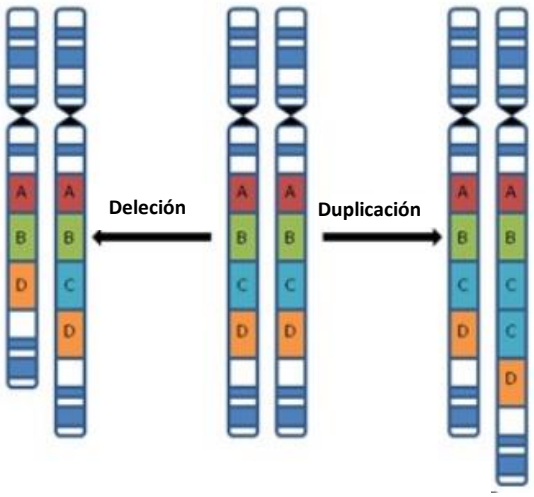



**Figura 16. Alteraciones de número de copia (CNV): deleción y duplicación.**

Las aberraciones cromosómicas en forma de ganancias o pérdidas de número de copia son comunes en cáncer y su estudio es una forma de identificar y validar genes implicados en la tumorigenesis (Venkatraman 2007). Para detectar CNVs en todo el genoma de individuos con un fenotipo concreto (como el síndrome de Lynch o el CCR familiar) existen técnicas como los arrays de genotipado de SNPs y los arrays de hibridación genómica comparada o CGH (*comparative genomic hybridization*). Los arrays de genotipado de SNPs consisten en fragmentar una muestra de ADN y marcarla con fluorescencia para después hibridarla con un *array* de sondas específicas de SNPs de todo el genoma. Si hay una CNV se detectará una descompensación en la señal fluorescente en varias sondas contiguas en el genoma que estará reflejando una descompensación alélica debido a una duplicación o deleción. De forma similar, la hibridación genómica comparada o CGH consiste a grandes rasgos en fragmentar el ADN de una muestra control y una muestra a estudiar, marcar cada una con un fluoróforo distinto y a continuación hibridarlas con un array de ADN. La intensidad de la señal de un fluoróforo comparado con el otro permitirá detectar con precisión las regiones ganadas o perdidas en la

muestra a estudiar en relación con la muestra control. Mediante estas técnicas es posible detectar CNVs con una gran resolución, permitiendo identificar genes concretos. (Holcomb y Trask 2007) (**Figura 17**).

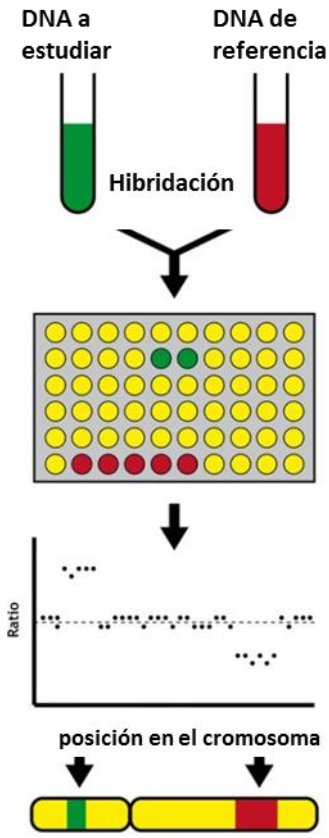



**Figura 17.Técnica de hibridación genómica comparada para detectar CNV en el genoma de una muestra problema en comparación a una muestra de referencia**. Imagen adaptada de http://www.phgfoundation.org/array_cgh_ld

Para detectar amplificaciones o deleciones de forma dirigida a regiones específicas del genoma se utilizan técnicas como el MLPA (*Multiplex Ligation-dependent Probe Amplification*) que consiste en la la ligación de dos fragmentos de ADN específicos con primers en los extremos que al unirse forman una sonda que se amplificará y producirá señal. Si hay una CNV se detectarán cambios en la señal en esa zona respecto a la muestra control (**Figura 18**).


**Figura 18. Reacción de MLPA.** Adaptado de www.mlpa.com.

A raíz de estas metodologías se detectó que deleciones en el gen *EPCAM* provocaban síndrome de Lynch mediante una metilación aberrante del gen *MSH2* (Ligtenberg 2009) Esto es debido a que el promotor de *MSH2* está situado a continuación de *EPCAM* y se transcribe en el mismo sentido. En consecuencia, delecciones en el extremo *3'* de *EPCAM* que incluyan la cola de poliadenilación alterarán la terminación de la transcripción y esto causará un silenciamiento epigenético de la región promotora de *MSH2*. Con el objetivo de encontrar nuevos genes de predisposición al CCR familiar, se han buscado CNVs raras como mecanismo mutacional, siguiendo distintas estrategias (**Tabla 5**). En una cohorte de pacientes con CCR de aparición temprana con un patrón de herencia recesivo se detectó una duplicación del extremo 5' del gen *PTPRJ,* la cual causaba una disminución de su expresión debido a que su promotor quedaba hipermetilado (Venkatachalam 2010). Otro estudió realizó genotipado de SNPs en 41 pacientes con CCR familiar y encontró CNVs raras en los genes *CDH18, GREM1 y BCR*, así como en zonas portadoras de los microARNs mir-491 y mir-646 (Venkatachalam 2011). El papel del gen *GREM1* como gen de predisposición a CCR fue corroborado por Jaeger y col cuando estudió mediante análisis de ligamento combinado con arrays de genotipado una familia con fuerte agregación para CCR, encontrando una duplicación de 40 kb delante del gen *GREM1*, la cual aumentaba su expresión. Adicionalmente, la búsqueda de CNVs raras permitió la detección de una delección que incluía el gen *BMPR1A* en un paciente con CCR de aparición temprana, sugiriendo que este gen podría estar relacionado también con el fenotipo mencionado (Fernandez-Rozadilla 2013b). Otro gen de predisposición a CCR encontrado con estas técnicas ha sido BUB1, el cual contenía una deleción de 1,7 kb en un paciente con CCR de aparición temprana, y para el que su implicación fue validada mediante estudios funcionales (de Voer 2013). Otro estudio detectó que el gen *FOCAD* presentaba deleciones en pacientes con poliposis y CCR, pero no en controles (Weren 2015a). Muy recientemente se han buscado variantes de número de copia mediante arrays de genotipado en 45 pacientes con CCR familiar de tipo X y se han encontrado 35 CNVs raras, destacando una duplicación que engloba *GALNT11* y *KMT2C* (Villacis 2015).

**Tabla 5. Estudios de detección de CNV en CCR familiar.**

| Autores | Pacientes | Resultados | Estudios adicionales |
|---|---|---|---|
| Venkatachalam R y col Gastroenterology 2010 | CCR de aparición temprana | Duplicación delante del gen ***PTPRJ*** | Se observó metilación del promotor de *PTPRJ* y disminución de la expresión génica |
| Venkatachalam y col 2011 Int J Cancer | 41 pacientes con CCR familiar | CNVs raras en los genes ***CDH18, GREM1 y BCR***, así como en zonas portadoras de los microARNs **mir-491 y mir-646** | |
| Jaeger y col Nature Genetics 2012 | 1 familia con poliposis hereditaria mixta | duplicación de 40 kb delante del gen ***GREM1*** | Cribado en casos y controles permite detectar la misma duplicación en 1 caso adicional y 0 controles |
| Fernandez-Rozadilla y col Clin Genet 2013b | 32 pacientes con CCR de aparición temprana | Delección de 7,32 Mb incluyendo 27 genes, entre ellos ***BMPR1A*** | |
| De Voer y col Gastroenterology 2013 | 39 pacientes con CCR de aparición temprana | Deleción 1,7 Mb en ***BUB1*** | Células mutadas presentan inestabilidad cromosómica, estudios funcionales en células muestran que *BUB1* mantiene la correcta segregación cromosómica, función que se pierde al delecionarse. Además se detectan nuevas mutaciones en *BUB1* y *BUB3* en una cohorte de casos con CCR. Estos presentan también inestabilidad cromosómica. |
| Weren y col J Pathol. 2015a | 41 pacientes con CCR de aparición temprana | Deleción intragénica en el gen ***FOCAD*** | Se cribó el gen el casos y controles, encontrándose nuevas deleciones en 2 casos pero no es controles |
| Villacis y col Int J Cancer. 2016 | 45 pacientes con CCR familiar de tipo X | 35 CNVs raras destacando una duplicación que engloba ***GALNT11*** **y** ***KMT2C*** | |

## 5.4- Secuenciación dirigida (Sanger)

Esta aproximación consiste en encontrar mutaciones causantes de predisposición mediante la secuenciación por Sanger de una región del genoma previamente seleccionada en base a diferentes criterios, como que esté descrita en estudios de desequilibrio de ligamiento (Guda 2013), o que sea un hit de GWAS como *BMP4* (Lubbe 2010) (**Tabla 6**). Otra estrategia sería limitar la búsqueda de genes a aquellos relacionados funcionalmente con los causantes de los síndromes de CCR hereditario, como por ejemplo el cribado de los genes *NUDT1* y *OGG1*, que participan en la misma vía de reparación del ADN que *MUTYH* (Garre 2011). Otra aproximación sería seleccionar genes ya implicados en un fenotipo similar al que se está estudiando, como en el caso del gen *BRCA2*, implicado previamente en cáncer de mama y ovario (Garre 2014). La limitación de esta estrategia es que no permite desvelar nuevos mecanismos causales alejados del conocimiento ya disponible de la enfermedad y es muy sesgada. A modo de validación,

series adicionales de pacientes con CCR familiar han sido secuenciadas para candidatos previamente identificados por Sanger como *GALNT12* y *UNC5C* (Segui 2014, Küry 2014, Mur 2015) sin identificarse nuevas mutaciones candidatas de causar predisposición. En cambio se han encontrado nuevas mutaciones probablemente patogénicas en cohortes adicionales de CCR familiar en los genes *BMPR1A* (Greenman 2007) y *BRCA2* (Yurgelun 2015) reforzando su papel como candidatos de predisposición.

**Tabla 6. Estudios de secuenciación Sanger dirigidos a encontrar mutaciones de predisposición a CCR en genes candidatos**.

| Autor | Pacientes | Genes/ Regiones secuenciadas | Motivo de elección | Resultados |
|---|---|---|---|---|
| Lipton y col. J Med Genet. 2003 | 47 pacientes con poliposis adenomatosa no filiada | *APC2*, *GSK3-β*, conductina, *SMAD4*, y *BMPR1A* | Implicación previa de la vía de Wnt y TGF-B en el desarrollo de adenomas | Una variante probablemente deletérea en *BMPR1A* |
| Guda y col. PNAS 2009 | 272 pacientes con CCR y 192 controles | *GALNT12* | Localización en 9q21-33, cerca de la región previamente identificada en estudios de ligamiento | Siete variantes deletéreas que afectan la función proteica en los casos con CCR. En controles seis variantes que no afectan su función |
| Lubbe y col Human Mutation 2010 | 504 pacientes con CCR y 524 controles | *BMP4* | Gen previamente asociado a CCR en estudios de GWAS | Diversas variantes en casos y controles, destacando en un caso una variante probablemente deletérea que afectaría la interacción con *BMPR1A* |
| Garre y col. Clin Cancer Res. 2011 | 42 familias con CCR familiar de tipo X | *MUTYH*, *OGG1* y *NUDT1* | Genes de la vía de señalización BER, implicada en CCR | Se detectaron dos variantes raras (*OGG1* p.*R46Q, MUTYH p.*G382D) con segregación correcta con CCR. *OGG1* p.R46Q afecta al splicing |
| Coissieux y col Gastroenterology 2011 | 235 pacientes con CCR y 500 controles | *UNC5C* | La expresión de este gen está reducida en muchos tumores colorrectales, además ratones mutados presentan mayor incidencia de CCR | Se detectan cuatro variantes candidatas, que son cribadas en casos y controles adicionales, permaneciendo una sola de las variantes (p.Ala628Lys) asociada a CCR. Estudio funcionales muestran que la variante reduce la apoptosis |
| Garre y col Clin Genet. 2014 | 48 pacientes con CCR familiar de tipo X | *BRCA2* | Implicación previa del gen en predisposición a otras neoplasias | Tres variantes de cambio aminoacídico probablemente deletéreas y una variante frameshift, todas con segregación familiar correcta y LOH |

## 5.5- Secuenciación de nueva generación

### 5.5.1- Concepto

La secuenciación de nueva generación (NGS, *Next generation sequencing)* es un conjunto de nuevas tecnologías posteriores a la secuenciación por Sanger, las cuales se basan en la preparación de la muestra de ADN a secuenciar, la propia secuenciación y el análisis bioinformático de los datos obtenidos (Metzker 2010). Es un método que permite producir cantidades enormes de datos a un precio relativamente reducido, realizándose la secuenciación de miles de fragmentos del ADN genómico simultáneamente y de manera masiva. Debido al gran volumen de datos generado, el análisis es la mayor dificultad que entrañan estas tecnologías, pues después de la secuenciación la variante de interés será mostrada junto con miles de otras variantes presentes en el genoma de cada organismo. Así, corresponderá a la persona que diseñe el experimento crear estrategias para seleccionar de forma correcta las variantes causantes del fenotipo de interés. El funcionamiento de esta tecnología puede variar en algunos pasos según el tipo de aparato secuenciador que se utilice. Primeramente, la preparación de la muestra requiere la fragmentación del ADN genómico y la unión de fragmentos de ADN de secuencia conocida llamados "adaptadores" en los extremos (**Figura 19**).

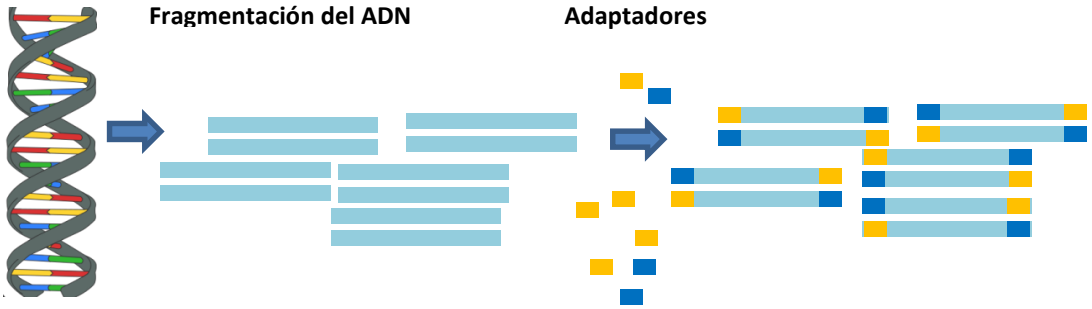



**Figura 19. Generación de librerías a partir del ADN muestral.** El ADN es fragmentado y a cada fragmento se le unen adaptadores de secuencia conocida.

El siguiente paso para la mayoría de secuenciadores consiste en la amplificación por PCR de cada fragmento de ADN genómico, que permitirá detectar con mayor intensidad la señal fluorescente producida durante la secuenciación. Para amplificar el ADN existen dos métodos: la amplificación en fase sólida (usada por los secuenciadores de Illumina) y la PCR en emulsión (usada por las tecnologías de secuenciación 454 de Roche). La primera consiste en hibridar los fragmentos con sus adaptadores a otros adaptadores unidos a una superficie sólida, lo que

provocará la retención de los fragmentos a esta superficie. Mediante primers complementarios a los adaptadores se procede entonces a amplificar los fragmentos generando miles de clústeres de amplificación (**Figura 20 A**).

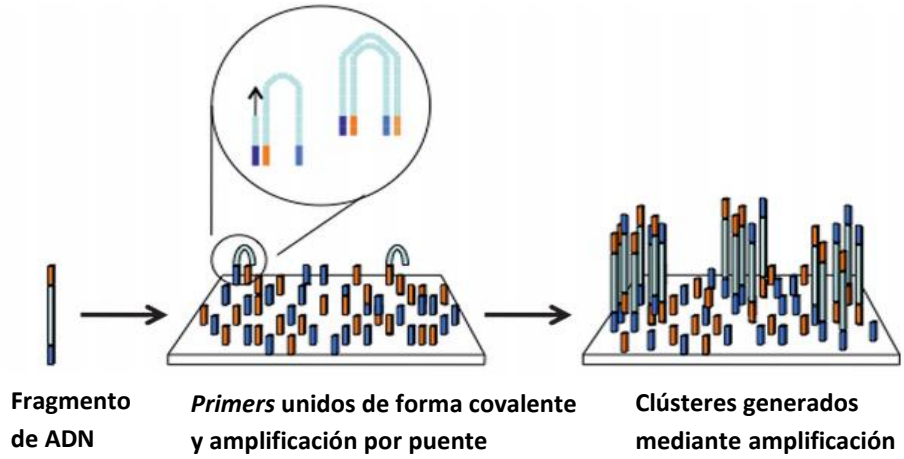



**Figura 20 A. Amplificación en fase sólida usada por los secuenciadores Illumina.** Los primers *forward* y *reverse* son unidos covalentemente y a una gran densidad a una superficie. Los fragmentos de ADN muestral con los adaptadores unidos son hibridados a los primers, y amplificados in situ mediante la unión del extremo libre del ADN al primer complementario de la superficie. Los múltiples ciclos de desnaturalización y extensión resultan en la amplificación localizada en clústeres a lo largo de la superficie. Figura adaptada de Casey (2013).

En el caso de la PCR en emulsión, cada fragmento de ADN de cadena simple es capturado en una microgota con los reactivos necesarios para la amplificación por PCR, generando en cada gota miles de copias de un fragmento (**Figura 20 B**).

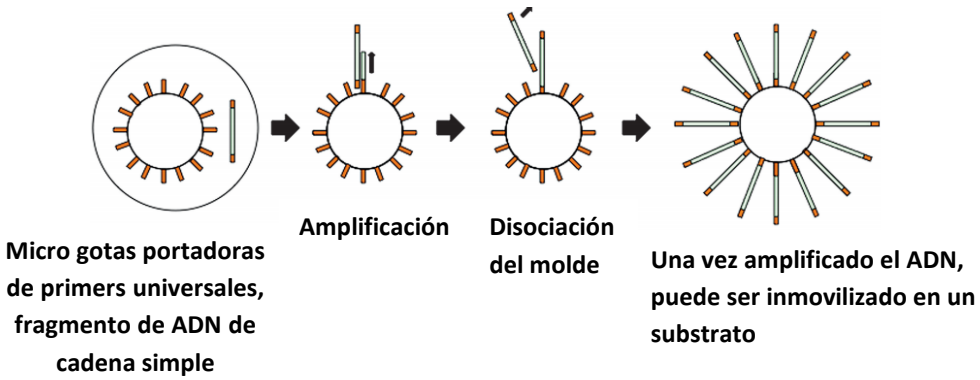



**Figura 20 B. PCR de emulsión utilizada en secuenciadores como 454 y Life Technologies**. En la PCR de emulsión, los fragmentos con los adaptadores son separados en cadenas de ADN sencilla y capturadas en micro gotas en condiciones que favorezcan la captura de una molécula de ADN en cada micro gota. En estas micro gotas están todos los reactivos necesarios para la amplificación por PCR, incluyendo primers universales complementarios a los adaptadores de los fragmentos del ADN muestral. A continuación, las moléculas de ADN ya amplificadas y separadas espacialmente se unen físicamente a un soporte sólido para la secuenciación. Figura adaptada de Casey (2013).

Posteriormente a la amplificación por PCR de cada fragmento de ADN genómico, se realiza la secuenciación. Ésta se produce mediante nucleótidos marcados de manera reversible con un terminador (el cual impide que continúe la reacción de polimerización del ADN) y un fluoróforo, que es diferente para los distintos nucleótidos (adenina, guanina, timina o citosina) (**Figura 21**).

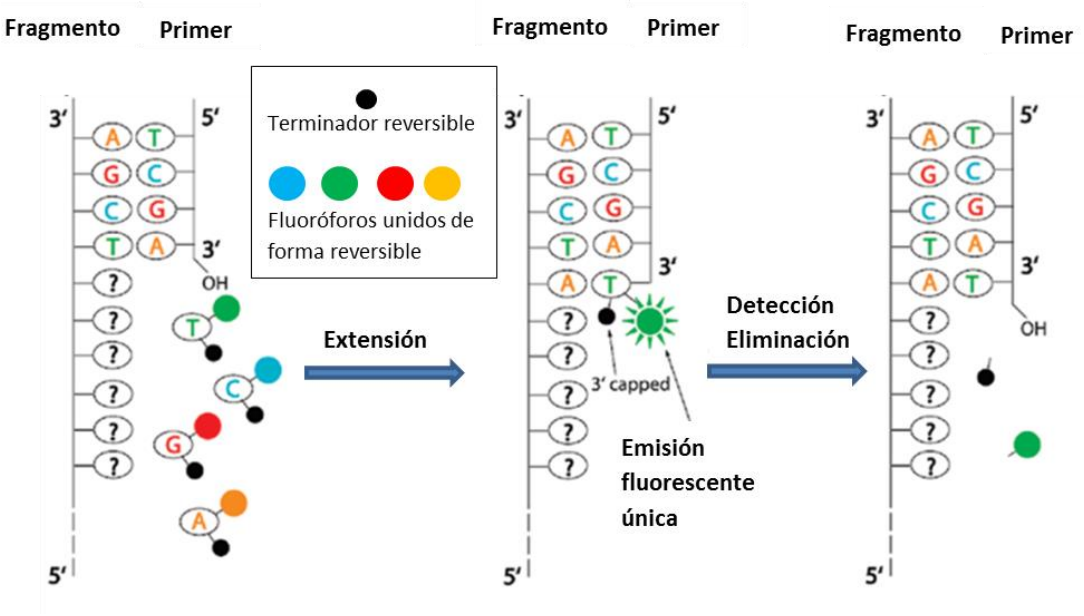



**Figura 14. Secuenciación por síntesis.** Se añaden los diferentes nucleótidos (T,C,G,A) marcados con diferentes fluoróforos cuya señal es captada por un sistema de imagen paralelamente para cada fragmento. Una molécula terminadora impide la adicción del siguiente nucleótido hasta que la señal fluorescente es captada. A continuación se elimina el fluoróforo y el terminador, y se añaden más nucleótidos marcados, repitiendo el ciclo de extensión, detección de la señal y eliminación del fluoroforo y el terminador hasta que se secuencia todo el fragmento. Imagen adaptada de Guo (2009).

El nucleótido marcado que sea complementario a la primera base del clúster se unirá impidiendo que se añadan más nucleótidos a continuación y emitirá una señal fluorescente. Esta señal será detectada mediante un sistema de captura de luz, que identificará qué nucleótido se ha unido en función del fluoróforo. Después se eliminará el fluoróforo terminador del primer nucleótido y se unirá el siguiente nucleótido marcado que se identificará por su señal fluorescente, así hasta que se haya secuenciado todo el clúster. Este proceso ocurre simultáneamente en todos los clústeres de ADN, y se obtendrá la secuencia de los distintos fragmentos amplificados. Finalmente, las secuencias de los fragmentos serán ordenadas mediante un programa informático utilizando una secuencia de un genoma de referencia y los nucleótidos que difieran de esta secuencia serán reportados como variantes. Otro programa realizará la anotación de las variantes, que consiste en reportar información

sobre éstas, como su frecuencia en bases de datos de variación genética o el efecto de la variante en la secuencia de aminoácidos (sinónima, cambio aminoacídico, splicing, pérdida de la pauta de lectura). También se anotan predicciones bioinformáticas de diferentes herramientas in silico como SIFT o Polyphen (Kumar 2009, Adzhubei 2013) que serán útiles para saber si una variante de cambio de aminoácido (en inglés *missense*) tiene mayor o menor probabilidad de afectar la función proteica. En base a la información disponible de las variantes y otros datos como el patrón de herencia esperado, o si se han secuenciado las variantes de otros individuos con la misma condición, se procederá a elegir la variante o variantes candidatas de producir un fenotipo determinado.

### 5.5.2- Secuenciación del genoma completo y secuenciación del exoma completo

Con el fin de detectar variación genética patogénica, es posible secuenciar el genoma completo (*whole-genome sequencing*, WGS) o capturar y secuenciar únicamente regiones específicas de interés (secuenciación dirigida). Por otro lado, la secuenciación del exoma se centra en el genoma codificante (1% del genoma), regiones intrónicas flanqueantes y algunos ARN no codificantes (Ng 2009). Esta es una estrategia más económica que la secuenciación del genoma completo y es útil para detectar variantes de alta penetrancia, ya que en las enfermedades mendelianas aproximadamente el 85% de las mutaciones causantes están situadas en regiones codificantes (Botstein 2003). Otra ventaja de centrarse en el exoma aparte del abaratamiento en los costes es que la cobertura media de las regiones secuenciadas es mayor. La cobertura de una región del genoma es el número de fragmentos secuenciados superpuestos en esa región, y su aumento comporta una mayor fiabilidad de las variantes detectadas. Por último, el secuenciar únicamente el genoma codificante conlleva una mayor simplicidad del análisis bioinformático. Sin embargo, el no poder detectar variantes fuera de los exones y los intrones colindantes hace que algunas variantes patogénicas situadas en *enhancers* distantes u otros elementos reguladores de la expresión génica sean pasadas por alto. Además, la inferencia de variantes de número de copia es más difícil en el caso de la secuenciación del exoma comparado con la secuenciación del genoma completo (Zhao 2013). Aparte de la secuenciación del exoma, otra aplicación de la NGS dirigida es la utilización de paneles de genes ya establecidos como causantes de una enfermedad para el diagnóstico clínico, como por ejemplo los genes reparadores causantes del síndrome de Lynch (Pritchard 2012). También pueden usarse paneles constituidos por múltiples genes candidatos de predisposición a una enfermedad con el fin de ampliar el conocimiento de estos genes mediante el cribado en múltiples casos y el análisis de las mutaciones detectadas. Esta estrategia para caracterizar genes de predisposición sería igual al cribado mutacional por

Sanger pero permitiría seleccionar un número mucho mayor de genes debido al menor coste de secuenciación. Recientemente, se publicó un cribado de 1.260 individuos afectos con CCR de tipo X mediante un panel de genes implicados en predisposición a distintos tipos de cáncer, identificándose los genes *BRCA1* y *BRCA2* como portadores de mutaciones en 15 casos. Además también se encontraron mutaciones en *ATM*, *BARD1*, *BRIP1*, *CHEK2*, *NBN*, *PALB2*, *RADS1C*, así como en *APC* y *STK11* (Yurgelun 2015).

### 5.5.3-Secuenciación del exoma completo para detectar nuevos genes de predisposición al cáncer colorrectal familiar

#### 5.5.3.1- Selección de familias

La elección de los individuos que serán secuenciados con el fin de encontrar variantes de predisposición a una enfermedad es un paso crítico para el éxito del estudio. Para las enfermedades con heterogeneidad genética como el cáncer, se pueden usar distintas estrategias, incluyendo entre otras la selección de familias con una fuerte agregación para la enfermedad o la secuenciación de casos esporádicos con aparición temprana de la patología, siendo ambas posibilidades sugestivas de predisposición genética. Si nos centramos en familias con varios miembros afectados, la secuenciación del mayor número de afectos posibles nos permitirá elegir las variantes compartidas entre estos, reduciendo el número de falsos positivos. No obstante, hay que tener en cuenta que a partir de una edad avanzada es muy posible la presencia de fenocopias, casos de cáncer esporádico debidos al envejecimiento y a factores ambientales compartidos en la familia. Para minimizar el riesgo de secuenciar fenocopias, es recomendable elegir individuos con una edad de aparición del cáncer menor a la que se esperaría en casos esporádicos. En caso de que una enfermedad presente penetrancia completa y se disponga de muestra de familiares no afectos con una edad superior a la del desarrollo de la patología, es posible secuenciar éstos no afectos para descartar las variantes que tengan en común con sus familiares afectos. En el caso del cáncer esta estrategia no es recomendable ya que para la mayoría de genes de predisposición conocidos no existe penetrancia completa. Por otro lado, si se secuencian casos esporádicos de aparición temprana, se seleccionarán genes que tengan variantes que afecten la secuencia proteica en diferentes individuos de las distintas familias.

#### 5.5.3.2- Filtrados

Típicamente se detectan de 20.000 a 50.000 variantes por exoma (Gilissen 2012). Para distinguir la variante patogénica entre los miles de falsos positivos es necesario establecer

ciertos criterios de selección. Para eliminar errores de secuenciación será necesario excluir las variantes que no tengan un mínimo de cobertura, es decir que la posición del genoma donde se encuentre la variante haya sido secuenciada un mínimo de veces, existiendo por tanto un número de fragmentos genómicos superpuestos en esa posición (típicamente 5x-10x). Además si buscamos mutaciones germinales, cierto porcentaje del total de estos fragmentos genómicos debe ser portador de la variante (típicamente más del 20% de los fragmentos para variantes heterocigotas y más del 80% para las homocigotas) (Gilissen 2012) (**Figura 22**).

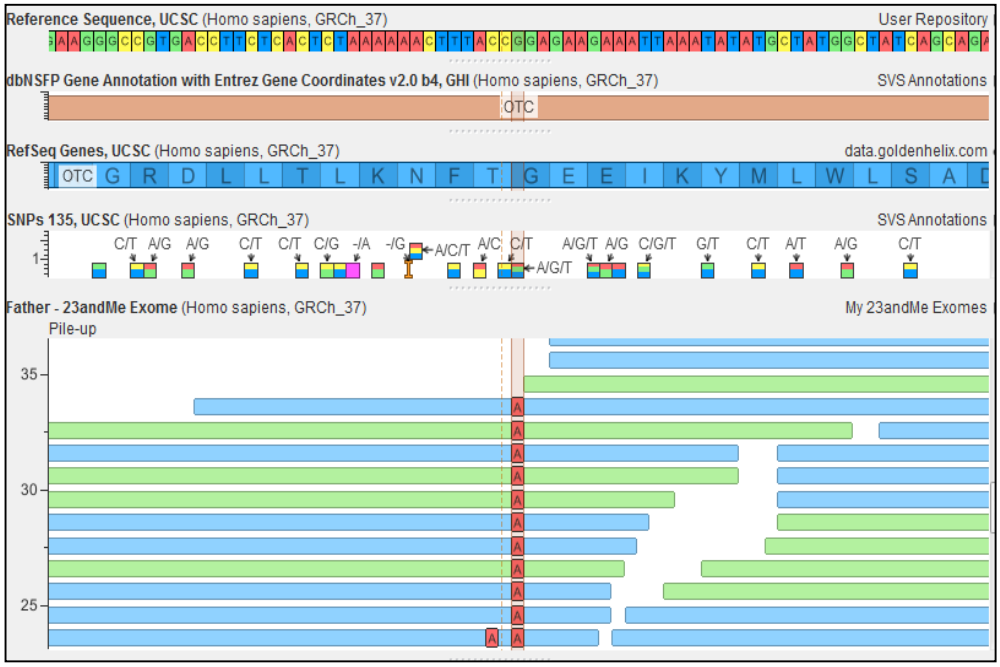



**Figura 22. Detección de variantes en fragmentos de ADN secuenciados y ordenados con la secuencia del genoma humano de referencia GRCh_37.** Las posiciones donde el nucleótido secuenciado difiere con el nucleótido de referencia son marcadas como variantes. La cobertura para una posición nucleotídica o una región determinada es el número de fragmentos diferentes que se solapan en esa posición o región. En el caso del ADN germinal, las variantes detectadas en menos del 20% de los fragmentos son susceptibles de ser falsos positivos. Imagen extraída de http://blogs.nature.com/news/2012/12/expert-tours-his-own-exome-and-finds-mainly-false-alarms.html.

Otros filtros que se pueden hacer dependen del modo de herencia, la heterogeneidad genética, la penetrancia y la frecuencia de la enfermedad en la población general. Si se presenta un patrón de herencia mendeliano, se podrán seleccionar las variantes heterocigotas en el caso de un patrón de herencia dominante y las homocigotas o en heterocigosis compuesta en el caso de un patrón recesivo (**Figura 23**). Sin embargo, las variantes en las regiones no pseudoautosómicas en el cromosoma X en el caso de herencia dominante también deberán tenerse en cuenta. En los hombres estas variantes serán anotadas como homocigotas

pero deberá evitarse su filtrado. Como se ha mencionado anteriormente, la disponibilidad de varios afectos secuenciados dentro de una familia permite seleccionar las variantes compartidas entre éstos. Si es una familia lo suficientemente grande como para realizar análisis de desequilibrio de ligamiento, seleccionar las variantes situadas dentro de esas regiones en desequilibrio de ligamento con la enfermedad permite acotar aún más el número de candidatas. En el caso de secuenciarse distintos individuos con aparición temprana de la enfermedad se seleccionarán los genes que tengan variantes en más de un individuo como candidatos. Si existe penetrancia completa se pueden secuenciar individuos controles y excluir las variantes que estén compartidas con los individuos afectos.

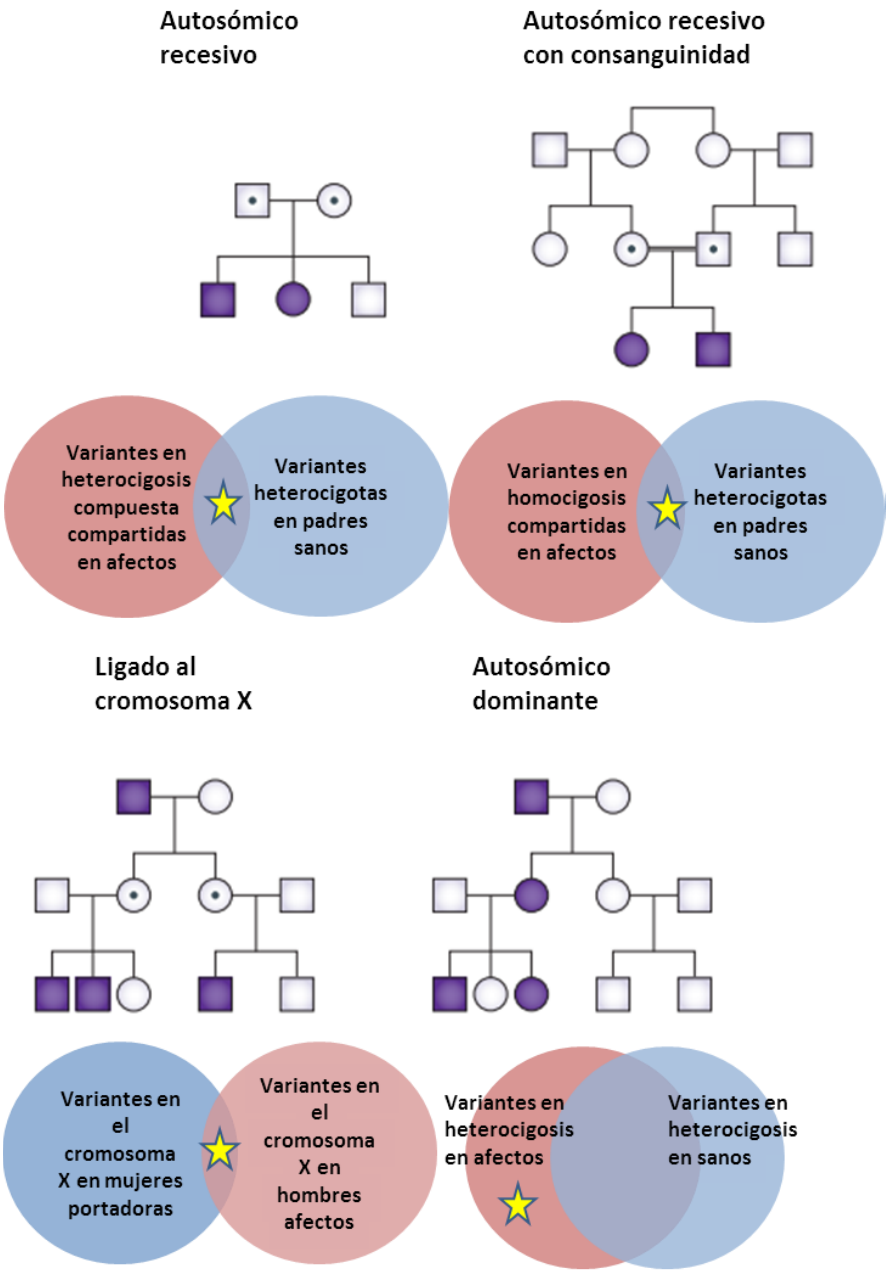


**Figura 23. Diferentes *pedigrees* representativos de los diferentes tipos de mutación patogénica y la transmisión de su herencia**. Los afectos están indicados con un relleno de color morado y los portadores sanos, con un círculo negro en el centro. Las estrellas resaltan el conjunto de variantes entre las cuales se espera que esté la mutación causante de la enfermedad. Figura adaptada de Bovcott (2013).

La frecuencia de la enfermedad en la población general también es un criterio importante a considerar. Para una enfermedad rara y genéticamente heterogénea como es el CCR hereditario/familiar, donde además ya existen genes mayoritarios identificados, la frecuencia de las variantes en la población general se espera que sea baja, especialmente cuando existe herencia dominante. Otro filtrado típico es la selección de las variantes que afecten la secuencia de proteínas, como las *missense* (substitución aminoacídica), *frameshift* (cambio de la pauta de lectura), variantes de *splicing* canónico (+1,-1,+2,-2 nucleótidos de distancia de los exones) o de terminación de la proteína, excluyendo las variantes sinónimas, o las situadas en los UTR o en zonas intrónicas más alejadas del exón.

#### 5.5.3.3- Priorización

El filtrado concluye con la obtención de un conjunto de variantes para cada individuo o familia, que posiblemente sea aún demasiado extenso para distinguir cual es la causante del fenotipo a estudiar. Una aproximación lógica para reducir el número de variantes es priorizar aquellas situadas en genes previamente implicados en la enfermedad. Además, debido a que las proteínas implicadas en una enfermedad determinada tienden a interactuar entre ellas físicamente, otra estrategia de priorización es seleccionar variantes en genes cuyos productos interaccionan con los de genes implicados previamente en la enfermedad (Lim 2006). Adicionalmente, el conocimiento de las vías de señalización implicadas en la enfermedad puede ser útil para priorizar los genes cuya función biológica esté relacionada con esas vías. Por otro lado, las variantes missense pueden priorizarse en función de si las predicciones bioinformáticas anotadas (como Polyphen, SIFT etc.) indican una alta probabilidad de alteración de la función proteica. Finalmente, las variantes missense también pueden priorizarse según si afectan dominios importantes (como por ejemplo el dominio de unión al DNA en una polimerasa) o residuos con modicaciones translacionales (glicosilaciones, fosforilaciones, etc). Una vez seleccionadas la variante o variantes de mayor interés, estudios posteriores deberán confirmar su papel en el desarrollo de la enfermedad.

#### 5.5.3.4- Validación por Sanger

Pese a la gran precisión de las tecnologías NGS existe un mínimo de error asociado, y debido a la gran cantidad de datos generados, la posibilidad de detectar falsos positivos es una realidad que debe subsanarse mediante la validación de las variantes candidatas mediante una tecnología alternativa, como por ejemplo la secuenciación por Sanger (Ulahannan 2013).

#### 5.5.3.5- Segregación

El análisis de segregación de las variantes candidatas en los familiares no secuenciados permite comprobar si la variante candidata en aquella familia segrega correctamente con la enfermedad. Así, los miembros afectos deben ser portadores y los no afectos con una edad suficientemente avanzada para haber expresado la enfermedad no deben ser portadores. No obstante, se debe ser cuidadoso debido a que en ocasiones puede haber penetrancia incompleta, lo que provoca que individuos sanos de edad avanzada puedan ser portadores y nunca expresar la enfermedad, pese a tener el alelo mutado que confiera una gran predisposición potencial a padecerla.

#### 5.5.3.6- Otros estudios: Pérdida de heterocigosidad, estudios de *splicing*, estudios de expresión, estudio del perfil mutacional somático

En el caso del cáncer hereditario, la pérdida de heterocigosidad (LOH) del alelo normal en el ADN tumoral puede ser sugestiva de que el gen candidato es un supresor tumoral, pues esta pérdida de heterocigosidad sería el segundo hit necesario para inactivar el gen, iniciando la tumorigenesis. No obstante en genes supresores de tumores el segundo hit inactivador puede producirse también por mutación puntual y por metilación. Además algunos supresores de tumores presentan haploinsuficiencia y no es necesaria la pérdida del alelo normal para desencadenar la tumorigenesis, por tanto la presencia de LOH no debe usarse como criterio de validación sino como una información adicional.

Los estudios de splicing servirán para validar el efecto las variantes situadas en los sitios donadores y aceptores de exones. Cuando la variante candidata de predisposición está situada en zonas intrónicas que se espera que afecten el correcto ensamblaje de los intrones se pueden realizar estudios del splicing. El splicing aberrante en las personas portadoras de la variante candidata se puede comprobar mediante la extracción de ARN y secuenciación dirigida de los exones flanqueantes en el ADN, para detectar alteraciones. Por último, los estudios de expresión permiten estudiar si efectivamente hay una reducción en la cantidad de transcrito o de proteína en los portadores respecto a los no portadores. En el caso de haberse

detectado LOH, se puede ver si el tumor tiene una expresión reducida respecto al tejido sano. Para cuantificar la expresión génica a nivel de ARN se puede realizar PCR cuantitativa del ADNc, y para la pérdida de expresión proteica se puede realizar inmunohistoquímica o *Western blot*.

#### 5.5.3.7- Cribado en casos y controles

Los estudios caso-control son útiles para ayudar a definir el impacto de la variante en el riesgo a causar la enfermedad. La identificación de la misma variante (o una variante con un efecto equivalente en el mismo gen) en casos adicionales pero no en controles apoya la existencia de una causalidad gen-enfermedad. En el caso de genes con una penetrancia incompleta, encontraremos un enriquecimiento de variantes patogénicas en casos respecto a controles, pero siempre habrá una pequeña fracción de individuos sanos mutados.

#### 5.5.3.8- Estudios funcionales

El hecho de identificar un cambio genético poco frecuente que afecte la expresión o la función proteica no es una evidencia suficientemente sólida como para relacionarlo directamente con la enfermedad (MacArthur 2014). Realizar estudios funcionales es el paso definitivo que demostrará que la variante candidata es efectivamente la mutación causante de la predisposición genética a la enfermedad estudiada. Un ejemplo es realizar ensayos en cultivos celulares o en modelos animales donde la variante se introduce por mutagénesis dirigida, y a continuación se estudia si el organismo modelo muestra un fenotipo similar a la enfermedad estudiada. Además de corroborar el papel patogénico de la mutación, también se realizan ensayos funcionales para identificar cual es el mecanismo molecular por el que se desencadena la patología. Se pueden estudiar las interacciones entre el producto del gen mutado y otras proteínas pertenecientes a una vía de señalización determinada mediante inmunoprecipitación de proteínas seguida de *Western blot*. En el caso de factores de transcripción se puede estudiar si la unión con el DNA está afectada por la variante mediante inmunoprecipitación de la cromatina (*Chromatin Immunoprecipitation* o ChIP) en el caso de disponer de un cultivo celular portador de la variante a estudiar.

### 6- Estudios previos de selección de genes candidatos de predisposición al CCR

Con el fin de identificar variantes de alta penetrancia que predispongan a CCR familiar, diferentes estudios de secuenciación del exoma se han realizado en diferentes cohortes de pacientes afectados, siguiendo distintas estrategias de filtrado y priorización, que han permitido en cada caso generar una lista de variantes en genes candidatos que potencialmente

causarán la enfermedad. Esta estrategia es útil para focalizar los esfuerzos de la investigación en una dirección como puede ser la detección de variantes en el mismo gen en cohortes independientes de individuos afectos. Además, la detección de diferentes genes candidatos de una misma vía de señalización puede poner sobre aviso a los investigadores que esa vía puede ser importante para el desarrollo de la enfermedad. Varios estudios se han realizado en base a esta estrategia de identificación de genes, y así se han señalado varios candidatos de predisposición a CCR familiar (**Tabla 7**). A continuación se detallarán los estudios realizados y los genes candidatos identificados hasta el momento.

**Tabla 7. Estudios de secuenciación de exomas completos en el campo del CCR familiar. Cada estudio ha permitido la identificación de distintos genes candidatos de predisposición.**

| Autores | cohorte | Criterios de selección | Genes identificados | Estudios posteriores |
|---|---|---|---|---|
| Smith y col 2013 | 50 pacientes con CCR esporádico incluyendo 18 casos con aparición temprana | Lista de 1138 genes con función implicada en cáncer + inactivación bialélica en ADN tumoral | *FANCM, LAMB4, PTCHD3, LAMC3* y *TREX2* | Cribado mutacional de *FANCM* en casos y controles |
| DeRycke y col 2013 | 40 casos de 16 familias con agregación para CCR | Variantes compartidas dentro de cada familia | *CENPE* y *KIF23* | Segregación en familiares adicionales |
| Gylfe y col 2013 | 96 casos independientes con CCR familiar | Variantes de pérdida de función compartidas en dos o más individuos | *UACA, SFXN4, TWSG1, PSPH, NUDT7, ZNF490, PRSS37, CCDC18, PRADC1, MRPL3*, y *AKR1C4.* | LOH y cribado mutacional en casos y controles |
| Tanskanen y col 2015 | 22 pacientes con CCR de aparición temprana | Variantes en genes mutados en una cohorte de 95 casos de CCR con agregación familiar o variantes con inactivación bialélica | *ADAMTS4, CYTL1, SYNE1, MCTP2, ARHGAP12, ATM, DONSON* y *ROS1* | |
| Goldberg y col 2015 | 2 pacientes con fallo ovárico, CCR y poliposis de 1 familia consanguinia | Variantes deletéreas en homocigosis | *MCM9* | Cribado de 170 casos adicionales |
| Spier y col 2016 | 7 pacientes con poliposis para los que además del ADN germinal se secuencia el ADN de tejido polipósico. | Variantes en genes con inactivación bialélica o mutados en una cohorte de 191 casos adicionales con poliposis | *DSC2, PIEZO1* y *ZSWIM7* | |
| De Voer y col 2016 | 55 pacientes con CCR de aparición temprana | Genes con variantes en dos o más individuos | *PTPN12* y *LRP6* | Cribado mutacional en casos y ensayo in vitro de la activación de Wnt en las variantes de *LRP6* |
| Thutkawkorapin y col 2016 | 3 pacientes de 1 familia con fuerte agregación para CCR y cáncer de estómago | Variantes con correcta segregación en 2 familiares adicionales con CCR | *DZIP1L, PCOLCE2, IGSF10, SUCNR1, OR13C8, EPB41L4B, SEC16A, NOTCH1, TAS2R7, SF3A1, GAL3ST1 y TRIOBP* | Cribado de los genes candidatos en 98 casos con CCR |

## 7. Nuevos genes de predisposición germinal al cáncer colorrectal

Algunos genes han sido seleccionados como candidatos de predisposición a CCR a partir de estudios de secuenciación de nueva generación o mediante detección de CNVs y posteriormente se han confirmado mediante estudios funcionales. Estos estudios han permitido dilucidar el mecanismo por el cual una mutación determinada en un gen causa la pérdida de funcionalidad de la proteína codificada, afectando a un mecanismo celular implicado en la tumorigenesis. Este último paso es necesario para concluir que un gen está implicado en predisposición a una enfermedad como puede ser el CCR familiar. Los genes identificados hasta el momento se detallan a continuación.

### 7.1- *GREM1*

En una familia con poliposis hereditaria mixta se realizó un estudio desequilibrio de ligamiento combinado con análisis de CNV con el fin de encontrar la causa de predisposición genética. Así, se detectó una duplicación de 40 kb situada *upstream* del gen *GREM1* (Jaeger 2012). Usando RT-qPCR e inmunohistoquímica se demostró que los portadores presentaban una expresión aumentada de *GREM1* y una distribución ectópica en el epitelio colónico normal. Se demostró también que mediante inmunoprecipitación de la cromatina que la duplicación contenía regiones de unión a factores de transcripción activadores. Además se hicieron ensayos con una línea celular transfectada con un plásmido portador de la duplicación y se midió la activación del gen *GREM1* mediante ensayos con luciferasa, demostrando que la introducción del fragmento duplicado en esas células era capaz de activar la transcripción más de cuatro veces (Jaeger 2012). Otro estudio detectó en una familia con poliposis atenuada con herencia dominante una duplicación de 16 kb que contenía sitios de unión de factores de transcripción y mediante estudios funcionales se demostró que predisponía a la enfermedad con herencia dominante (Rohlin 2016). Previamente al descubrimiento de estas variantes de alta penetrancia, se había reportado mediante un estudio de GWAS una variante de baja penetrancia en *GREM1*, que confería un riesgo modesto de padecer CCR (Tomlinson 2008). *GREM1* actúa como oncogén inhibiendo la vía de señalización no canónica de BMP (*bone morphogenetic protein*), lo que provoca una inhibición de la diferenciación celular y la activación del fenotipo de célula madre pluripotente por todo el epitelio colónico, desencadenando el desarrollo de pólipos. (**Figura 24**). La vía de BMP se ha visto previamente implicada en la poliposis juvenil, que está causada por mutaciones en *BMPR1A* y *SMAD4*.

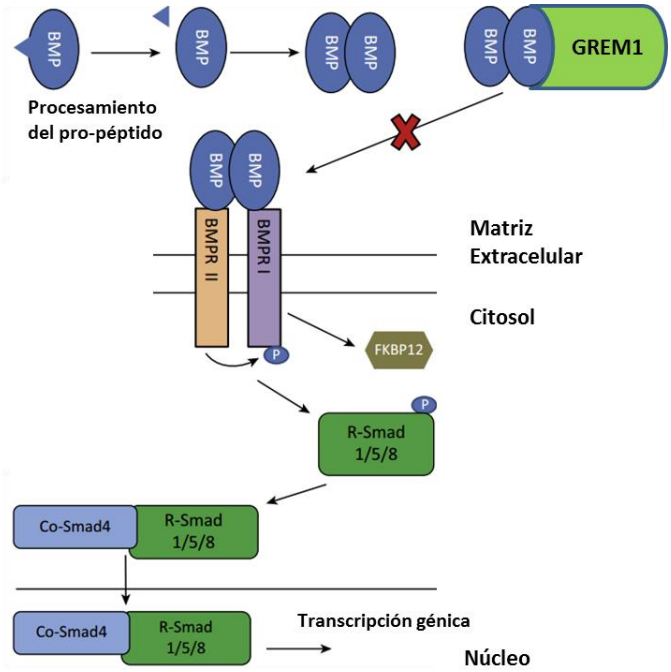



**Figura 21. Cascada de señalización de BMP y su regulación por GREM1**. Los pro-pétidos son procesados para generar dímeros maduros que se unen a dos copias de los receptores de BMP de tipo I y II. Esta unión conlleva la fosforilación de los receptores de tipo I, que provoca su activación y la fosforilación por este receptor de las proteinas Smad1/5/8. Estas dimerizan con Smad4 y se acumulan en el núcleo donde median cambios en la expresión génica. La regulación de esta vía ocurre mediante antagonistas extracelulares como GREM1 y NOG y de forma intracelular mediante otros antagonistas. Figura adaptada de Brazil (2015).

## 7.2- *BUB1/BUB3*

Con el objetivo de detectar nuevos genes de predisposición al CCR, se hicieron estudios de detección de CNV a nivel genómico en 39 pacientes con CCR de aparición temprana (de Voer 2013). En un paciente se detectó una microdeleción de aproximadamente 1,75 Mb en el gen *BUB1*, que resultaba en una pérdida de expresión de la proteína codificada. El análisis citogénetico en fibroblastos y linfocitos del portador reveló un gran número de alteraciones cromosómicas. Mediante la generación de una línea celular mutante *BUB1*+/-se detectó un número aberrante de ganancias y pérdidas cromosómicas y se confirmó así que este gen es necesario para el mantenimiento de la estabilidad cromosómica y que un solo alelo no es suficiente para mantener su función (haploinsuficiencia). Posteriormente se secuenció el exoma de 33 pacientes adicionales con CCR de aparición temprana y se detectaron dos variantes de pérdida de función en *BUB1* y una variante de cambio de aminoácido en el gen *BUB3*, que está relacionado funcionalmente con el anterior. Finalmente, un cribado de casos y controles en estos genes permitió la detección de dos variantes adicionales en *BUB3* en casos pero no en controles. Estos casos también presentaban alteraciones cromosómicas en el ADN de sus linfocitos. Así quedó demostrado que mutaciones germinales en estos dos genes

predisponen a CCR de aparición temprana con herencia autosómica dominante. Es importante señalar que además de CCR algunos de los pacientes portadores presentaban otros tumores primarios, como ovario, mama, pulmón, linfoma, riñón, esófago y estómago, ampliándose el espectro fenotípico de predisposición para estos genes. En cuanto a su función biológica, *BUB1* y *BUB3* participan en el punto de control del ensamblaje del huso mitótico además de ser necesarias para el correcto alineamiento de los cromosomas, por tanto son necesarias para la correcta segregación cromosómica durante la mitosis (Johnson 2005, Zhang 2014). El mecanismo por el cual se desarrolla el cáncer es por la generación de CIN, que deriva de una incorrecta segregación derivada de la pérdida de estas proteínas.

### 7.3- *POLE/POLD1*

#### 7.3.1- Identificación

Los genes *POLE* y *POLD1* fueron establecidos como causantes de predisposición a CCR mediante la secuenciación del genoma completo en familias con agregación para poliposis múltiple y CCR de aparición temprana (Palles 2013). Los datos obtenidos de la secuenciación fueron contrastados con los resultados del análisis de desequilibrio de ligamiento que estaba disponible para algunas familias, lo que limitó en éstas la selección de variantes candidatas a las regiones en desequilibrio de ligamiento. Así, se encontró la variante p.L424V en el gen *POLE* en una familia y la variante p.S478N en el gen *POLD1* en dos familias independientes. Un cribado posterior de estas variantes en 3.085 casos con CCR y 10.755 controles permitió la detección de la variante en el gen *POLE* en 12 casos adicionales y la variante en el gen *POLD1* en un caso adicional, estando ambas variantes ausentes en controles. A continuación se procedió a analizar el efecto de la mutación p.S478N en la capacidad correctora de errores de las polimerasas mediante mutagénesis dirigida en *Schizosaccharomyces pombe*, mutando el residuo equivalente en los genes ortólogos de esta levadura y demostrando que estos cambios realmente alteraban la función proteica, ya que provocaban un fenotipo hipermutador en la levadura. Este fenotipo se caracteriza por una alta tasa de cambios en el genoma, lo que hace que una levadura con un fenotipo determinado, como la incapacidad de crecer en medios deplecionados de adenina, revierta ese fenotipo y obtenga la capacidad de formar colonias en ese medio. Para la mutación p.L424V ya existían estudios funcionales realizados por grupos independientes en levadura que respaldaban su patogenicidad (Murphy 2006). De esta manera estos dos genes fueron establecidos como causantes de predisposición a CCR hereditario.

### 7.3.2- Estudios posteriores

Con el fin de conocer mejor el espectro mutacional y las características clínicas asociadas a este síndrome, definido como síndrome asociado a la actividad reparadora de la polimerasa (Briggs 2013) se han realizado diferentes estudios. (**Figura 25**, **Tabla 8**). Las mutaciones descubiertas en *POLE* y *POLD1* por Palles y col han sido reportadas un total de 24 y 4 veces respectivamente, con una frecuencia que va del 0% al 1,5% en las diferentes cohortes estudiadas. Además, se han descubierto nuevas mutaciones con evidencia funcional de patogenicidad en *POLE* (p.Y458F, p.W347C, p.D368V) y en *POLD1* (p.L474P, p.D316H, p.D316G, p.R409W) (Valle 2014, Chubb 2015, Hansen 2015, Bellido 2015, Aoude 2015). En la mayoría de los casos ya existían estudios realizados en levadura, E.coli, el virus del herpes simplex o en bacteriófago T4, que demostraban que estas mutaciones afectaban la función de las polimerasas (Derbyshire 1991, Soengas 1992, Abdus Sattar 1996, Hwang 1997, Murphy 2006), a excepción del estudio de Aoude y col, que realizó el ensayo funcional posterior al descubrimiento de la mutación p.W347C en *POLE,* demostrando su patogenicidad. Otro descubrimiento ha sido la ampliación del espectro de neoplasias de este síndrome. Mutaciones en *POLE* predisponen a adenomas múltiples, CCR, endometrio, ovario, cerebro, páncreas, intestino delgado y melanoma cutáneo (Rholin 2014, Hansen 2015, Bellido 2015, Aoude 2015), mientras que mutaciones en *POLD1* predisponen a adenomas múltiples, CCR, endometrio y mama (Valle 2014, Chubb 2015, Bellido 2015). Además se ha encontrado que alteraciones genéticas en *POLE* pueden predisponer a tumores colorrectales con MSI debido a que facilitan la acumulación de mutaciones somáticas en los genes reparadores de Lynch (Elsayed 2015).

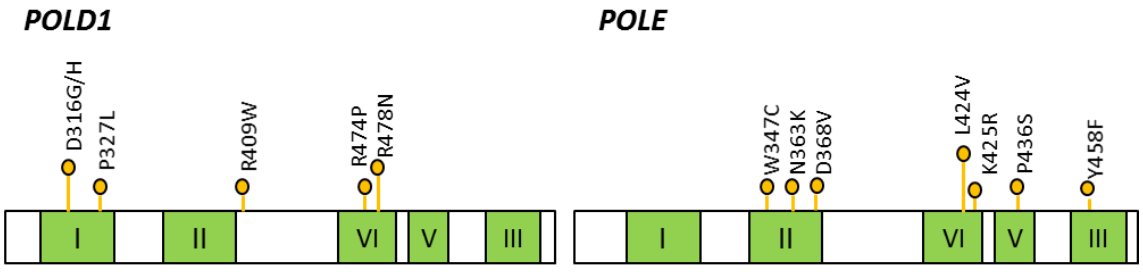



**Figura 22. Mutaciones germinales encontradas en el dominio exonucleasa de POLD1 y POLE.** Los motivos conservados en la evolución exo I-V del dominio están marcados en verde.

**Tabla 8. Estudios que han permitido identificar mutaciones en los genes *POLE* Y *POLD1* posteriores a su descubrimiento como genes de predisposición a CCR.**

| Autores | Método | Pacientes | Resultados | Conclusiones principales |
|---|---|---|---|---|
| Valle y col. 2014 | Sondas Taqman o secuenciación Sanger del fragmento que contiene las mutaciones | 858 pacientes con poliposis y/o CCR de aparición temprana | **1 caso con *POLE* p.L424V** de novo y 1 caso con ***POLD1* p.L474P** (nueva mutación, con estudios funcionales ya hechos en levadura) | Las mutaciones en *POLE* pueden tener herencia de novo. Se amplía el espectro mutacional de *POLD1* |
| Rholin y col 2014 | WES | 1 familia con agregación para CCR, adenomas y otras neoplasias | Variante nueva en ***POLE* p.N363K** potencialmente patogénica sin estudios funcionales | Sin estudios funcionales, solo se puede sugerir que algunas variantes en *POLE* pueden causar un fenotipo con tumores múltiples |
| Elsayed y col 2015 | Sondas Taqman | 1188 casos con CCR familiar o poliposis | **3 casos con *POLE* p.L424V** (1 caso con herencia de novo) 2 de estos casos con pérdida de MSH2/MSH6 y tumores MSI | Mutaciones germinales en *POLE* pueden causar alteraciones somáticas en los genes de Lynch, provocando tumores MSI |
| Spier y col 2015 | Secuenciación dirigida (NGS) de *POLE* y otros 7 genes de las polimerasas | 266 pacientes con poliposis y CCR de aparición temprana | **4 casos con *POLE* L424V**. Otros casos con variantes potencialmente patogénicas, incluyendo Pro436Ser | Sin estudios funcionales, solo se puede sugerir que variantes en otros genes de las polimerasas pueden causar predisposición a CCR de aparición temprana |
| Chubb y coll. 2015 | WES | 626 casos con CCR de aparición temprana | **2 casos con *POLE* L424V**, **1 caso con *POLD1* S478N**. 6 casos con nuevas variantes en *POLE* y *POLD1,* resaltando **p.D368V** en *POLE, existiendo evidencia invitro de su patogenicidad* | Se descubre una nueva mutación en *POLE* |
| Hansen y col. 2015 | WES | 1 familia con agregación para adenomas y diferentes neoplasias | **1 caso con variante nueva en *POLE* p.Y458F**, con ensayos funcionales existentes demostrando su patogenicidad | Se confirma que mutaciones en *POLE* predisponen a un fenotipo con múltiples tumores |
| Bellido y col 2015 | Secuenciación dirigida (NGS) de *POLE* y *POLD1* | 456 casos con CCR y 88 casos con poliposis | **1 caso con *POLE* L424V, 4 mutaciones en *POLD1*** con estudios funcionales probando su patogenicidad **(p.L474P,p.D316H, p.D316G, p.R409W)** y 2 variantes potencialmente patogénicas en *POLD1* | Se descubren nuevas mutaciones en *POLD1* |
| Aoude y col 2015 | WES y WGS (fase de descubrimiento)+ secuenciación dirigida (NGS) *POLE* (validación) | 34 familias con melanoma + 1243 casos con melanoma cutáneo, algunos con agregación familiar para otras neoplasias | **1 caso con mutación nueva en *POLE* W347C,** validada realizando estudios funcionales en levaduras. 10 casos con variants nuevas o raras en el dominio exonucleasa de *POLE*, incluyendo p.Q520R | Mutaciones en *POLE* predisponen a melanoma cutáneo |
| Rohlin y col 2016 | WES, estudio de ligamiento y CNV para una familia y secuenciación Sanger para 76 pacientes. | 1 familia con poliposis y CCR y 76 pacientes independientes con poliposis, síndrome de Gardner o CCR | 1 caso con una variante nueva potencialmente patogénica en ***POLE* p.K425R** sin estudios funcionales | Se encuentra una variante potencialmente patogénica en *POLE* |

### 7.3.3- Dominio exonucleasa y actividad correctora de errores durante la replicación del ADN

Como se ha mencionado anteriormente, estos genes pertenecen a la subunidad catalítica de las polimerasas épsilon y delta. La subunidad catalítica de estas proteínas incluye el dominio polimerasa y el dominio exonucleasa, comprendiendo este último los aminoácidos 267-471 en POLE y 304-562 en POLD1. El dominio exonucleasa tiene la función de corregir errores en el ADN generados durante su replicación y su mal funcionamiento provocaría la acumulación de mutaciones somáticas, desencadenando la tumorigenesis (Shevelev y Hübscher 2002). Hasta el momento se ha pensado que únicamente las mutaciones germinales dentro de este dominio eran susceptibles de causar predisposición a CCR y son las únicas para las cuales existen estudios funcionales respaldando su patogenicidad. No obstante se han encontrado familias afectadas con variantes fuera de este dominio, como *POLE* p.Q520R, que son sugestivas de causar la enfermedad, pero también sería necesario realizar estudios funcionales para dilucidar su efecto (Aoude 2015). Hasta ahora se había establecido que la función principal de POLE era la replicación de la cadena rápida del ADN mientras que POLD1 replicaba la cadena retardada (Johansson y Dixon 2013). Sin embargo, un estudio muy reciente concluyó que POLD1 participaba en la replicación de las dos cadenas mientras que POLE tiene un papel importante en la eliminación de errores introducidos por POLD1 durante la replicación de la cadena rápida (Johnson 2015) (**Figura 26**).

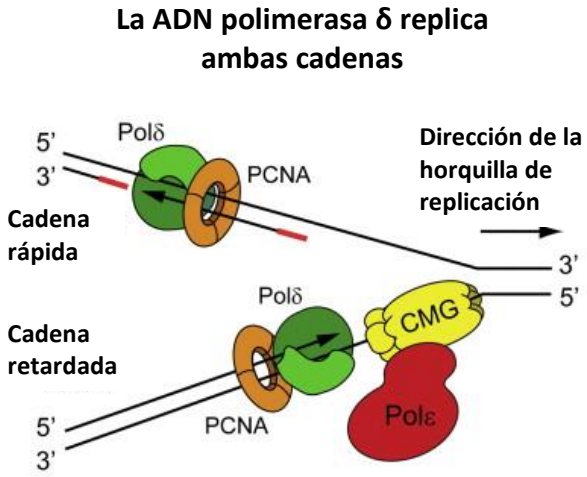



**Figura 26. Nuevo modelo potencial donde la ADN polimerasa δ replica las dos cadenas en condiciones normales y, cuando existe daño en el ADN en la cadena rápida, la polimerasa ε se coloca en su lugar.** La ADN polimerasa a se acopla con la primasa para sintetizar un primer ARN-ADN en la cadena retardada que es reconocido por RFC y PCNA para cambiar a la polimerasa replicativa. Imagen adaptada de Johnson (2015).

### 7.3.4- Efecto de las mutaciones en el ADN: perfil tumoral

El estudio del ADN tumoral en los pacientes con mutaciones en *POLE* y *POLD1* demostró que algunos tumores no tenían LOH, planteando la pregunta de si estos genes se comportaban como supresores de tumores clásicos. Varios tumores mostraban mutaciones en *APC, KRAS, BRAF* y *FBXW7.* Se observaba además un numero anormalmente alto de alteraciones genéticas somáticas (fenotipo conocido como hipermutador), con un porcentaje mayor de transversiones G>T/ C>A y C>T/ G>A, especialmente en el contexto TCT>TAT y TCG>TTG. (The Cancer Genome Atlas Network 2012, Palles 2013, Rayner 2016). Además, se ha reportado que el mal funcionamiento de estas polimerasas puede provocar la inactivación somática del gen *MSH2*, generando tumores con MSI (Elsayed 2015).

## 7.4- *RPS20*

Mediante un estudio de desequilibrio de ligamiento combinado con secuenciación del exoma completo en una familia con CCR familiar de tipo X se identificó la mutación p.V50Sfs*23 en *RPS20* (Nieminen 2014). Este gen codifica para la proteína S20, que forma parte de la subunidad ribosómica pequeña 40S. Mediante ensayos celulares se demostró que la inhibición de la expresión de *RPS20* provocaba un aumento de los ARNs precursores 21S y 18S asociado con un procesamiento tardío de estos. El mecanismo exacto de tumorigenesis por el cual este desfase en el procesamiento de los ARNs precursores de la subunidad pequeña del ribosoma provoca el desarrollo de tumores colorrectales es desconocido. La frecuencia en que este gen se encuentra mutado en familias con CCR familiar de tipo X también está por resolver aún. La familia RPS comprende 80 genes, 11 de los cuales se han implicado a un síndrome llamado Anemia de Blackfan Diamond, causante de aplasia de los glóbulos rojos y predisposición a distintos tipos de cáncer. A raíz del descubrimiento de Nieminen y col. otro grupo reportó que habían identificado dos familias afectadas con este síndrome con CCR que tenían mutaciones en *RPS19*, señalándolo como otro posible gen de predisposición (Kessel 2014).

## 7.5- *SEMA4A*

Este gen se encontró en una familia austríaca con CCR familiar de tipo X con tumores extracolónicos, en la cual se secuenció el exoma completo en combinación con estudios de desequilibrio de ligamiento (Schulz 2014). Al estudiar la segregación en miembros afectados y sanos de esa familia se observó que la variante p.V78M segregaba con la enfermedad aunque presentaba penetrancia incompleta (había personas sanas portadoras). Se hizo un cribado mutacional en 53 pacientes pertenecientes a familias independientes con CCR familiar de tipo

X y se encontraron dos variantes adicionales en *SEMA4A* (p.G484A y p.S326F), también en familias con tumores extracolónicos. Adicionalmente se encontró que el polimorfismo p.P682S estaba enriquecido en la cohorte de CCR familiar de tipo X en comparación con 1.138 controles o con bases de datos de variación genética (OR 6.79, 95% CI 2.63-17.52). Además se estudió el efecto de la missense V78M en la línea celular HCT-116, y se observó una mayor sensibilidad a la vía de PIK3K y la vía de las MAP quinasas en el caso de la proteína mutada en comparación con la wild-type, demostrando que esta mutación afectaba la función proteica. Un estudio posterior contradijo parte de estos resultados mediante el cribado de las variantes p.P682S y p.G484A en una cohorte mayor, sin encontrar diferencias significativas entre casos de CCR y controles sanos (Kinnersley 2016). Sin embargo, los autores del primer estudio sugieren que su cohorte a estudiar es clínicamente distinta al tratarse de CCR familiar de tipo X y a su vez reafirman que la variante V78M si que estaría implicada en predisposición, ya que hay evidencia funcional que lo demuestra. Sin embargo coinciden en que el espectro mutacional de este gen debe ser mejor caracterizado antes de añadirse al diagnóstico clínico (Sill 2016). Las semaforinas se han relacionado con el desarrollo neuronal, la morfogénesis y la modulación de la respuesta inmune, aunque también se han implicado en carcinogénesis, promoviéndola o inhibiéndola en función del contexto (Neufeld 2012, Rehman 2013).

### 7.6-*NTHL1*

Mediante la secuenciación del exoma de 51 individuos pertenecientes a 48 familias con poliposis adenomatosa múltiple y sin alteraciones en *APC* ni *MUTYH,* se encontró la variante c.268C>T (p.Q90*) en homocigosis (Weren 2015b). El patrón de transmisión observado era autosómico recesivo para poliposis, con al menos un miembro afectado de CCR por familia, y con otros tumores primarios incluyendo endometrio en todas las mujeres portadoras de la mutación bialélica, observándose también en algunos portadores menangiomas, cáncer de próstata, linfomas, carcinoma de las células basales, cáncer de mama, cáncer de duodeno y hamartomas del tracto biliar. La mutación p.Q90* se encuentra en una frecuencia relativamente alta (0.15%) en las bases de datos de la población general, pero siempre en heterocigosis. Se cribaron 149 casos adicionales con poliposis pero no se detectó la mutación concluyendo que su frecuencia es baja pero que en homocigosis presenta penetrancia completa para adenomas múltiples. Los perfiles tumorales presentaban un número bajo de mutaciones (no hipermutados) y con una alta frecuencia de transversiones G:C>A:T. Adicionalmente otro estudio identificó una familia con poliposis y múltiples tumores primarios con un cambio que afectaba el splicing (c.709+1G>A) y la misma mutación truncante p.Q90*(Rivera 2015). Este gen codifica para una proteína del sistema de reparación del ADN

por escisión de bases que reconoce y repara el daño oxidativo en las pirimidinas y además tiene actividad glicosilasa de las guaninas alteradas por los radicales de oxígeno.

### 7.7- *FAN1*

Este gen se seleccionó como candidato de predisposición a partir de la secuenciación del exoma en una familia de CCR familiar de tipo X que conllevó la identificación de la mutación C>A (p.C47*) (Seguí 2015). Con el fin de encontrar más cambios en este gen que respalden su papel como causante de predisposición se cribaron 176 familias adicionales con CCR familiar de tipo X. El resultado fue el hallazgo de otra alteración genética de pérdida de función (p.R952*) y tres mutaciones de cambio de aminoácido (p.D140Y, p.P340S y p.R591W). Las células portadoras de los cambios p.C47* y p.D140Y eran sensibles a la mitomicina C. Este compuesto provoca *crosslinks* (uniones covalentes entre cadenas) en el ADN y causa apoptosis en las células que no son capaces de repararlos. Estudios funcionales posteriores permitieron generar una línea celular *knockout* de *FAN1* e introducir un vector expresando el alelo normal de *FAN1*, la mutación *FAN1* p.D140Y o un vector vacío. La sensibilidad a la mitomicina C era similar en el caso del vector vacío que en el vector con *FAN1* mutado, demostrando que el cambio provocaba pérdida de función del sistema reparador de los *crosslinks* en el ADN. La secuenciación del exoma tumoral de los portadores de las mutaciones p.C47* y p.D140Y reveló un perfil no hipermutado con un exceso de transversiones T:A>G:C (10.5%) y C:G>G:C (12.5%). No se encontraron segundos hits somáticos (ni LOH ni mutaciones somáticas). Este gen participa en la vía de la anemia de Fanconi que detecta daños en el ADN y mediante recombinación homóloga los repara.

### 7.8- *BLM*

Este gen codifica para una helicasa dependiente de ATP que desenrolla el ADN de cadena sencilla y de cadena doble en dirección 3'-5'. También participa en la resección del ADN durante la reparación de las brechas de doble hebra, desenrolla el ADN y recluta otras proteínas que cortan el ADN de cadena simple. Mutaciones bialélicas en *BLM* son causantes del síndrome de Bloom, que está caracterizado por deficiencias en el crecimiento, alteraciones en la pigmentación de la piel, inestabilidad cromosómica y predisposición a diferentes tipos de cáncer, incluyendo el CCR y el desarrollo de pólipos adenomatosos (German 1995, Lowy 2001). Aproximadamente un tercio de las personas con síndrome de Bloom tienen ascendencia judía y son portadoras de la mutación fundadora *BLM*(ash) (p. Y736fs) en homocigosis. En contraste con este síndrome, se ha postulado que el tener el cambio *BLM*(ash) en heterocigosis puede comportar un riesgo moderado-bajo de padecer CCR. Mediante un estudio de asociación en

casos y controles, se reportó que los portadores de un alelo *BLM*(ash) tenían un riesgo aumentado de 2.3-2.8 veces de padecer la neoplasia (Gruber 2002), mientras que otro estudio concluía que la mutación no confería un riesgo significativo a desarrollarla (Cleary 2003). Más adelante, este gen se reportó como candidato de predisposición a cáncer de mama mediante secuenciación del exoma en familias con agregación para este cáncer, en el que se encontró una familia con la alteración *BLM* c.2695C>T p.R899*, que se transmitía con herencia autosómica dominante, correspondiendo a una variante de alta penetrancia. Un cribado posterior permitió la detección de otra familia con cáncer de mama portadora de la misma mutación (Thompson 2012). Finalmente, la secuenciación del exoma en 55 casos con CCR de aparición temprana permitió la detección de dos casos, uno con la mutación p.R899*, previamente reportada en cáncer de mama y otro con la alteración c.3558+1G>T, ambas en heterocigosis y con un patrón de herencia dominante. A continuación este mismo grupo realizó un cribado en casos y controles y detectó 3 casos adicionales con las mutaciones p.R899*, p.Q548* y c.98+1G>A. Así, la frecuencia de alteraciones genéticas de pérdida de función resultó significativamente mayor en la cohorte de CCR de aparición temprana que en controles o en bases de datos de variación genética (de Voer 2015).

# OBJETIVOS

**Objetivo principal**

El objetivo general de esta tesis doctoral es la búsqueda de la causa de predisposición genética germinal en pacientes con cáncer colorrectal familiar, incluyendo casos pertenecientes a familias con una fuerte agregación para esta neoplasia, así como pacientes con multiples pólipos o cáncer colorrectal de aparición temprana. Las familias seleccionadas no presentarán alteraciones en los genes hereditarios conocidos y en la mayoría de los casos el patrón de herencia será compatible con herencia autosómica dominante. Se buscarán variantes genéticas de alta penetrancia en nuevos genes candidatos de predisposición para esta enfermedad para de esta forma ampliar el conocimiento existente sobre esta entidad. En un futuro, los resultados de esta tesis doctoral se prevé que se podrían incluir como parte de los programas de diagnóstico genético predictivo para esta enfermedad.

**Objetivos específicos**

- Identificación de nuevos genes candidatos implicados en cáncer colorrectal familiar mediante secuenciación completa del exoma.

- Implementación de una metodología automática de filtrado y priorización de variantes genéticas a partir de los datos de secuenciación del exoma.

- Búsqueda de nuevas vías de señalización relacionadas con la aparición de cáncer colorrectal familiar.

- Cribado de los genes *POLE* y *POLD1* en pacientes con múltiple polipos o CCR de aparición temprana.

# RESULTADOS

**ARTÍCULO 1**

**Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer**

Esteban-Jurado C, Vila-Casadesús M, Garre P, Lozano JJ, Pristoupilova A, Beltran S, Muñoz J, Ocaña T, Balaguer F, López-Cerón M, Cuatrecasas M, Franch-Expósito S, Piqué JM, Castells A, Carracedo A, Ruiz-Ponte C, Abulí A, Bessa X, Andreu M, Bujanda L, Caldés T, Castellví-Bel S. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. Genet Med 2015;17:131-42.

ORIGINAL RESEARCH ARTICLE | Genetics inMedicine

*Open*

# Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer

Clara Esteban-Jurado, MSc[1], Maria Vila-Casadesús, MSc[2], Pilar Garre, MD, PhD[3], Juan José Lozano, PhD[2], Anna Pristoupilova, MSc[4,5], Sergi Beltran, PhD[4], Jenifer Muñoz, MSc[1], Teresa Ocaña, MSc[1], Francesc Balaguer, MD, PhD[1], Maria López-Cerón, MD, PhD[1], Miriam Cuatrecasas, MD, PhD[6], Sebastià Franch-Expósito, MSc[1], Josep M. Piqué, MD, PhD[1], Antoni Castells, MD, PhD[1], Angel Carracedo, MD, PhD[7], Clara Ruiz-Ponte, PhD[7], Anna Abulí, PhD[8], Xavier Bessa, MD, PhD[8], Montserrat Andreu, MD, PhD[8], the EPICOLON Consortium, Luis Bujanda, MD, PhD[9], Trinidad Caldés, PhD[3] and Sergi Castellví-Bel, PhD[1]

**Purpose:** Colorectal cancer is an important cause of mortality in the developed world. Hereditary forms are due to germ-line mutations in *APC*, *MUTYH*, and the mismatch repair genes, but many cases present familial aggregation but an unknown inherited cause. The hypothesis of rare high-penetrance mutations in new genes is a likely explanation for the underlying predisposition in some of these familial cases.

**Methods:** Exome sequencing was performed in 43 patients with colorectal cancer from 29 families with strong disease aggregation without mutations in known hereditary colorectal cancer genes. Data analysis selected only very rare variants (0–0.1%), producing a putative loss of function and located in genes with a role compatible with cancer. Variants in genes previously involved in hereditary colorectal cancer or nearby previous colorectal cancer genome-wide association study hits were also chosen.

**Results:** Twenty-eight final candidate variants were selected and validated by Sanger sequencing. Correct family segregation and somatic studies were used to categorize the most interesting variants in *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1*.

**Conclusion:** We identified new potential colorectal cancer predisposition variants in genes that have a role in cancer predisposition and are involved in DNA repair and the cell cycle, which supports their putative involvement in germ-line predisposition to this neoplasm.

*Genet Med* advance online publication 24 July 2014

**Key Words:** colorectal neoplasm; genetic variant; genetic predisposition to disease; hereditary disease; next-generation sequencing

## INTRODUCTION

Colorectal cancer (CRC) is a very common disease, and its associated mortality rate is quite significant in the developed world. It is estimated that around 5% of the general population will be diagnosed with CRC. Also, as life expectancy increases, the number of CRC cases is also presumed to increase. As an illustrative example, there will be ~473,200 new CRC diagnoses and ~233,900 deaths related to this neoplasm in Europe during 2015.[1]

Germ-line predisposition and environmental factors affect CRC susceptibility, as established for many other complex diseases. Importantly, the inherited germ-line contribution is known to influence about 35% of all cases.[2] Included in this previous group, the Mendelian CRC syndromes are the best characterized CRC cases because an inherited cause corresponds to 5% of total CRC cases. Lynch syndrome and familial adenomatous polyposis are the most frequent forms of Mendelian CRC syndromes. Classic hereditary CRC syndromes are mainly due to germ-line mutations in *APC*, *MUTYH*, and the mismatch repair genes (i.e., *MLH1*, *MSH2*, *MSH6*, *PMS2*).[3,4] Finding the causative mutation in familial CRC also has implications that apply to genetic counseling practices that are of critical importance for the analyzed family. Once it is established in a particular family which individuals are carriers and which are noncarriers, prevention strategies can be directed more precisely to those individuals carrying the causative mutation and who are therefore at risk of developing CRC and other related malignancies. On the other hand, noncarriers can be spared excessive clinical monitoring.

[1]Servei de Gastroenterologia, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, University of Barcelona, Barcelona, Spain; [2]Plataforma de Bioinformática, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Barcelona, Spain; [3]Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos, Madrid, Spain; [4]Centre Nacional d'Anàlisi Genòmica, Parc Científic de Barcelona, Barcelona, Spain; [5]Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; [6]Department of Pathology, Hospital Clinic, Barcelona, Spain; [7]Galician Public Foundation of Genomic Medicine, Centro de Investigación Biomédica en Red de Enfermedades Raras, Genomics Medicine Group, Hospital Clínico, University of Santiago de Compostela, Galicia, Spain; [8]Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, Barcelona, Spain; [9]Gastroenterology Department, Hospital Donostia – Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Country University (UPV/EHU), San Sebastián, Spain. Correspondence: Sergi Castellví-Bel (sbel@clinic.ub.es)

Submitted 8 April 2014; accepted 10 June 2014; advance online publication 24 July 2014. doi:10.1038/gim.2014.89

In addition to hereditary forms, around 30% of CRC cases also present familial aggregation but an unknown inherited cause. Among these cases, familial CRC type X can be used as an example in which the clinical criteria of Lynch syndrome are fulfilled but no alteration of the mismatch repair system is found.[5] Then, the hypothesis of rare high-penetrance mutations in genes yet to be discovered is a very likely explanation for the underlying predisposition in a portion of these familial CRC cases. Therefore, past efforts in this direction included some low-throughput sequencing studies in familial CRC cases of some plausible candidate genes such as *EPHB2*, *GALNT12*, *PTPRJ*, *BMP4*, and *BMPR1A*.[6–10] Next-generation sequencing technologies added a new unbiased approach to facilitate the identification of new genes responsible for predisposition to human disease. Palles et al.[11] recently reported the identification of germ-line mutations in the *POLE* and *POLD1* genes in individuals with multiple colorectal adenomas, carcinoma, or both, or early onset of this disease using whole-genome sequencing. Smith et al.[12] recently performed exome sequencing in a cohort of patients with sporadic CRC enriched for early onset, and variants in genes showing biallelic inactivation were selected. In addition, exome sequencing was completed in CRC familial cases and shared variants were selected within families in an additional study.[13] Finally, a Finnish cohort of familial CRC was also sequenced in order to find rare truncating variants present in two or more cases.[14]

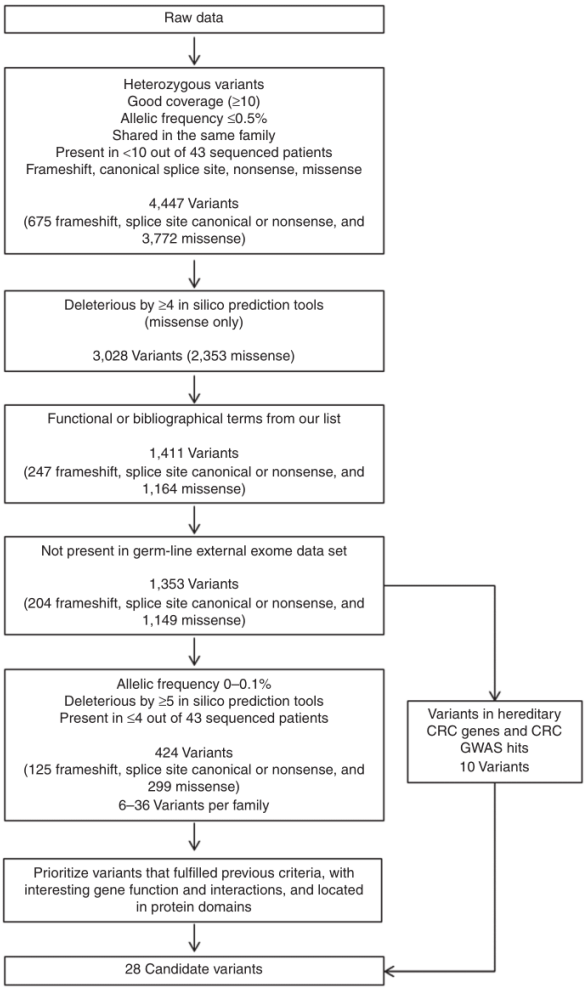



**Figure 1 Schematic of the data analysis steps after whole-exome sequencing.** Forty-three patients with colorectal cancer (CRC) from 29 families with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance were sequenced. Variants remaining after each filtering step are indicated. GWAS, genome-wide association study.

Accordingly, the aim of our study was to find rare predisposition variants in new genes by performing exome sequencing in patients with familial CRC compatible with an autosomal dominant inheritance and without an alteration in the previously known hereditary CRC genes. In doing so, our final goal is to facilitate genetic counseling and to be able to correctly address prevention strategies in these families.

## MATERIALS AND METHODS

### Patients

Forty-three CRC patients from 29 families with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance were selected. Alterations in *APC* or the mismatch repair genes, and homozygous or compound heterozygous mutations in *MUTYH* were previously excluded. Families were chosen based on the following criteria: three or more relatives with CRC, two or more consecutive affected generations, and at least one case of CRC diagnosed before the age of 60. In two families, advanced adenomas (i.e., size ≥1 cm, villous architecture, or high-grade dysplasia) were taken into account as early disease presentation. In addition, other extracolonic cancers were considered in six families. Fourteen families were collected in high-risk CRC clinics (Hospital Clínico San Carlos in Madrid, Hospital Clinic in Barcelona, and Hospital Donostia in San Sebastián), and two patients with CRC were selected to be sequenced from among available affected individuals, preferentially those most distantly related. On the other hand, 15 families were chosen from the EPICOLON Consortium[15] and one patient with CRC per family was selected to be sequenced. This study was approved by the institutional ethics committee of each participating hospital. Written informed consent was obtained at CRC diagnosis on a systematic basis.

Germ-line DNA samples used for exome sequencing were obtained from peripheral blood, whereas formalin-fixed, paraffin-embedded tumor DNA was isolated in some cases for loss of heterozygosity (LOH) studies using the QIAamp DNA Blood Kit or QIAamp Tissue Kit (Qiagen, Redwood City, CA) and following the manufacturer's instructions.

### Exome sequencing

Quality control was applied to DNA samples (3–5 μg needed per reaction at a concentration of 50–300 ng/μl measured

by PicoGreen, A260/280 = 1.7–2, integrity check by agarose electrophoresis). The whole exome was characterized by using the HiSeq2000 platform (Illumina, San Diego, CA) and SureSelectXT Human All Exon V4 for exon enrichment (Agilent, Santa Clara, CA). Initial DNA shearing was performed using the Covaris S2 equipment, achieving an optimal range in the size distribution of fragments. Library size and concentration were checked by capillary electrophoresis (Bioanalyzer 2100; Agilent). Adapters with different indexes for each sample were incorporated during enrichment, allowing samples to be multiplexed before sequencing. After enrichment, the indexed libraries were pooled and massively parallel sequenced using a paired-end 2×75–base pair (bp) read length protocol.

### Data analysis

Base calling and quality control were performed using the Real-Time Analysis software sequence pipeline (Illumina). Sequence reads were trimmed to keep only those bases with a quality >10 and then mapped to the human genome build (hg19/GRCh37) using Genome Multitool,[16] allowing up to four mismatches. Reads not mapped by Genome Multitool were submitted to a last round of mapping with BLAT-like Fast Accurate Search Tool.[17] Uniquely mapping nonduplicate read pairs were locally realigned with Genome Analysis Toolkit.[18] The SAMtools suite (http://samtools.sourceforge.net) was used to call single-nucleotide variants and short insertions/deletions, taking into account all reads per position.[19] Variants with high strand bias ($P > 0.001$ in at least one sample) or regions with low mappability (identified with the Genome Multitool mappability tool as having 75-bp reads and two mismatches)[16] were filtered out. Variant annotation took into account data available in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), the 1000 Genomes Project (http://www.1000genomes.org), the Exome Variant Server (http://evs.gs.washington.edu), and the Geuvadis European Exome Variants Server (http://geevs.crg.eu) and from an in-house database (100 whole genomes of Spanish ancestry from Centre Nacional d'Anàlisi Genòmica (http://www.cnag.cat)). Functional consequences of variants were also predicted by SnpEff (http://snpeff.sourceforge.net) (stop codon, frameshift, splicing, missense, synonymous), as well as by position (coding, intronic, exon–intron junction, untranslated regions). Regarding missense changes, six bioinformatic predictions for pathogenicity were available (PhyloP (http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt), SIFT (Sorting Intolerant From Tolerant; http://sift.bii.a-star.edu.sg), PolyPhen (http://genetics.bwh.harvard.edu/pph2), MutationTaster (http://www.mutationtaster.org), GERP (Genomic Evolutionary Rate Profiling; http://mendel.stanford.edu/SidowLab/downloads/gerp), LRT (likelihood ratio test)).

Because a dominant inheritance pattern was expected, homozygous variants were removed, except for chromosome X nonpseudoautosomal regions in male samples. When analyzing two affected individuals from the same family, only shared variants were selected. Variants with low sequencing coverage (<10) and those with an allelic frequency ≥0.5% in the 1000 Genomes Project, Exome Variant Server, Geuvadis European Exome Variants Server, or the Centre Nacional d'Anàlisi Genòmica in-house database were filtered out. Variants present in >10 of the 43 individuals in our data set were discarded because they most likely corresponded to polymorphisms. Also, only variants predicted to have a strong effect on gene function (frameshift, splice-site canonical, nonsense, and missense) were chosen. Regarding missense variants, we used six bioinformatics tools to select for a deleterious amino acid change, namely, PhyloP (score >0.85), SIFT (score <0.05), PolyPhen (score >0.85), GERP (score >2), Mutation Taster (score >0.5), and LRT (score >0.9), and only those with four or more deleterious predictions were further considered.

Biological functions and pathways of the genes containing variants were annotated with terms and previous bibliography according to NCBI Gene (http://www.ncbi.nlm.nih.gov/gene), Gene Ontology (http://www.geneontology.org/GO), KEGG (http://www.genome.jp/kegg/), and Reactome (http://www.reactome.org/PathwayBrowser/). A list of cancer terms was created from these previous databases (**Supplementary Table S1** online) and used to select variants from among genes that had those terms annotated. All previous filters were performed using an automated pipeline encoded with R software (http://CRAN.R-project.org). CRC specificity of this pipeline regarding function and bibliography was tested by comparing our data with an external germ-line exome sequencing data set with equivalent coverage, which included the same number of patients with chronic lymphocytic leukemia from the International Cancer Genome Consortium (https://www.icgc.org/). Also, variants present in both data sets were filtered out.

Once a variant list per sequenced CRC patient was generated, a thorough manual annotation using NCBI Gene corroborated variant genome position and annotated protein interactions. The amino acid position of missense variants in functional domains, disulfide bonds, or posttranslational modifications was verified, as well as their effect on protein tridimensional structure, when available, using NCBI Protein (http://www.ncbi.nlm.nih.gov/protein) and UniProtKB (http://www.uniprot.org/). Also, their conservation in 46 vertebrates was checked (comparative alignment UCSC (https://genome.ucsc.edu/)).

### Variant prioritization

Once all previous information was available, variant prioritization selected those variants more plausible to be causative of CRC genetic predisposition when they fulfilled more stringent criteria (0–0.1% allelic frequency; present in ≤4 individuals in our data set; ≥5 missense pathogenicity predictions; gene terms and bibliography compatible with cancer; interesting interactions and protein information; and amino acid species conservation). It is noteworthy that variants in genes previously involved in hereditary CRC were carefully checked, as were those genes near previous CRC genome-wide association studies (GWAS) hits (**Supplementary Table S2** online) with less strict criteria (missense considered deleterious by four or more bioinformatics tools). As previously specified, thresholds to select variants were

**Table 1** Description of the final 28 prioritized variants, including gene category, frequency, and functional information

| Gene | Mutation | Category | Genotype frequency (EVS) | In silico | LD region | Domain/region | Interactions | Function | OMIM |
|---|---|---|---|---|---|---|---|---|---|
| *CDKN1B* | c.195G>T(p.Q65H) | New | 0/6503 | 5 | N | Cyclin-dependent kinase inhibitor region | AKT1, CCND2, CDKN1A, TSC2, CDK2, CCNA1, CCNA2, CCNE1, CDK4, CCND1 | DNA damage response, induction of apoptosis, negative regulation of cell proliferation | 600778; Multiple endocrine neoplasia, hereditary prostate cancer |
| *XRCC4* | c.497_498delTG (p.V166Efs*3) | New | 0/6503 | FS | N | — | LIG4, APFL, APTX, BIN1, CAAP1, CHD3, DDX5, ERG, JUN, PRKDC | DNA ligation involved in DNA repair, double-strand break repair | 194363; Peripheral B-cell lymphoma |
| *EPHX1* | c.293G>A(p.R98Q) | New | 0/6503 | 6 | N | Epoxide hydrolase *N* terminus region | CDKN2AIP, ATP4A, NRF1, SUMO1 | Metabolism of xenobiotics by cytochrome P450 | 132810; Lymphoproliferative disorder |
| *NFKBIZ* | c.2153_2154dupAT (p.*719Ifs*10) | New | 2/6258 | FS | N | — | NFKB1, NFKB2, STAT3 | TNF-α/NF-κB signaling pathway | 608004 |
| *SMARCA4* | c.295C>T(p.R99W) | New | 0/6503 | 5 | N | Interaction with SS18L1/CREST | PMS2, CDKN1A, STK11, MYC, BRCA1, BCL6, TP53, BRCA1, IRF1, CREBBP | Regulation of cell growth, regulation of mitotic cell cycle, chromatin remodeling | 603254; Rhabdoid tumor predisposition syndrome, SCCOHT (ref. 31) |
| *BARD1* | c.1811-2A>G | New | 0/6503 | SP | N | BRCT1 domain (exon 9 skipping) | MSH2, BRIP1, MLH1, MSH6, TP53, BRCA2, BRCA1 | DNA repair, apoptosis, cell cycle arrest | 601593; Breast cancer susceptibility, ovarian cancer susceptibility |
| *BRIP1* | c.1702_1703delAA (p.N568Wfs*9) | New | 0/6503 | FS | N | — | BARD1, PMS2, PMS1, MLH1, BRCA1, LMNA, BLM | Double-strand break repair, DNA damage checkpoint | 605882; Breast cancer susceptibility, ovarian cancer susceptibility |
| *RB1* | c.491A>G(p.K164R) | New | 0/6503 | 4 | N | DUF3452 domain | CDKN1A, RUNX2, PIK3R1, BRCA1, EP300, MYC, SMARCA4, TP53 | Negative regulation of cell growth and epithelial cell proliferation, cell cycle checkpoint | 614041; Retinoblastoma, pinealoma, sarcoma, melanoma (ref. 43) |
| *AKR1C4* | c.667C>T(p.R223*) | New[a] | 0/6503 | NS | 10p15.3–p15.1 (ref. 39) | — | BAG3 | Metabolism of xenobiotics by cytochrome P450 | 600451; CRC |
| *CARD9* | c.1228G>T(p.E410*) | New | 0/6503 | NS | 9q33.3–q34.3 (ref. 39) | — | BCL10, CSNK2A1 | Regulation of apoptotic process | 607212; Innate immunity |
| *NSMCE2* | c.346delT(p.S116Lfs*18) | New | 2/6249 | FS | N | — | NSMCE1, RAD21, SMC1A, SMC3, SMC5, SMC6, STAG2, SUMO1 | Double-strand break repair via homologous recombination, mitosis, telomere maintenance | — |

CRC, colorectal cancer; GWAS, genome-wide association study; FS, frameshift; N, no; NS, nonsense; SP, canonical splice site; SCCOHT, ovarian small-cell carcinoma of hypercalcemic type.

[a]Category: new genes corresponded mainly to those not previously involved in CRC predisposition, except for *AKR1C4* (Smith et al.[12], Gylfe et al.[14]), *CCDC18* (Gylfe et al.[14]) and ATM (Gala et al.[38]), with seminal evidence. [b]This same variant was previously found in two juvenile polyposis patients (Sayed et al.[45] and Greenman et al.[46]).

EVS: Exome Variant Server (version ESP6500SI-V2). 0/6503 indicates absence of the variant in the database. This data set is listed as representative of variant frequency. The 1000 Genomes Project and Geuvadis European Exome Variants Server were also checked but showed similar results.

In silico: number of deleterious predictions by bioinformatics tools used.

LD region: previously reported linkage disequilibrium regions for familial CRC are listed if the variant was located within.

Domain/region: the protein domain or region where the variant is located. If the position affects a disulfide bond or corresponds to a modified residue, the information is also listed.

Interactions: those reported with proteins with a role in cancer and other relevant partners.

OMIM: entry number (if available), hereditary involvement in malignancies and in some cases also somatic contribution.

**Table 1** Continued on next page

**Table 1** Continued

| Gene | Mutation | Category | Genotype frequency (EVS) | In silico | LD region | Domain/region | Interactions | Function | OMIM |
|---|---|---|---|---|---|---|---|---|---|
| *BMPR1A* | c.1327C>T(p.R443C) | Hereditary | 7/6496 | 6 | N | Protein kinase domain | BMP4, SMAD1, RUNX2, SMURF1, MYC, BMP2, BMP7, BMP6, BMPR1B, BAMBI | Epithelial cell proliferation, cell differentiation, BMP signaling pathway | 601299; Juvenile polyposis syndrome[b], hereditary mixed polyposis syndrome, CRC |
| *CCDC18* | c.3268G>A(p.R1036H) | New[a] | 1/5839 | 4 | N | Coiled coil/SMC prok B region | UBC | Antigen found differentially expressed in sarcoma | — |
| *MYC* | c.906C>A(p.H302Q) | GWAS | 4/6499 | 4 | N | Myc amino-terminal region | BRCA1, CDKN1A, MLH1, BMPR1A, MSH2, POLE, POLD1, SMARCA4 | Response to DNA damage stimulus, cell cycle arrest, Wnt signaling, TGF-β signaling | 190080; Burkitt lymphoma, CRC, prostate cancer |
| *POLE* | c.5473A>T(p.H1810L) | Hereditary | 0/6503 | 4 | 12q24.32 (ref. 40) | DUF1744 domain | MDM2, MYC, RAD17, PCNA, XRCC5 | DNA repair, mitotic cell cycle, nucleotide excision repair, telomere maintenance | 174762; CRC |
| *TSC2* | c.5116C>T(p.R1706C) | New | 3/6495 | 6 | N | Rap-GAP domain | AKT1, CDKN1B, GSK3B, MAPK1, CCND1, CCND2, CCND3 | Negative regulation of cell proliferation, negative regulation of Wnt receptor signaling pathway | 191092; Tuberous sclerosis, renal-cell cancer, angiomyolipoma, subependymal giant-cell astrocytoma, rhabdomyoma (ref. 43) |
| *RAD52* | c.590_593dupAACC (p.S199Tfs*88) | New | 0/6503 | FS | N | — | MCM2, ERCC4, RAD51 | DNA repair, double-strand break repair via homologous recombination | 600392 |
| *BMP4* | c.1001C>T(p.A334V) | GWAS | 0/6503 | 6 | 14q23.1–14q24.1 (ref. 41) | TGF-β region | BMPR1A, BMPR1B, BMPR2 | Apoptosis, negative regulation of cell proliferation, BMP signaling pathway | CRC (ref. 9) |
| *DUSP4* | c.82G>A(p.A28T) | GWAS | 0/6503 | 4 | N | Regulatory rhodanese domain of dual specificity phosphatases | MAPK1, MAPK14, MAPK3, MAPK8, MAPK9 | Inactivation of MAPK activity, Toll signaling pathway | 602747 |

CRC, colorectal cancer; GWAS, genome-wide association study; FS, frameshift; N, no; NS, nonsense; SP, canonical splice site; SCCOHT, ovarian small-cell carcinoma of hypercalcemic type.

[a]Category: new genes corresponded mainly to those not previously involved in CRC predisposition, except for *AKR1C4* (Smith et al.[12], Gylfe et al.[14]), *CCDC18* (Gylfe et al.[14]) and ATM (Gala et al.[38]), with seminal evidence. [b]This same variant was previously found in two juvenile polyposis patients (Sayed et al.[45] and Greenman et al.[46]).

EVS: Exome Variant Server (version ESP6500SI-V2). 0/6503 indicates absence of the variant in the database. This data set is listed as representative of variant frequency. The 1000 Genomes Project and Geuvadis European Exome Variants Server were also checked but showed similar results.

In silico: number of deleterious predictions by bioinformatics tools used.

LD region: previously reported linkage disequilibrium regions for familial CRC are listed if the variant was located within.

Domain/region: the protein domain or region where the variant is located. If the position affects a disulfide bond or corresponds to a modified residue, the information is also listed.

Interactions: those reported with proteins with a role in cancer and other relevant partners.

OMIM: entry number (if available), hereditary involvement in malignancies and in some cases also somatic contribution.

Table 1 Continued on next page

**Table 1** Continued

| Gene | Mutation | Category | Genotype frequency (EVS) | In silico | LD region | Domain/region | Interactions | Function | OMIM |
|---|---|---|---|---|---|---|---|---|---|
| *LAMA5* | c.4930C>T(p.R1644C) | GWAS | 0/6503 | 6 | 20q13.11–13.33 (ref. 42) | Laminin IV type A domain | LAMC1, SMAD2 | Regulation of cell proliferation, cell differentiation, regulation of cell adhesion | 601033 |
| *CETN2* | c3+2T>C | New | 0/6503 | SP | N | — | PARP1, POU5F1, PRMT6, RAD23B, SFI1, USP39 | Mitosis, centriole replication, regulation of cytokinesis | 300006 |
| *MAP9* | c.681dupA (p.A228Sfs*4) | New | 0/6503 | FS | N | — | EP300, TP53, SH3KBP1 | Cell cycle | 610070 |
| *ENG* | c.934C>T(p.A312V) | Hereditary | 0/6503 | 4 | 9q33.3–q34.3 (ref. 39) | Interaction with EGL | TGFBR1, BMP7, BMP2 | Regulation of cell proliferation, regulation of cell adhesion, BMP signaling pathway | Hamartomatous or hyperplastic/mixed polyposis (ref. 44) |
| *RUNX2* | c.622T>C(p.P208S) | New | 0/6503 | 6 | N | Runt domain | BMPR1A, EP300, RB1, SMURF2, SMURF1, JUN, SMAD2, SMAD3, SMAD6, XRCC6 | BMP signaling pathway | 600211; Osteosarcoma |
| *PIK3R1* | c.220G>A(p.V74I) | New | 0/6503 | 6 | N | SH3 domain | AKT1, PIK3CA, EGFR, BRCA1, EP300, GSK3B, TGFBR1 | Insulin receptor signaling pathway, EGF receptor signaling pathway, focal adhesion | 171833; Glioblastoma |
| *SMURF2* | c.838C>T(p.H280Y) | New | 0/6503 | 5 | N | WW2 domain | SMAD7, GSK3B, RUNX2, TGFBR1, SMAD1, SMAD2, SMAD3, SMAD5, SMAD6, NEDD4 | BMP signaling pathway, TGF-β-signaling pathway | 605532; Different tumor types |
| *ATM* | c.8327T>C(p.I2776T) | New[a] | 0/6503 | 6 | 11q22.3 (ref. 39) | PI3K catalytic domain | MLH1, MSH2, MSH6, STK11, SMAD7, BRCA2, BRCA1, TP53 | DNA repair, cell cycle arrest, mitotic cell cycle spindle assembly checkpoint | 607585; Ataxia-telangiectasia lymphoma, familial breast cancer, kidney cancer, multiple sessile serrated adenomas |
| *DHX9* | c.286T>G(p.V40G) | New | 0/6503 | 6 | N | DSRM1 domain | POLD1, AKT1, BRCA1, JUN, PRKDC | CRD-mediated mRNA stabilization, RNA splicing, innate immune response | 603115 |

CRC, colorectal cancer; GWAS, genome-wide association study; FS, frameshift; N, no; NS, nonsense; SP, canonical splice site; SCCOHT, ovarian small-cell carcinoma of hypercalcemic type.

[a]Category: new genes corresponded mainly to those not previously involved in CRC predisposition, except for *AKR1C4* (Smith et al.[12], Gylfe et al.[14]); *CCDC18* (Gylfe et al.[14]) and ATM (Gala et al.[38]), with seminal evidence. [b]This same variant was previously found in two juvenile polyposis patients (Sayed et al.[45] and Greenman et al.[46]).

EVS: Exome Variant Server (version ESP6500SI-V2). 0/6503 indicates absence of the variant in the database. This data set is listed as representative of variant frequency. The 1000 Genomes Project and Geuvadis European Exome Variants Server were also checked but showed similar results.

In silico: number of deleterious predictions by bioinformatics tools used.

LD region: previously reported linkage disequilibrium regions for familial CRC are listed if the variant was located within.

Domain/region: the protein domain or region where the variant is located. If the position affects a disulfide bond or corresponds to a modified residue, the information is also listed.

Interactions: those reported with proteins with a role in cancer and other relevant partners.

OMIM: entry number (if available), hereditary involvement in malignancies and in some cases also somatic contribution.

applied for sequencing coverage, allelic frequency, presence in our data set, predictions by bioinformatics tools, presence in the functional and bibliography term list, and absence in the external exome set. On the other hand, there were no thresholds for some other additional variant/gene information that was used if available to further select for variants present within each family. This information included protein function and interactions; amino acid position in functional domains, disulfide bonds, or posttranslational modification sites; effect on protein tridimensional structure; and amino acid species conservation. Therefore, variants also complying with these last criteria were considered more interesting functionally and were further selected as final candidates. Some studied CRC families had up to four variants prioritized, whereas other families had none.

### Variant validation, segregation analysis, and tumor loss of heterozygosis

Exome sequencing results for prioritized variants were validated using specific primers for polymerase chain reaction amplification designed using Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and Sanger sequencing (GATC Biotech, Cologne, Germany).

Segregation analysis of the prioritized variants was performed in additional family members (those with CRC and advanced adenoma) when germ-line DNA was available. When possible, somatic LOH was studied in tumor DNA of patients carrying the selected variants. LOH was tested by comparing Sanger sequencing results for germ-line and tumor DNA of the same individual. In addition, microsatellite markers within and around the gene of interest were used when LOH of the wild-type allele was suspected. Sanger and microsatellite markers results were always concordant. Primer details are listed in **Supplementary Table S3** online.

### Network analysis

Ingenuity Pathway Analysis (IPA; Qiagen; http://www.qiagen.com/ingenuity) was used to perform a core analysis to check the putative enrichment for canonical pathways, disease and biological functions, and molecular networks among the 18 final candidate genes carrying variants that either fulfilled CRC family segregation or could not be tested (variants without correct family segregation were not included). IPA was run with an experimentally observed filter, aiming to obtain information based on confirmed data. The IPA networks generation algorithm transformed the gene list into a network set using Global Molecular Network connections and Ingenuity Pathways Knowledge Base.

## RESULTS

Whole-exome sequencing was performed in 43 patients with CRC from 29 families (2 affected relatives from 14 families and 15 unrelated patients with CRC) with strong disease aggregation compatible with an autosomal dominant pattern of inheritance but without mutations in known hereditary CRC.

After sequencing, mean coverage was >95× in all samples. Raw data were analyzed using an automatic pipeline that selected only very rare variants (0–0.1%) producing a putative loss of function and located in genes with a role compatible with cancer. Also, variants in genes previously involved in hereditary CRC or nearby previous CRC GWAS hits were prioritized (**Figure 1**). Initial filtering removed variants in homozygosis, those with low coverage, those not shared in the same family, those with a frequency ≥0.5% and those present in ≥10 of the 43 individuals in our data set. On the other hand, frameshift, nonsense, canonical splice-site, and missense variants were selected (4,447 variants: 675 frameshift, splice-site canonical, or nonsense and 3,772 missense). When missense variants complying with most pathogenicity prediction tools (designated deleterious by at least 4 of 6 tools) were selected, 2,353 remained. Of these, 1,411 variants annotated with functional or bibliographical terms from our cancer list were selected.

CRC specificity of this pipeline regarding function and bibliography was tested by comparing our set with an external germ-line exome sequencing data set for a different disease. After applying frequency, heterozygosity, function, and bibliography filters, a *t* test was used to compare the mean number of frameshift, splice-site canonical, or nonsense variants per individual in the two exome data sets. In doing so, our pipeline selected more variants in our exome data set ($mean_{CRC} = 41.87$; $mean_{external} = 34.05$; $P = 3.75 \times 10^{-10}$), supporting the CRC specificity of our pipeline.

After checking the aforementioned pipeline specificity, we continued with variant filtering, and 1,353 variants that were not present in the external data set were further considered. At this stage, 10 variants in genes previously implicated in CRC predisposition and CRC GWAS hits that fulfilled previous criteria had been selected as final candidates. Stricter filtering was applied to prioritize variants in new genes, including allelic frequency 0–0.1%, presence in ≤4 of the 43 individuals in our data set, and compliance with most pathogenicity prediction tools for missense classification (designated deleterious by ≥5 tools), leaving 424 selected variants (125 frameshift, splice-site canonical, or nonsense and 299 missense), ranging from 6 to 36 variants per family. Filtering for the 10 variants in genes previously implicated in CRC predisposition and CRC GWAS hits was the same except for pathogenicity prediction tools for missense classification (designated deleterious by ≥4 tools). Finally, as previously specified, thresholds regarding sequencing coverage, allelic frequency, presence in our data set, prediction by bioinformatics tools, presence in functional and bibliography term lists, and absence from the external exome set were applied in order to select variants. On the other hand, there were no thresholds for some other additional variant/gene information that was used, if available, to further select for variants present within each family. This information included protein function and interactions; amino acid position in functional domains, disulfide bonds, or posttranslational modification sites; effect on protein tridimensional structure; and amino acid species conservation. Therefore, variants also complying with these last

**Table 2** Results for the final 28 prioritized variants regarding Sanger validation, family segregation, and somatic status

| Gene | Category | Mutation | Samples | Sanger | CRC seg | AA seg | LOH | COSMIC |
|---|---|---|---|---|---|---|---|---|
| *CDKN1B* | New | c.195G>T(p.Q65H) | FAM14 | Y | Y 2/2 | Y 2/2 | Y | 7/636 (1.1%) |
| *XRCC4* | New | c.497_498delTG(p.V166Efs*3) | FAM10 | Y | Y 2/2 | | Y[b] | 9/577 (1.56%) |
| *EPHX1* | New | c.293G>A(p.R98Q) | FAM11 | Y | Y 2/2 | | Y[b] | 18/576 (3.13%) |
| *NFKBIZ* | New | c.2153_2154dupAT (p.*719Ifs*10) | FAM8 | Y | Y 4/4 | | N | 11/576 (1.91%) |
| *SMARCA4* | New | c.295C>T(p.R99W) | FAM3 | Y | Y 2/2 | Y 1/1 | N | 60/646 (9.29%) |
| *BARD1* | New | c.1811-2A>G | H458 | Y | Y 2/2 | | N | 18/577 (3.12%) |
| *BRIP1* | New | c.1702_1703delAA (p.N568Wfs*9) | H463 | Y | NA | | Y[b] | 28/636 (4.4%) |
| *RB1* | New | c.491A>G(p.K164R) | H466 | Y | NA | | Y | 126/713 (17.67%) |
| *AKR1C4* | New[a] | c.667C>T(p.R223*) | H467 | Y | NA | | N | 13/576 (2.26%) |
| *CARD9* | New | c.1228G>T(p.E410*) | H456 | Y | NA | | N | 13/578 (2.25%) |
| *NSMCE2* | New | c.346delT(p.S116Lfs*18) | H465 | Y | NA | | N | 3/602 (0.5%) |
| *BMPR1A* | Hereditary | c.1327C>T(p.R443C) | H460 | Y | NA | | N | 23/757 (3.04%) |
| *CCDC18* | New[a] | c.3268G>A(p.R1036H) | H460 | Y | NA | | N | 22/597 (3.69%) |
| *MYC* | GWAS | c.906C>A(p.H302Q) | H460 | Y | NA | | N | 6/636 (0.94%) |
| *POLE* | Hereditary | c.5473A>T(p.H1810L) | H462 | Y | NA | | NA | 40/577 (6.93%) |
| *TSC2* | New | c.5116C>T(p.R1706C) | I139 | Y | NA | | NA | 28/635 (4.41%) |
| *RAD52* | New | c.590_593dupAACC (p.S199Tfs*88) | FAM13 | Y | Y 3/3 | N 0/1 | N | 9/577 (1.56%) |
| *BMP4* | GWAS | c.1001C>T(p.A334V) | FAM3 | Y | Y 2/2 | N 0/1 | N | 8/577 (1.39%) |
| *DUSP4* | GWAS | c.82G>A(p.A28T) | H468 | Y | N 1/2 | | | 5/577 (0.87%) |
| *DUSP4* | GWAS | c.82G>A(p.A28T) | H466 | Y | NA | | | 5/577 (0.87%) |
| *LAMA5* | GWAS | c.4930C>T(p.R1644C) | FAM8 | Y | N 3/4 | | | 51/576 (8.85%) |
| *CETN2* | New | c.3+2T>C | FAM8 | Y | N 2/4 | | | 4/577 (0.69%) |
| *MAP9* | New | c.681dupA(p.A228Sfs*4) | FAM8 | Y | N 2/4 | | | 25/576 (4.34%) |
| *ENG* | Hereditary | c.934C>T(p.A312V) | H469 | Y | N 1/2 | | | 9/635 (1.42%) |
| *RUNX2* | New | c.622T>C(p.P208S) | FAM1 | Y | N 2/3 | | | 13/577 (2.25%) |
| *PIK3R1* | New | c.220G>A(p.V74I) | FAM4 | Y | N 3/4 | | | 83/961 (8.64%) |
| *SMURF2* | New | c.838C>T(p.H280Y) | FAM12 | Y | N 3/4 | | | 16/577 (2.77%) |
| *ATM* | New[a] | c.8327T>C(p.I2776T) | H470 | Y | N 1/2 | | | 210/707 (29.7%) |
| *DHX9* | New | c.286T>G(p.V40G) | FAM2 | N | | | | 21/576 (3.65%) |

AA seg, advanced adenoma segregation; CRC seg, colorectal cancer segregation; GWAS, genome-wide association study; LOH, loss of heterozygosity; N, no; NA, not available; Y, yes.

[a]Category: new genes corresponded mainly to those not previously found to be involved in CRC predisposition, except for *AKR1C4* (Smith et al.[12]; Gylfe et al.[14]), *CCDC18* (Gylfe et al.[14]), and *ATM* (Gala et al.[38]), with seminal evidence.

Samples: identifier for individuals in which the variant was found in exome sequencing. Samples with the header "FAM" belong to families with two sequenced individuals who share the same variant.

Sanger: refers to Sanger sequencing confirmation of the variant found by exome sequencing.

CRC seg: proportion of CRC cases within the family that carry the variant, including additional CRC cases when available.

AA seg: proportion of advanced adenomas cases within the family that carry the variant.

LOH: depletion of the wild-type allele in tumor DNA in comparison with the germ line.

COSMIC: proportion of somatic mutations found in large-intestine carcinoma for the gene carrying the variant (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/).

[b]Indicates results for both Sanger sequencing and microsatellite markers.

criteria were considered more interesting functionally and were further selected as final candidates. Some CRC families had up to four selected variants, whereas other families had none. The final 28 prioritized variants are shown in Table 1.

Candidate variants were subsequently validated by Sanger sequencing, and, if confirmed, segregation was studied in additional affected family members when available (Table 2). A variant in *DHX9* was not confirmed (1 of 28 prioritized variants).

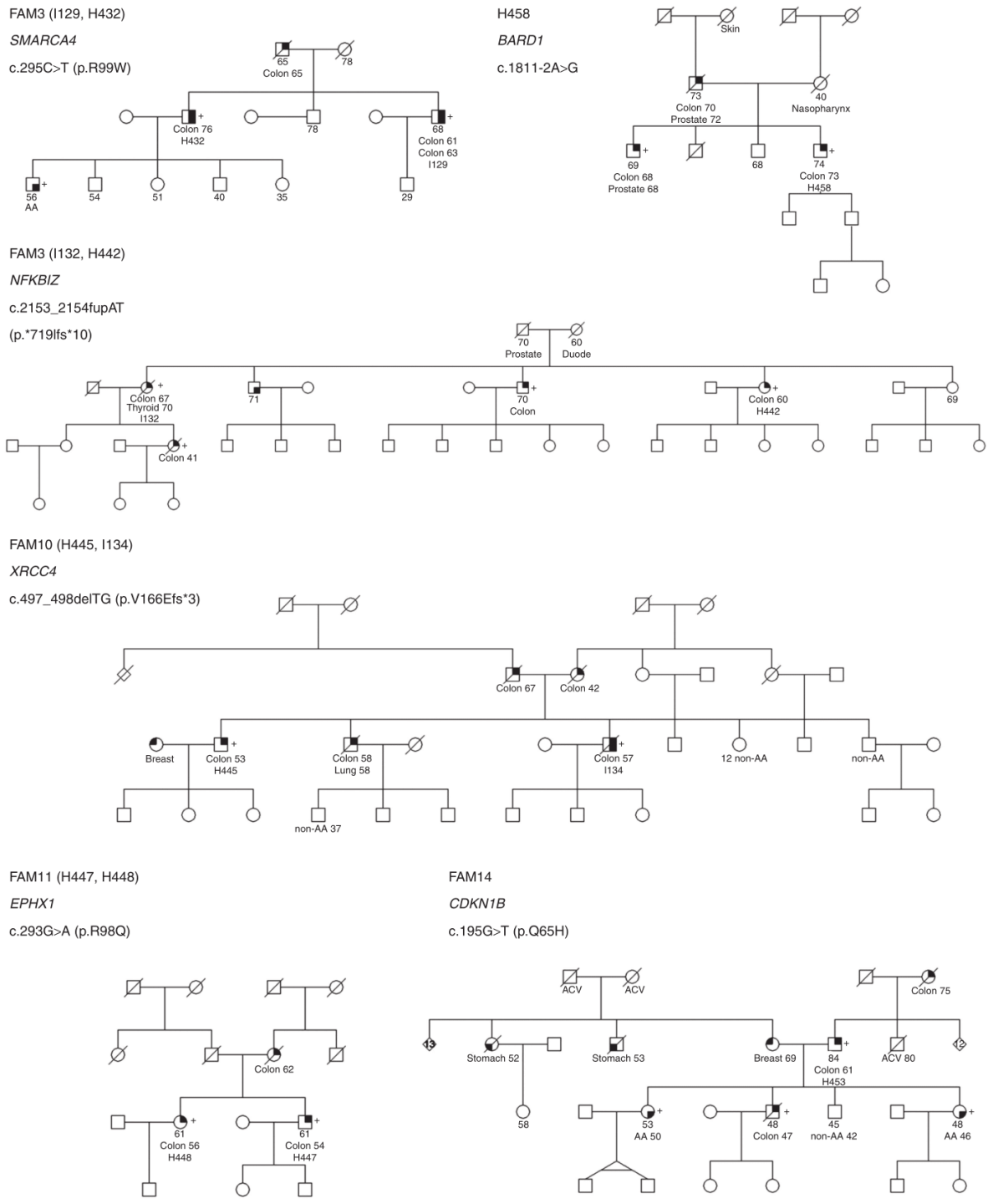



**Figure 2 Pedigrees from families FAM3, FAM8, FAM10, FAM11, FAM14, and H458 are shown.** Filled symbols indicate those affected by colorectal cancer (upper right quarter), adenoma(s) (lower right quarter), stomach cancer (lower left quarter), or breast cancer (upper left quarter). Colon, breast, stomach, thyroid, lung, prostate, and nasopharynx refer to the type of cancer. (+), mutation carrier; (–), wild type. AA, advanced adenoma; ACV, cerebrovascular accident; Duode, duodenum carcinoma; non-AA, nonadvanced adenoma.

LOH in tumor DNA was analyzed in variants with correct disease segregation when possible (Table 2; **Supplementary Figure S1** online). Among the 28 prioritized variants, the best candidates for being involved in CRC genetic predisposition included those located in genes such as *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1* because they segregated correctly with disease presentation (Figure 2; the rest of families are shown in **Supplementary Figure S2** online). Regarding variants in these genes, it is expected that three of them abolish protein function and the other three are missense changes with strongly deleterious in silico predictions. Family segregation and tumor LOH of the wild-type allele was positive for variants in *CDKN1B*, *XRCC4*, and *EPHX1*. Other interesting variants were found in *BRIP1*, *RB1*, *AKR1C4*, *CARD9*, *NSMCE2*, *BMPR1A*, *CCDC18*, *MYC*, *POLE*, and *TSC2*, although segregation analysis was not feasible. Nevertheless, tumor LOH of the wild-type allele was present for the *BRIP1* and *RB1* variants. It is noteworthy that variants in *BMP4* and *RAD52* showed correct family segregation for CRC, but they did not correlate with advanced adenoma presentation, although they can still be considered interesting candidates. As reported in the COSMIC database, somatic mutations in sporadic CRC were more common for the *RB1*, *SMARCA4*, and *POLE* genes (Table 2). Candidate variants within genes previously implicated in CRC predisposition and CRC GWAS hits included those located in *AKR1C4*, *BMPR1A*, *CCDC18*, *MYC*, *POLE*, *BMP4*, *DUSP4* (present in two independent families), *LAMA5*, *ENG*, and *ATM*. The variant in the *BMP4* gene segregated with CRC but not with advanced adenoma. *DUSP4*, *LAMA5*, *ENG*, and *ATM* variants did not segregate with disease, whereas disease segregation could not be tested for variants in *AKR1C4*, *BMPR1A*, *CCDC18*, *MYC*, and *POLE*. The *POLE* variant did not correspond to those previously reported and did not fall within the exonuclease or polymerase domains.[11]

In addition, we performed IPA to test for a putative enrichment for canonical pathways, disease and biological functions, and molecular networks among the 18 final candidate genes carrying variants that either fulfilled CRC family segregation or for which segregation analysis was not possible. A relevant network that contains 9 of the 18 genes was obtained with an overrepresentation of the DNA Replication, Recombination and Repair, Cell Cycle, Connective Tissue Development and Function terms (**Supplementary Figure S3a** online). On the other hand, when testing for canonical pathways in our set, the "Role of BRCA1 in DNA damage response" network included some of our more interesting candidates, such as *SMARCA4*, *BARD1*, *BRIP1*, and *RB1* (**Supplementary Figure S3b** online).

## DISCUSSION

Exome sequencing in 43 patients with CRC from 29 families with strong disease aggregation identified new potential CRC predisposition variants in *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1*.

CDKN1B (p27, Kip1) binds to cyclin E/A-CDK2 and cyclin D-CDK4 complexes and hinders their activation. By doing so it exerts control on cell cycle progression.[20] The c.195G>T (p.Q65H) mutation is located inside the cyclin-dependent kinase inhibitor region, particularly in the β-hairpin (residues 61–71), which interacts with CDK2.[21] Therefore, this variant most likely affects the normal interaction between CDKN1B and CDK2, causing a deregulation in cell cycle progression. Interestingly, germ-line mutations in this gene have been previously implicated in multiple endocrine neoplasia.[22] Moreover, a polymorphism in this gene has been significantly associated with hereditary prostate cancer.[23]

XRCC4 is involved in the repair of DNA double-strand breaks by nonhomologous end joining and the completion of V(D)J recombination events, along with DNA ligase IV and the DNA-dependent protein kinase.[24] The c.497_498delTG (p.V166Efs*3) mutation is predicted to abolish protein function, and it is likely to contribute to genomic instability and tumorigenesis.

The EPHX1 enzyme converts epoxides produced by the degradation of aromatic compounds to trans-dihydrodiols, which afterward are conjugated and excreted from the body. Thus, EPHX1 can be considered an important biotransformation protein.[25] The affected residue of the c.293G>A (p.R98Q) mutation is located in the epoxide hydrolase *N*-terminus region. Because EPHX1 alleles can have a differential efficiency in procarcinogen detoxification, it can be postulated that they may affect cancer risk in a specific manner.[26]

*NFKBIZ* is involved in inflammatory response through regulation of nuclear factor-κB transcription factor complexes.[27,28] The c.2153_2154dupAT (p.*719Ifs*10) mutation disrupts a stop codon, producing an abnormally long *C*-terminal region. This could affect the interactions with nuclear factor-κB complexes that bind to that region, altering the transcriptional regulation of its target genes and leading to cancer predisposition.

The SMARCA4 protein is a component in the large SNF/SWI complex involved in chromatin remodeling. This complex is necessary to activate the transcription of genes that are usually repressed by chromatin.[29] The c.295C>T (p.R99W) mutation is located in the region necessary for the interaction with SS18L1, which inhibits transcription of c-FOS and is required for dendritic growth and branching in cortical neurons. It can be hypothesized that this variant may cause predisposition to CRC by impairing this network and causing abnormal cell proliferation. Germ-line mutations in this gene can cause rhabdoid tumor predisposition syndrome type 2[30] and small-cell carcinoma of the ovary, hypercalcemic type.[31]

BARD1 interacts with the well-known BRCA1 protein. Both proteins, along with others, participate in several cellular pathways involved in DNA damage repair, ubiquitination, and transcriptional regulation to preserve genomic stability.[32] The c.1811-2A>G mutation is predicted to cause exon 9 skipping, disrupting the BRCT1 domain, which is postulated to participate in ligand binding according its structure.[33] This domain is highly homologous to the BRCA1 BRCT1 domain, which is considered to bind substrates of DNA damage response kinases such as *ATM*. Moreover, tumor-associated mutations in the

BRCT domains of BRCA1 abolish binding to phosphorylated substrates.[34] Thus, disruption of this BARD1 domain likely affects its capacity to interact with other proteins, abolishing its tumor suppressor function. Germ-line mutations in this gene predispose to breast and ovarian cancer,[35] and its expression has been involved in differential CRC prognosis.[36]

Focusing on the best candidates to be involved in CRC genetic predisposition (*CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1*), it is remarkable that, as highlighted by the IPA analysis and previous studies, most of them have been formerly involved in DNA repair, cell cycle, and predisposition to germ-line cancer, which supports their putative involvement in genetic predisposition to CRC as well. Among them, mutated *BARD1* and *BRIP1* have been found in the germ-line DNA of breast cancer patients described in several reports.[35,37] In addition, DNA repair constitutes a cellular mechanism with proven importance in the genetic predisposition for CRC.[3]

Among those variants within genes previously involved in CRC predisposition or located in CRC GWAS hits, it is remarkable that so far three independent studies, including ours, have identified interesting variants in the *AKR1C4* gene.[12,14]

Taken together, we could conclude that our results highlight some interesting candidates for CRC germ-line predisposition, with an overrepresentation of genes involved in DNA repair and the cell cycle. We identified several putative new genes predisposing to CRC and some with previous involvement in cancer predisposition, including *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1*, that deserve to be considered in additional familial CRC cohorts with an unknown hereditary cause. Furthermore, once their role in hereditary CRC is confirmed, more complex functional studies would be warranted to help understand the molecular mechanism of disease predisposition.

## SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

## ACKNOWLEDGMENTS

We are sincerely grateful to the Centre Nacional d'Anàlisi Genòmica and the Biobank of Hospital Clínic–IDIBAPS, Barcelona, for technical help, and the International Cancer Genome Consortium for access to exome data set. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona. CEJ and JM are supported by a contract from CIBERehd. MVC is supported by Ministerio de Educación, Cultura y Deporte (FPU12/05138). PG and SCB are supported by a contract from the Fondo de Investigación Sanitaria (JR13/00013 and CP 03-0070, respectively). CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III. This work was supported by grants from the Fondo de Investigación Sanitaria/FEDER (10/00641, 11/00219, 11/00681, RD12/0036/006, 13/02588), the Ministerio de Economía y Competitividad (SAF2010-19273), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), COST Action BM1206 (SCB and CRP), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255).

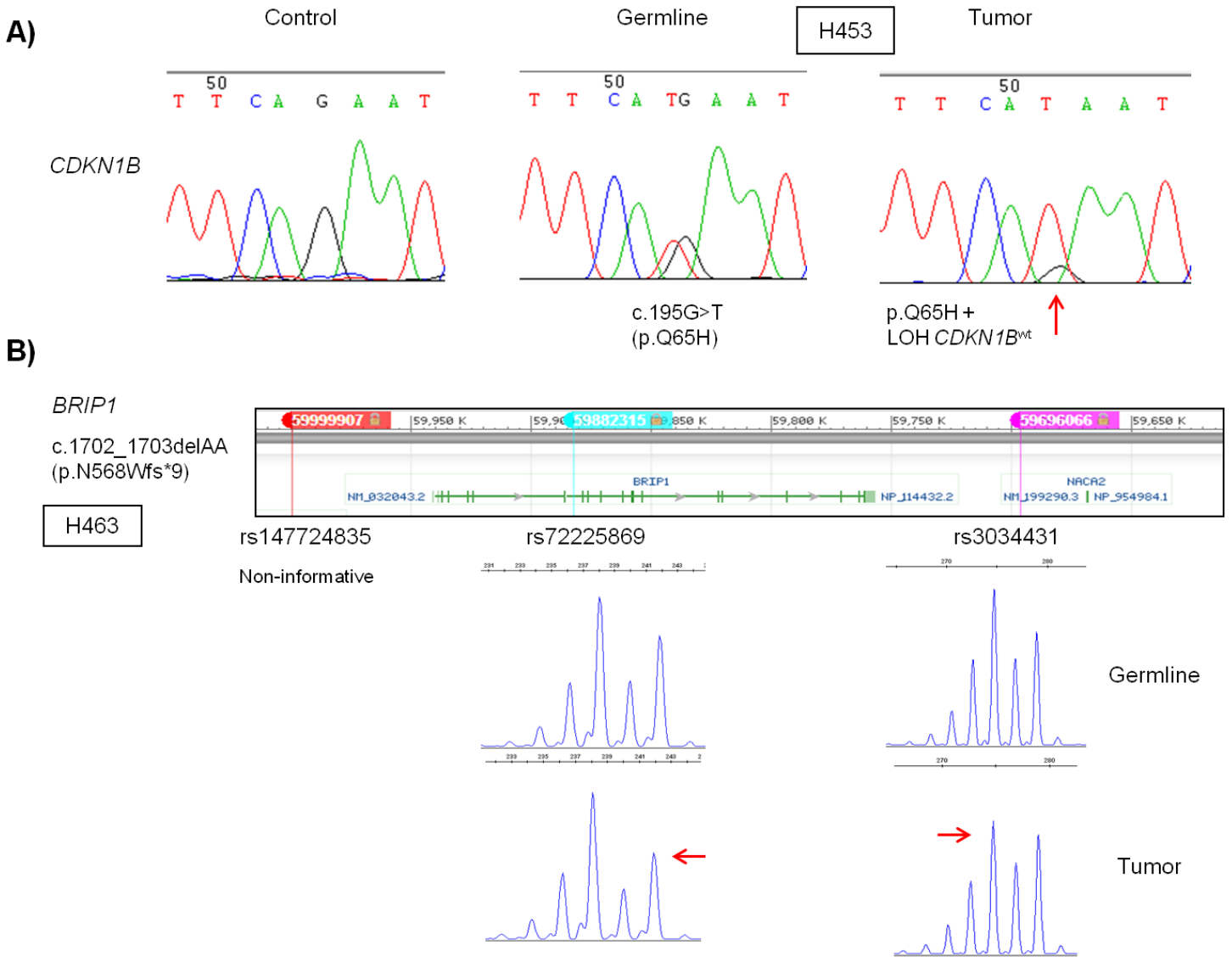
## DISCLOSURE

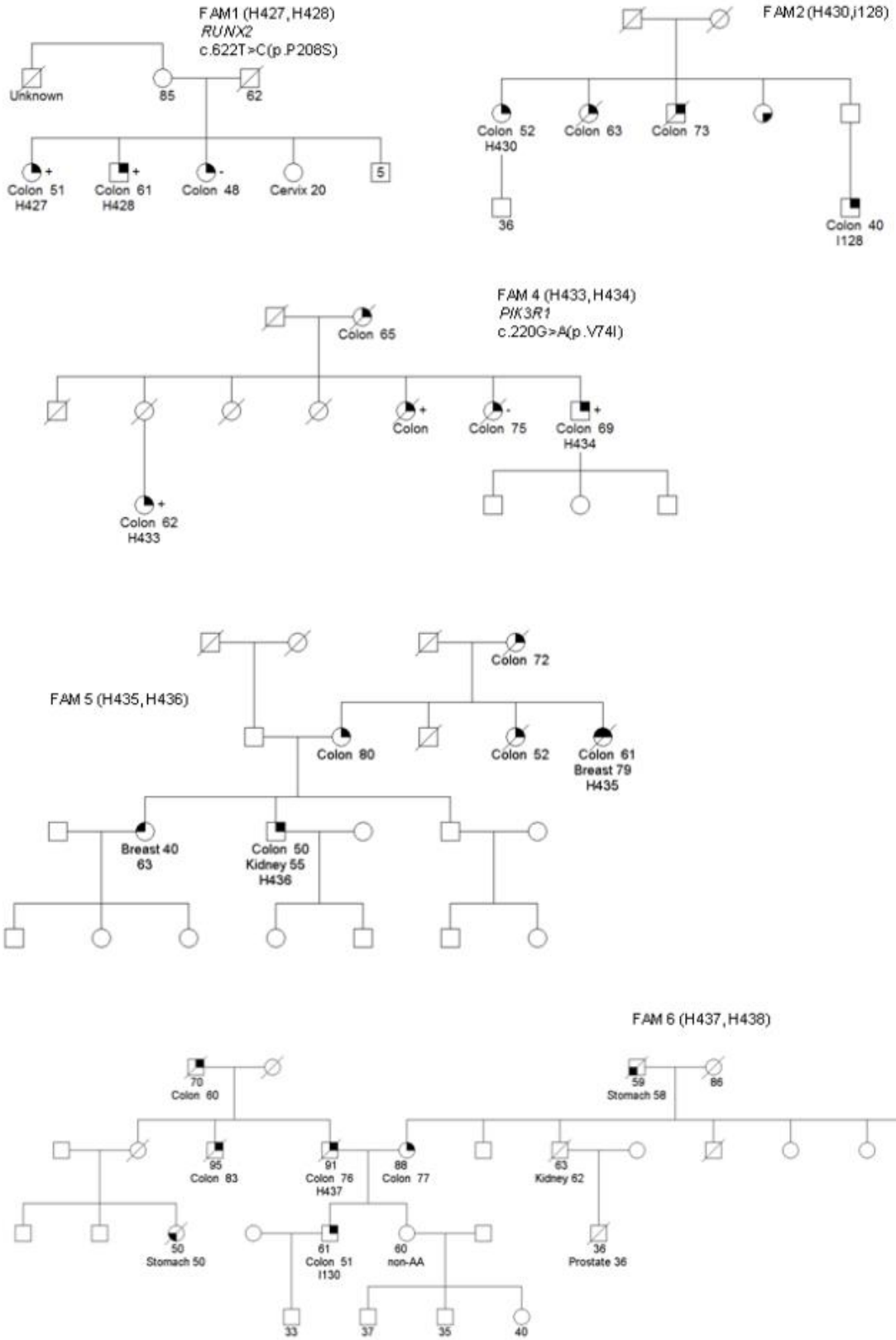
The authors declare no conflict of interest.

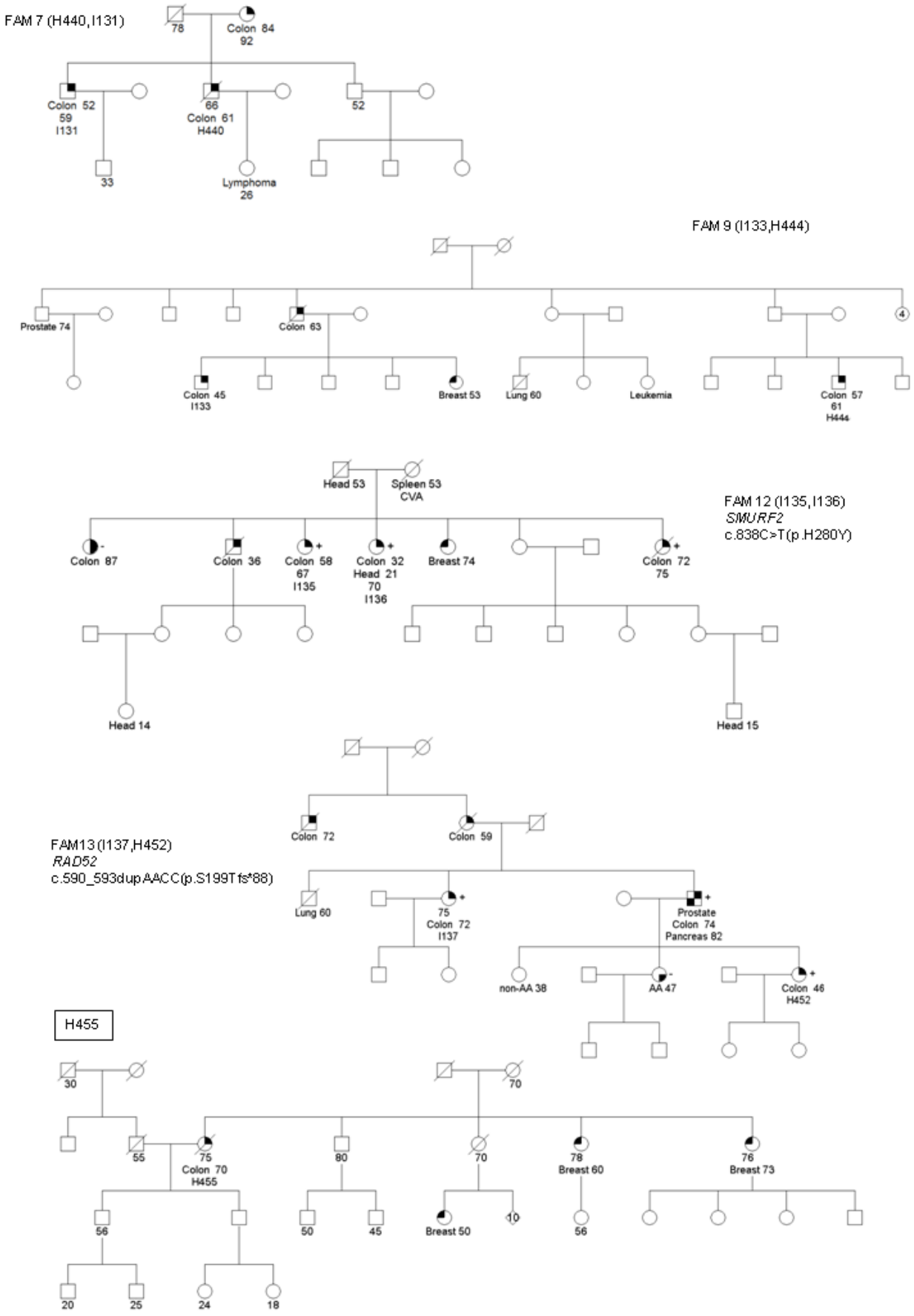
## REFERENCES

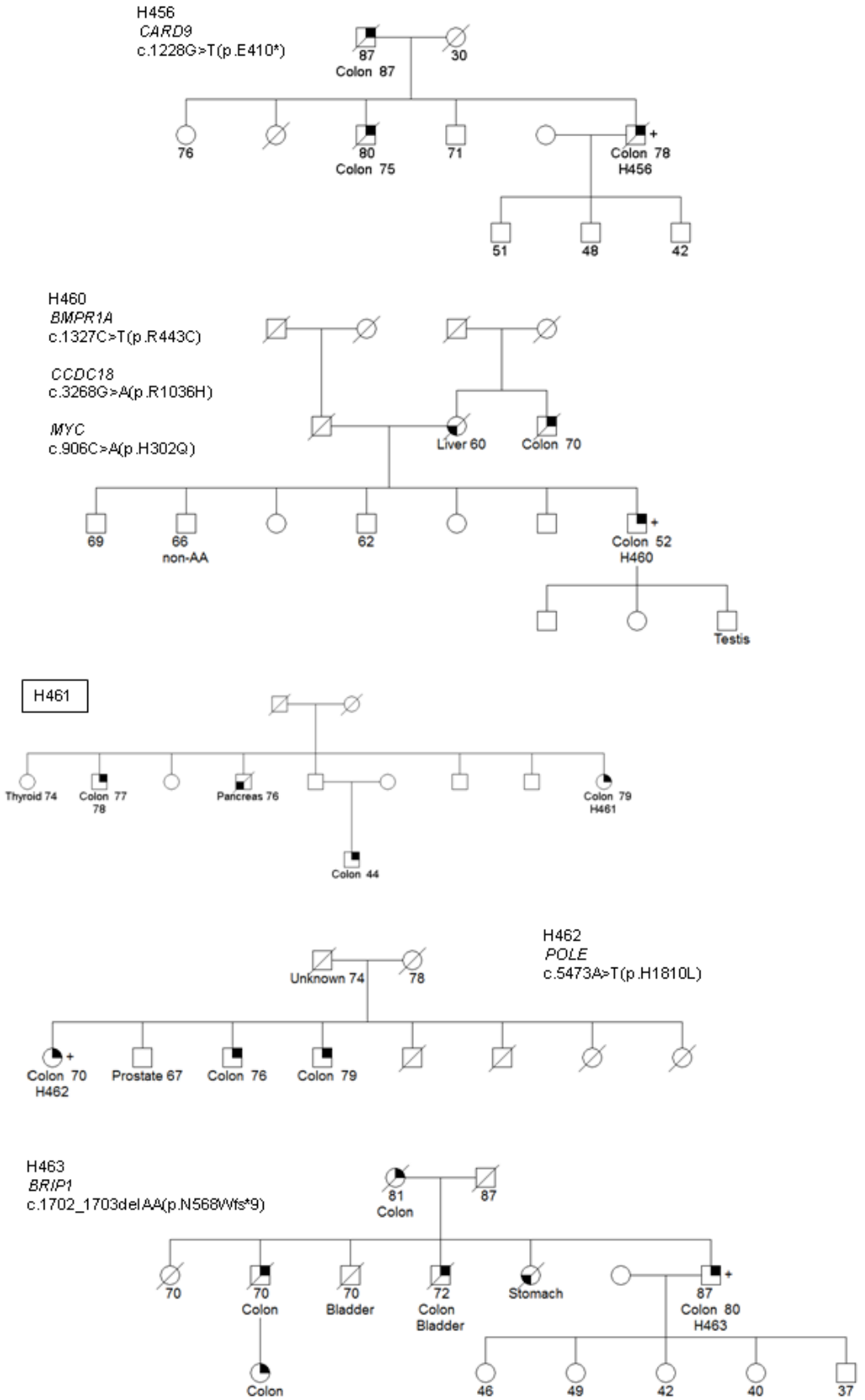
1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–2917.
2. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.
3. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010;138:2044–2058.
4. Castells A, Castellví-Bel S, Balaguer F. Concepts in familial colorectal cancer: where do we stand and what is the future? *Gastroenterology* 2009;137:404–409.
5. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA* 2005;293:1979–1985.
6. Zogopoulos G, Jorgensen C, Bacani J, et al. Germline EPHB2 receptor variants in familial colorectal cancer. *PLoS One* 2008;3:e2885.
7. Guda K, Moinova H, He J, et al. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci USA* 2009;106:12921–12925.
8. Venkatachalam R, Ligtenberg MJ, Hoogerbrugge N, et al. Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 2010;139:2221–2224.
9. Lubbe SJ, Pittman AM, Matijssen C, et al. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. *Hum Mutat* 2011;32:E1928–E1938.
10. Nieminen TT, Abdel-Rahman WM, Ristimäki A, et al. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 2011;141:e23–e26.
11. Palles C, Cazier JB, Howarth KM, et al.; CORGI Consortium; WGS500 Consortium. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013;45:136–144.
12. Smith CG, Naven M, Harris R, et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* 2013;34:1026–1034.
13. DeRycke MS, Gunawardena SR, Middha S, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev* 2013;22:1239–1251.
14. Gylfe AE, Katainen R, Kondelin J, et al. Eleven candidate susceptibility genes for common familial colorectal cancer. *PLoS Genet* 2013;9:e1003876.
15. Piñol V, Castells A, Andreu M, et al.; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005;293:1986–1994.
16. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* 2012;9:1185–1188.
17. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. *PLoS One* 2009;4:e7767.
18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
19. Li H, Handsaker B, Wysoker A, et al.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
20. Lee J, Kim SS. The function of p27 KIP1 during tumor development. *Exp Mol Med* 2009;41:765–771.
21. Russo AA, Jeffrey PD, Patten AK, Massagué J, Pavletich NP. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* 1996;382:325–331.
22. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, et al. Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *Proc Natl Acad Sci USA* 2006;103:15558–15563.

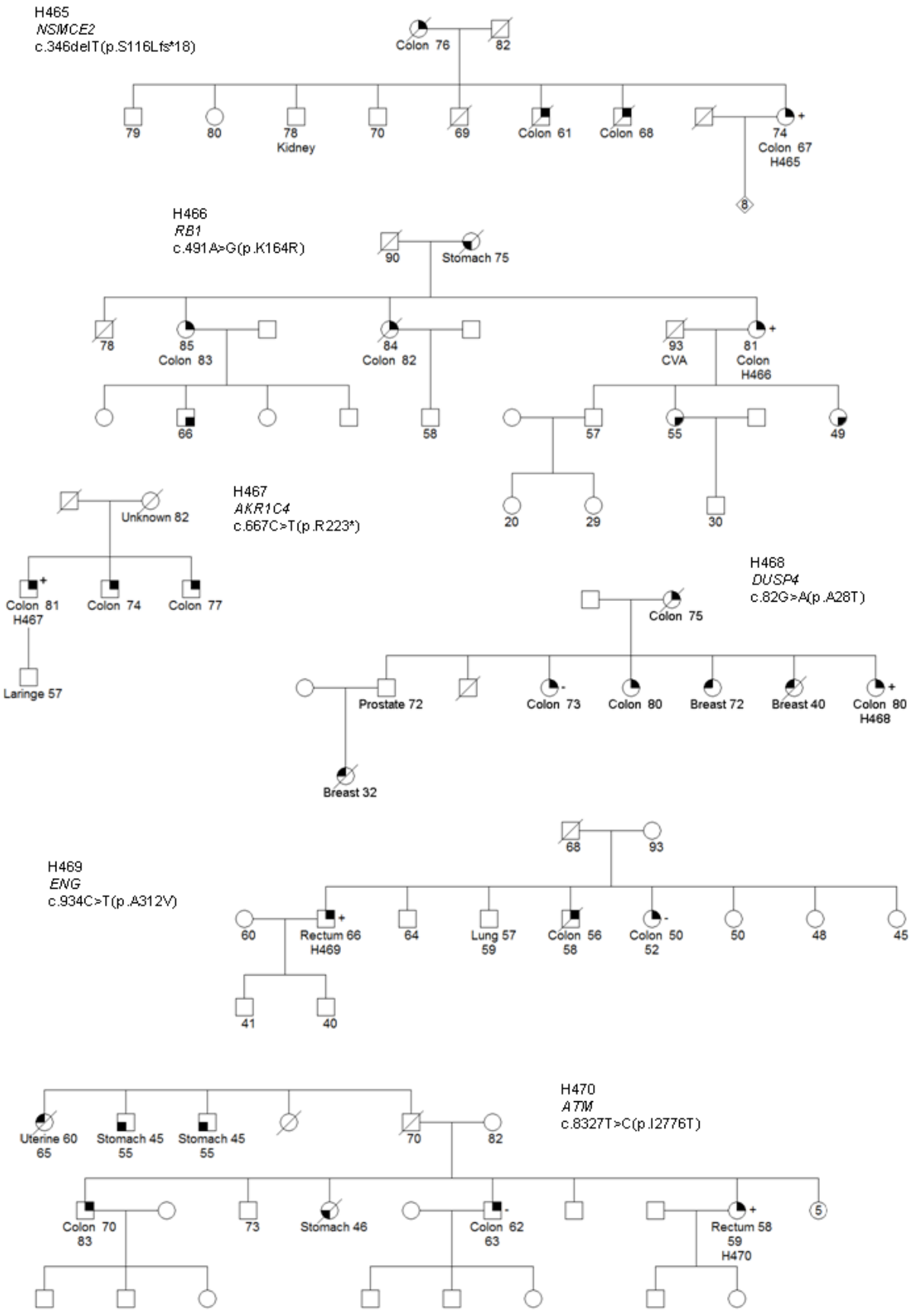
23. Chang BL, Zheng SL, Isaacs SD, et al. A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer. *Cancer Res* 2004;64:1997–1999.
24. Li Z, Otevrel T, Gao Y, et al. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* 1995;83:1079–1089.
25. Fretland AJ, Omiecinski CJ. Epoxide hydrolases: biochemistry and molecular biology. *Chem Biol Interact* 2000;129:41–59.
26. Liu F, Yuan D, Wei Y, et al. Systematic review and meta-analysis of the relationship between EPHX1 polymorphisms and colorectal cancer risk. *PLoS One* 2012;7:e43821.
27. Totzke G, Essmann F, Pohlmann S, Lindenblatt C, Jänicke RU, Schulze-Osthoff K. A novel member of the IkappaB family, human IkappaB-zeta, inhibits transactivation of p65 and its DNA binding. *J Biol Chem* 2006;281:12645–12654.
28. Cowland JB, Muta T, Borregaard N. IL-1beta-specific up-regulation of neutrophil gelatinase-associated lipocalin is controlled by IkappaB-zeta. *J Immunol* 2006;176:5559–5566.
29. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* 2011;11:481–492.
30. Schneppenheim R, Frühwald MC, Gesk S, et al. Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. *Am J Hum Genet* 2010;86:279–284.
31. Witkowski L, Carrot-Zhang J, Albrecht S, et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat Genet* 2014;46:438–443.
32. Irminger-Finger I, Jefford CE. Is there more to BARD1 than BRCA1? *Nat Rev Cancer* 2006;6:382–391.
33. Birrane G, Varma AK, Soni A, Ladias JA. Crystal structure of the BARD1 BRCT domains. *Biochemistry* 2007;46:7706–7712.
34. Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* 2003;302:636–639.
35. Ratajska M, Antoszewska E, Piskorz A, et al. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res Treat* 2012;131:89–97.
36. Sporn JC, Hothorn T, Jung B. BARD1 expression predicts outcome in colon cancer. *Clin Cancer Res* 2011;17:5451–5462.
37. Cantor SB, Bell DW, Ganesan S, et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* 2001;105:149–160.
38. Gala MK, Mizukami Y, Le LP, et al. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. *Gastroenterology* 2014;146:520–529.
39. Saunders IW, Ross J, Macrae F, et al. Evidence of linkage to chromosomes 10p15.3-p15.1, 14q24.3-q31.1 and 9q33.3-q34.3 in non-syndromic colorectal cancer families. *Eur J Hum Genet* 2012;20:91–96.
40. Cicek MS, Cunningham JM, Fridley BL, et al. Colorectal cancer linkage on chromosomes 4q21, 8q13, 12q24, and 15q22. *PLoS One* 2012;7:e38175.
41. Djureinovic T, Skoglund J, Vandrovcova J, et al. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. *Gut* 2006;55:362–366.
42. Laiho P, Hienonen T, Karhu A, et al. Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. *Oncogene* 2003;22:2206–2214.
43. Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 2014;505:302–308.
44. Ngeow J, Heald B, Rybicki LA, et al. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterol* 2013;144:1402–1409.
45. Sayed MG, Ahmed AF, Anderson ME, et al. Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis. *Ann Surg Oncol* 2002;9:901–906.
46. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007;446:153–158.


A)
Control
Germline
H453
Tumor
CDKN1B
50
T T C A G A A T
50
T T C A TG A A T
50
T T C A T A A T
c.195G>T
(p.Q65H)
p.Q65H +
LOH CDKN1B^wt
B)
BRIP1
c.1702_1703delAA
(p.N568Wfs*9)
H463
59999907
59,950 K
59882315
59,850 K
59,800 K
59,750 K
59696066
59,650 K
BRIP1
NACA2
NM_032043.2
NP_114432.2
NM_199290.3
NP_954984.1
rs147724835
rs72225869
rs3034431
Non-informative
Germline
Tumor


FAM1 (H427,H428)
RUNX2
c.622T>C(p.P208S)
Unknown
85
62
Colon 51
H427
Colon 61
H428
Colon 48
Cervix 20
5
FAM2 (H430,i128)
Colon 52
H430
Colon 63
Colon 73
36
Colon 40
I128
FAM 4 (H433,H434)
PIK3R1
c.220G>A(p.V74I)
Colon 65
Colon
Colon 75
Colon 69
H434
Colon 62
H433
FAM 5 (H435,H436)
Colon 72
Colon 80
Colon 52
Colon 61
Breast 79
H435
Breast 40
63
Colon 50
Kidney 55
H436
FAM 6 (H437,H438)
70
Colon 60
59
Stomach 58
86
95
Colon 83
91
Colon 76
H437
88
Colon 77
63
Kidney 62
50
Stomach 50
61
Colon 51
I130
60
non-AA
36
Prostate 36
33
37
35
40


FAM 7 (H440,I131)
78
Colon 84
92
Colon 52
59
I131
66
Colon 61
H440
52
33
Lymphoma
26
FAM 9 (I133,H444)
Prostate 74
Colon 63
Colon 45
I133
Breast 53
Lung 60
Leukemia
4
Colon 57
61
H444
Head 53
Spleen 53
CVA
FAM 12 (I135,I136)
SIMURF2
c.838C>T(p.H280Y)
Colon 87
Colon 36
Colon 58
67
I135
Colon 32
Head 21
70
I136
Breast 74
Colon 72
75
Head 14
Head 15
FAM13 (I137,H452)
RAD52
c.590_593dupAACC(p.S199Tfs*88)
Colon 72
Colon 59
Lung 60
75
Colon 72
I137
Prostate
Colon 74
Pancreas 82
non-AA 38
AA 47
Colon 46
H452
H455
30
55
75
Colon 70
H455
80
70
78
Breast 60
76
Breast 73
56
50
45
Breast 50
10
56
20
25
24
18


H456
CARD9
c.1228G>T(p.E410*)
87
Colon 87
30
76
80
Colon 75
71
+
Colon 78
H456
51
48
42
H460
BMPR1A
c.1327C>T(p.R443C)
CCDC18
c.3268G>A(p.R1036H)
MYC
c.906C>A(p.H302Q)
Liver 60
Colon 70
69
66
non-AA
62
+
Colon 52
H460
Testis
H461
Thyroid 74
Colon 77
78
Pancreas 76
Colon 79
H461
Colon 44
H462
POLE
c.5473A>T(p.H1810L)
Unknown 74
78
+
Colon 70
H462
Prostate 67
Colon 76
Colon 79
H463
BRIP1
c.1702_1703delAA(p.N568Wfs*9)
81
Colon
87
70
70
Colon
70
Bladder
72
Colon
Bladder
Stomach
+
87
Colon 80
H463
Colon
46
49
42
40
37


H465
NSMCE2
c.346delT(p.S116Lfs*18)
Colon 76
82
79
80
78
Kidney
70
69
Colon 61
Colon 68
74
Colon 67
H465
8
H466
RB1
c.491A>G(p.K164R)
90
Stomach 75
78
85
Colon 83
84
Colon 82
93
CVA
81
Colon
H466
66
58
57
55
49
20
29
30
H467
AKR1C4
c.667C>T(p.R223*)
Unknown 82
Colon 81
H467
Colon 74
Colon 77
Laringe 57
H468
DUSP4
c.82G>A(p.A28T)
Colon 75
Prostate 72
Colon 73
Colon 80
Breast 72
Breast 40
Colon 80
H468
Breast 32
H469
ENG
c.934C>T(p.A312V)
68
93
60
Rectum 66
H469
64
Lung 57
59
Colon 56
58
Colon 50
52
50
48
45
41
40
H470
ATM
c.8327T>C(p.I2776T)
Uterine 60
65
Stomach 45
55
Stomach 45
55
70
82
Colon 70
83
73
Stomach 46
Colon 62
63
Rectum 58
59
H470
5

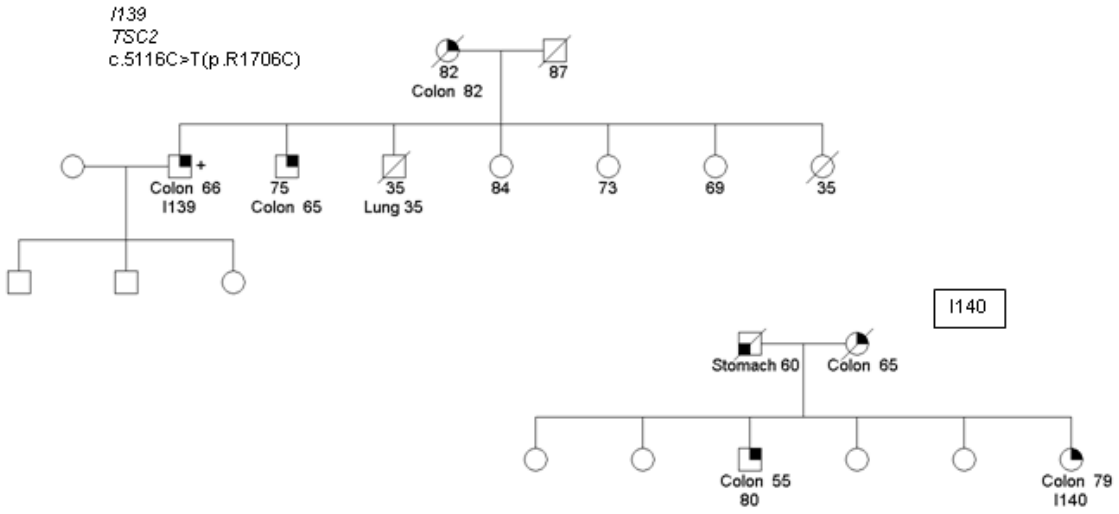

I139
TSC2
c.5116C>T(p.R1706C)
82
Colon 82
87
Colon 66
I139
75
Colon 65
35
Lung 35
84
73
69
35
I140
Stomach 60
Colon 65
Colon 55
80
Colon 79
I140

**Table S1.** Key terms used to prioritize genetic variants with functional annotations and/or bibliography related to cancer. Functional annotations listed corresponded to those selected from Gene Ontology, KEGG and REACTOME, whereas the Bibliography category referred to terms from NCBI gene summary and Gene Reference Into Function (GeneRIF). Noteworthy, other annotations containing listed key terms were also selected (e.g. “breast cancer” since it contains “cancer”).

| **Gene ontology biological process terms** |
|---|
| Apoptosis and cell death |
| Apoptotic process |
| Execution phase of apoptosis |
| Apoptotic signaling pathway |
| Mitochondrial outer membrane permeabilization |
| Release of cytochrome c from mitochondria |
| Apoptosome assembly |
| Cell death |
| Programmed cell death |
| Autophagic cell death |
| Macroautophagy |
| Microautophagy |
| Regulation of autophagy |
| Cytolysis |
| Necrotic cell death |
| Cell adhesion |
| Cell-cell adhesion |
| Homophilic cell adhesion |
| Cell-cell junction assembly |
| Cell-substrate junction assembly |
| Regulation of cell junction assembly |
| Adherens junction assembly |
| Cell adhesion mediated by integrin |
| Cell-substrate adhesion |
| Regulation of cell adhesion |
| Tight junction assembly |
| Cell cycle |
| Cell cycle |
| Cell cycle phase |
| Cell cycle phase transition |
| Mitotic cell cycle |
| Abortive mitotic cell cycle |
| Mitotic spindle organization |
| Mitotic nuclear envelope disassembly |
| Mitotic nuclear envelope reassembly |
| Establishment of mitotic spindle localization |
| Mitosis |
| Mitotic prophase |
| Mitotic metaphase |
| Mitotic anaphase |
| Mitotic telophase |
| Mitotic prometaphase |
| Mitotic sister chromatid segregation |
| Spindle assembly involved in mitosis |
| Mitotic spindle assembly checkpoint |
| Mitotic centrosome separation |
| Cell cycle arrest |
| Cell cycle checkpoint |
| DNA integrity checkpoint |
| Spindle assembly checkpoint |
| Cell differentiation |
| Cell differentiation |
| Epithelial cell differentiation |
| G1 to G0 transition involved in cell differentiation |
| Mesodermal cell differentiation |
| Stem cell differentiation |
| Cell development |
| Cell fate commitment |
| Negative regulation of cell differentiation |

| |
|---|
| Positive regulation of cell differentiation |
| Regulation of epithelial cell differentiation |
| **Cell growth** |
| Cell growth |
| Developmental cell growth |
| Regulation of cell proliferation |
| Regulation of epithelial cell proliferation |
| **Cellular response to DNA damage stimulus** |
| DNA repair |
| Single strand break repair |
| Mismatch repair |
| Postreplication repair |
| Double-strand break repair |
| DNA replication proofreading |
| Non-recombinational repair |
| Base-excision repair |
| Nucleotide-excision repair |
| UV-damage excision repair |
| Interstrand cross-link repair |
| DNA damage checkpoint |
| DNA damage response, detection of DNA damage |
| DNA damage induced protein phosphorylation |
| Replication fork processing |
| Telomere maintenance in response to DNA damage |
| Intrinsic apoptotic signaling pathway in response to DNA damage |
| Signal transduction in response to DNA damage |
| **Chromatin** |
| Chromatin organization |
| Helicase activity |
| Maintenance of DNA methylation |
| Chromosome organization |
| Kinetochore organization |
| Chromosome condensation |
| Chromosome decondensation |
| Chromosome breakage |
| Telomere organization |
| Maintenance of DNA repeat elements |
| Sister chromatid cohesion |
| Regulation of cohesin localization to chromatin |
| Centromere complex assembly |
| Regulation of chromosome organization |
| Maintenance of chromatin silencing |
| Telomere maintenance |
| Maintenance of fidelity involved in DNA-dependent DNA replication |
| DNA protection |
| Chromatin remodeling |
| **Pathways** |
| Wnt receptor signaling pathway |
| BMP signaling pathway |
| SMAD protein signal transduction |
| SMAD protein complex assembly |
| Negative regulation of MAPKKK cascade |
| Negative regulation of MAP kinase activity |
| Regulation of JUN kinase activity |
| Negative regulation of JNK cascade |
| Insulin receptor signaling pathway |
| Cellular response to TGF-beta stimulus |
| Cellular response to TGF-beta stimulus |
| Transforming growth factor beta receptor signaling pathway |
| Response to transforming growth factor beta stimulus |
| Epidermal growth factor receptor signaling pathway |
| Notch signaling pathway |
| Positive regulation of transcription of Notch receptor target |
| Notch receptor processing, ligand-dependent |
| Negative regulation of protein kinase B signaling cascade |
| Anterior/posterior pattern formation |
| Netrin-activated signaling pathway |
| Regulation of TOR signaling cascade |
| Target of rapamycin signaling pathway |
| Regulation of phosphatidylinositol 3-kinase cascade |
| NF-kappab import into nucleus |
| Regulation of NIK/NF-kappab cascade |
| Regulation of p52-dependent NF-kappab signaling |

| **KEGG pathway terms** |
|---|
| Base excision repair |
| Nucleotide excision repair |
| Mismatch repair |
| Homologous recombination |
| Non-homologous end-joining |
| Cell cycle checkpoint |
| Metabolism of xenobiotics by cytochrome P450 |
| Benzoate degradation |
| Aminobenzoate degradation |
| Fluorobenzoate degradation |
| Chloroalkane and chloroalkene degradation |
| Chlorocyclohexane and chlorobenzene degradation |
| Toluene degradation |
| Xylene degradation |
| Nitrotoluene degradation |
| Ethylbenzene degradation |

2

| |
|---|
| Styrene degradation |
| Atrazine degradation |
| Caprolactam degradation |
| DDT degradation |
| Bisphenol degradation |
| Dioxin degradation |
| Naphthalene degradation |
| Polycyclic aromatic hydrocarbon degradation |
| Furfural degradation |
| Steroid degradation |
| Metabolism of xenobiotics by cytochrome P450 |
| Focal adhesion |
| Adherens junction |
| Tight junction |
| Gap junction |
| Purine metabolism |
| Pyrimidine metabolism |
| Wnt signaling pathway |
| Notch signaling pathway |
| TGF-beta signaling pathway |
| Insulin signaling pathway |
| PI3K-Akt signaling pathway |
| Mucin type O-Glycan biosynthesis |

**Reactome terms**

| |
|---|
| DNA repair |
| Base excision repair |
| Cleavage of the damaged purine |
| Depurination |
| Removal of DNA patch containing abasic residue |
| Resolution of AP sites via the single-nucleotide replacement pathway |
| Resolution of Abasic Sites (AP sites), organism-specific biosystem |
| Gap-filling DNA repair synthesis and ligation in GG-NER |
| Gap-filling DNA repair synthesis and ligation in TC-NER |
| Transcription-coupled NER (TC-NER) |
| Repair synthesis for gap-filling by DNA polymerase in TC-NER |
| Apoptosis |
| Apoptotic execution phase |
| Apoptotic cleavage of cellular proteins |
| Mitotic anaphase |
| M phase |
| G1/S transition |
| DNA replication |
| G2/M transition |
| Integrin cell surface interactions |
| Adherens junctions interactions |
| Cell-Cell communication |
| Telomere maintenance |
| Extension of Telomeres |
| Chromosome maintenance |
| Signaling by BMP |
| Signaling by TGF beta |
| Signaling by Wnt |
| Beta-catenin phosphorylation cascade |
| Degradation of beta-catenin by the destruction complex |
| PKB-mediated events |
| TGF-beta receptor signaling |
| Constitutive PI3K/AKT Signaling in Cancer |
| Chromosome maintenance |
| O-linked glycosylation of mucins |

**Bibliography terms**

| |
|---|
| Cancer |
| Colon cancer |
| Rectal cancer |
| Tumor |
| Gastrointestinal tumor |
| Tumor suppressor |
| Oncogene |
| Carcinoma |
| Sarcoma |
| Adenoma |
| Polyposis |
| Polyps |
| Cowden |
| Inflammatory bowel disease |
| Crohn |
| Osteosarcoma |
| Glioma |
| Melanoma |
| Leukemia |
| Lymphoma |

**Table S2.** Hereditary genes and GWAS hits for colorectal cancer considered in the present study. Involvement with this disease's predisposition was seminal or very preliminary in some cases.

| Hereditary CRC genes | | |
|---|---|---|
| **Gene** | **Chromosome** | **Reference** |
| *APC* | 5q21-q22 | Groden J, et al. Cell 1991;66:589-600 |
| *MLH1* | 3p21.3 | Papadopoulos N, et al. Science 1994;263:1625-9 |
| *MSH2* | 2p21 | Fishel R, et al. Cell 1993;75:1027-38 |
| *MSH6* | 2p16 | Miyaki M, et al. Nat Genet 1997;17:271-2 |
| *PMS2* | 7p22.2 | Nicolaides N, et al. Nature 1994;371:75-80 |
| *MUTYH* | 1p34.1 | Al-Tassan N, et al. Nat Genet 2002;30:227-232 |
| *BMPR1A* | 10q22.3 | Howe J, et al. Nat Genet 2001;28:184-7 |
| *BMP4* | 14q22-q23 | Lubbe SJ, et al. Hum Mutat 2011;32:1928-38 |
| *PTPRJ* | 11p11.2 | Venkatachalam R, et al. Gastroenterology 2010;139:2221-4 |
| *GALNT12* | 9q22.33 | Guda K, et al. Proc Natl Acad Sci U S A 2009;106:12921-5 |
| *EPHB2* | 1p36.1-p35 | Kokko A, et al. BMC Cancer 2006;6:145 |
| *AXIN2* | 17q23-q24 | Lammi L, et al. Am J Hum Genet 2004;74:1043-50 |
| *UNC5C* | 4q21-q23 | Coissieux MM, et al. Gastroenterology 2011;141:2039-46 |
| *GREM1* | 15q13.3 | Jaeger E, et al. Nature Genet 2012;44:699-703 |
| *STK11* | 19p13.3 | Hemminki A, et al. Nature 1998;391:184-7 |
| *SMAD4* | 18q21.1 | Howe JR, et al. Science 1998;280:1086-8 |
| *PTEN* | 10q23.3 | Liaw D, et al. Nature Genet 1997;16:64-7 |
| *KLLN* | 10q23 | Bennett KL, et al. JAMA 2010;304:2724-31 |
| *POLE* | 12q24.3 | Palles C, et al. Nat Genet 2013;45:136-44 |
| *POLD1* | 19q13.3 | Palles C, et al. Nat Genet 2013;45:136-44 |
| *AKT1* | 14q32.32 | Orloff MS, et al. Am J Hum Genet 2013;92:76-80 |
| *PIK3CA* | 3q26.3 | Orloff MS, et al. Am J Hum Genet 2013;92:76-80 |
| *ENG* | 9q34.11 | Ngeow J, et al. Gastroenterology 2013;144:1402-9 |
| *BUB1* | 2q14 | de Voer RM, et al. Gastroenterology 2013;145:544-7 |
| *BUB3* | 10q26 | de Voer RM, et al. Gastroenterology 2013;145:544-7 |
| *PMS1* | 2q31.1 | Nicolaides N, et al. Nature 1994;371:75-80 |
| *CENPE* | 4q24-q25 | DeRycke MS, et al. CEBP 2013;22:1239-51 |
| *KIF23* | 15q23 | DeRycke MS, et al. CEBP 2013;22:1239-51 |
| *FANCM* | 14q21.2 | Smith CG, et al. Hum Mutat 2013;34:1026-34 |
| *LAMB4* | 7q31 | Smith CG, et al. Hum Mutat 2013;34:1026-34 |
| *PTCHD3* | 10p12.1 | Smith CG, et al. Hum Mutat 2013;34:1026-34 |
| *LAMC3* | 9q31-q34 | Smith CG, et al. Hum Mutat 2013;34:1026-34 |
| *TREX2* | Xq28 | Smith CG, et al. Hum Mutat 2013;34:1026-34 |
| *NOTCH3* | 19p13.2-p13.1 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *UACA* | 15q22-q24 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *SFXN4* | 10q26.11 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *TWSG1* | 18p11.3 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *PSPH* | 7p11.2 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *NUDT7* | 16q23.1 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |

| | | |
|---|---|---|
| *ZNF490* | 19p13.2 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *PRSS37* | 7q34 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *CCDC18* | 1p22.1 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *PRADC1* | 2p13.2 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *MRPL3* | 3q21-q23 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *AKR1C4* | 10p15.1 | Gylfe AE, et al. PLoS Genet 2013;9: e1003876 |
| *RNF43* | 17q22 | Gala MK, et al. Gastroenterology 2014;146:520-9 |
| *ATM* | 11q22-q23 | Gala MK, et al. Gastroenterology 2014;146:520-9 |
| *PIF1* | 15q22.31 | Gala MK, et al. Gastroenterology 2014;146:520-9 |
| *TELO2* | 16p13.3 | Gala MK, et al. Gastroenterology 2014;146:520-9 |
| *XAF1* | 17p13.1 | Gala MK, et al. Gastroenterology 2014;146:520-9 |
| *RBL1* | 20q11.2 | Gala MK, et al. Gastroenterology 2014;146:520-9 |
| **GWAS** | | |
| **Gene** | **Chromosome** | **Reference** |
| *MYC* | 8q24.21 | Tomlinson IP, et al. Nat Genet 2007;39:984-8 |
| *SMAD7* | 18q21.1 | Broderick P, et al. Nat Genet 2007;39:1315-7 |
| *EIF3H* | 8q23.3 | Tomlinson IP, Nat Genet 2008;40:623-30 |
| *GREM1* | 15q13.3 | Jaeger E, et al. Nat Genet 2008;40:26-8 |
| *C11orf93* | 11q23 | Tenesa A, et al. Nat Genet 2008;40:631-7 |
| *BMP4* | 14q22.2 | Houlston RS, et al. Nat Genet 2008;40:1426-35 |
| *CDH1* | 16q22.1 | Houlston RS, et al. Nat Genet 2008;40:1426-35 |
| *RHPN2* | 19q13 | Houlston RS, et al. Nat Genet 2008;40:1426-35 |
| *DUSP10* | 1q41 | Houlston RS, et al. Nat Genet 2010 Nov;42:973-7 |
| *DIP2B* | 12q13.13 | Houlston RS, et al. Nat Genet 2010 Nov;42:973-7 |
| *LAMA5* | 20q13.33 | Houlston RS, et al. Nat Genet 2010 Nov;42:973-7 |
| *BMP2* | 20p12.3 | Houlston RS, et al. Nat Genet 2010 Nov;42:973-7 |
| *TERC* | 3q26.2 | Jones AM, et al. Gut 2012;61:248-54. |
| *TERT* | 5p15.33 | Kinnersley B, et al. Br J Cancer 2012;107:1001-8 |
| *CDKN1A* | 6p21 | Dunlop MG, et al. Nat Genet 2012;44:770-6 |
| POLD3 | 11q13.4 | Dunlop MG, et al. Nat Genet 2012;44:770-6 |
| *SHROOM2* | Xp22.2 | Dunlop MG, et al. Nat Genet 2012;44:770-6 |
| *SLC5A9* | 1p33 | Fernandez-Rozadilla C, et al. BMC Genomics 2013 14:55 |
| *DUSP4* | 8p12 | Fernandez-Rozadilla C, et al. BMC Genomics 2013 14:55 |
| *CCND2* | 12p13.32 | Jia WH, et al. Nat Genet 2013;45:191-6 |
| *PITX1* | 5q31.1 | Jia WH, et al. Nat Genet 2013;45:191-6 |
| *HAO1* | 20p12.3 | Jia WH, et al. Nat Genet 2013;45:191-6 |
| *NABP1* | 2q32.3 | Peters U, et al. Gastroenterology 2013;144:799-807 |
| *LAMC1* | 1q25.3 | Peters U, et al. Gastroenterology 2013;144:799-807 |
| *TBX3* | 12q24.21 | Peters U, et al. Gastroenterology 2013;144:799-807 |

**Table S3.** PCR primers used for Sanger sequencing and tumor loss of heterozigosity (LOH) studies.

| Gene | Technique | Forward primer sequence 5' -3' | Reverse primer sequence 5' -3' | PCR size |
|---|---|---|---|---|
| *CDKN1B* | Sanger/LOH | AAGAGTTAACCCGGGACTTG | CTGACATCCTGGCTCTCCT | 214 |
| *XRCC4* | Sanger | ACTGAGGAAACAGGCTACATTT | CCTGAACATCCACATCCAGA | 642 |
| | LOH | TGCTAAGCAGAACTGCATGACT | TTCTCGTTCTTGAGCTGCATT | 185 |
| | STR 1 | TTCCAGCTTAGCTAACATATTGTAGAG | GGCTTCATAAGTGCAGCAAA | 245 |
| | STR 2 | TGCACAGATGTGAGCTGTTG | ACGTTCCAGACACAGTGCAA | 189 |
| *EPHX1* | Sanger/LOH | TGGCTTCAACTCCAACTACC | TGTGGAAGAAGGCTGTTCTC | 177 |
| | STR 1 | GGGTCCCTGATCTGCACTAA | CAGGTCACACTCCCTCCTCT | 167 |
| | STR 2 | TTGCTCAGCTCACATCCTTG | TCAACTCAAATGATCCCTACCA | 213 |
| | STR 3 | TCCCAAAAACCATGTCTTCC | GGCAGAGCGAGACTCAAAAA | 231 |
| *NFKBIZ* | Sanger/LOH | CGACGTATCCTGAAGGGAAA | TGCAAATCTGCCAAGAGAAA | 248 |
| *SMARCA4* | Sanger/LOH | ATGCATGAGAAGGGCATGT | GGTACTCCCTGAATCCATCC | 222 |
| *BRCA2* | Sanger/LOH | ATGGCGTTTCTAAACATTGC | AATTCTTCTTTTCCAGCCTTTC | 172 |
| *BARD1* | Sanger | TGTGTTTGCCAATATGGCTTT | ACAGGGCTTCACCGTGTTAG | 574 |
| | LOH | TTGATGGCCAGGTTAGAGAA | TTGAACTGCATCACCAGGAA | 216 |
| | STR 1 | CGCTGCAAATGCTCTTCAT | ACATACTCGCACACCCATGT | 214 |
| | STR 2 | AACCCGGGAAGTAAAGGTTG | GGCAGCCAACAGCATTCTA | 256 |
| | STR 3 | GGTCAGGTCTCAGCATGGTT | CCTCATTCCCCATTCAACAC | 237 |
| *BRIP1* | Sanger | TGGGAAGGTACCAGCTCTTT | CCTGCCTCAAAACAACAACA | 404 |
| | LOH | TGGGAAGGTACCAGCTCTTT | TTTTCTGTCGTGAACGTTTCTT | 239 |
| | STR 1 | GCTGATAGCTTGCTTTCAAAATG | CTGCTTCAAATCTGGGGAAA | 282 |
| | STR 2 | ACACGGATTCTCCTTTTGGA | CGCACTGGCTCATGCTTAT | 247 |
| | STR 3 | TGTGCAGAATTTCCAAAAAGC | CCACACTGGGAGACAGAATG | 199 |
| *RB1* | Sanger/LOH | CCAGTACCAAAGTTGATAATGCTA | TGAGCTAACATTAAAAGGGACAA | 220 |
| *AKR1C4* | Sanger/LOH | CCTTACCTCAACCAGAGCAA | TTACCCTGACCCATTGAGAA | 202 |
| *CARD9* | Sanger | GTCCGCAGAGCACACAAAT | CAGGAGGGGTTTGGTTAGGT | 483 |
| | LOH | AGCTGCAGGTGTTCCAGTGT | CAGGAGGGGTTTGGTTAGGT | 218 |
| *NSMCE2* | Sanger | GGATCATTGAAGGGCCACTA | AACTGGGGCTGTTCATCAAG | 476 |
| | LOH | GGTGAAAGAAGAACGTCCAGA | TTCAAAAAGCCGAACCAAAG | 193 |
| *BMPR1A* | Sanger | GCCCCAAGGAGAAAAAGAAG | TATAGCACAACGGGGAGGAC | 576 |
| | LOH | AGCCCTACATCATGGCTGAC | TGTGCCCACAAAAATTAAAAAC | 176 |
| *CCDC18* | Sanger/LOH | CAATTACAAAGCCTATTCTAAGCA | AGTTCCAATTCTGATTTCTCCA | 176 |
| *MYC* | Sanger/LOH | TCGATGTTGTTTCTGTGGAA | AGGATAGTCCTTCCGAGTGG | 176 |
| *POLE* | Sanger/LOH | GTGCACCTTCTCTTTCAGGA | TTCATCATGTTGTGGAGTGTG | 177 |
| *TSC2* | Sanger/LOH | AAGTCTCCCCAGACATGGA | GGTAGATATCGGTGGGGTTG | 240 |
| *RAD52* | Sanger | ACAGGACCTAGAAGGCGTGA | CAGAGGAAAGGAGGGGACTT | 345 |
| | LOH | TTTTCCAGGAATCGCAATCT | GGTCGGCAGCTGTTGTATCT | 175 |
| *BMP4* | Sanger | AGCAGCCAAACTATGGGCTA | GGGAACGTGTGTGTGTGG | 621 |
| | LOH | AAGCGTAGCCCTAAGCATCA | TGGTTGAGTTGAGGTGGTCA | 187 |

| | | | | |
|---|---|---|---|---|
| *DUSP4* | Sanger | AGTGTGCTCAAAAGGCTGAT | ACCGAACCTAGGATGTAGCC | 158 |
| *LAMA5* | Sanger | AGTGTCCTTACGGGCCTTTC | CCCTCCATATCCACGAACTG | 422 |
| *CETN2* | Sanger | TTCCTCCACCAAGAGGATTG | CAGGCCCAGTAGGTTGGTTA | 463 |
| *MAP9* | Sanger | TGGGCTTCATCTGGACTCTT | CACACACGCACACAAATTCA | 463 |
| *ENG* | Sanger | CTTCAAGCTCCCAGACACAC | AGCTCACACAGAGGTGCTTC | 211 |
| *RUNX2* | Sanger | TTGGAAATCTATGCTGCTGTG | TCCCGAGGTCCATCTACTGT | 176 |
| *PIK3R1* | Sanger | CCTGAAGAAATTGGCTGGTT | CAGTTTTCGAAGAACCTGGTG | 166 |
| *SMURF2* | Sanger | AGAACAGAGGACAACGCAAC | GTGAAATGGCTCTTTGATGG | 136 |
| *ATM* | Sanger | CTCAGCGAAGTGGTGTTCTT | AATTTTGGGTGTCACTCACC | 156 |
| *DHX9* | Sanger | AATCTCAGAATTACCCAGGTCAG | CAAGTGCTTCGCCATAGGTA | 218 |

LOH, loss of heterozigosity; PCR, polymerase chain reaction; STR, short tandem repeat.

**ARTÍCULO 2**

**The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer**

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

EJHG Open

European Journal of Human Genetics (2016), 1–5

www.nature.com/ejhg

SHORT REPORT

# The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer

Clara Esteban-Jurado[1], Sebastià Franch-Expósito[1], Jenifer Muñoz[1], Teresa Ocaña[1], Sabela Carballal[1], Maria López-Cerón[1], Miriam Cuatrecasas[2], Maria Vila-Casadesús[3], Juan José Lozano[3], Enric Serra[4], Sergi Beltran[4], The EPICOLON Consortium, Alejandro Brea-Fernández[5], Clara Ruiz-Ponte[5], Antoni Castells[1], Luis Bujanda[6], Pilar Garre[7], Trinidad Caldés[7], Joaquín Cubiella[8], Francesc Balaguer[1] and Sergi Castellví-Bel*[,1]

**Colorectal cancer (CRC) is one of the most common neoplasms in the world. Fanconi anemia (FA) is a very rare genetic disease causing bone marrow failure, congenital growth abnormalities and cancer predisposition. The comprehensive FA DNA damage repair pathway requires the collaboration of 53 proteins and it is necessary to restore genome integrity by efficiently repairing damaged DNA. A link between FA genes in breast and ovarian cancer germline predisposition has been previously suggested. We selected 74 CRC patients from 40 unrelated Spanish families with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance and without mutations in known hereditary CRC genes and performed germline DNA whole-exome sequencing with the aim of finding new candidate germline predisposition variants. After sequencing and data analysis, variant prioritization selected only those very rare alterations, producing a putative loss of function and located in genes with a role compatible with cancer. We detected an enrichment for variants in FA DNA damage repair pathway genes in our familial CRC cohort as 6 families carried heterozygous, rare, potentially pathogenic variants located in *BRCA2*/*FANCD1*, *BRIP1*/*FANCJ*, *FANCC*, *FANCE* and *REV3L*/*POLZ*. In conclusion, the FA DNA damage repair pathway may play an important role in the inherited predisposition to CRC.**
*European Journal of Human Genetics* (2016) **00,** 1–5. doi:10.1038/ejhg.2016.44

## INTRODUCTION

Colorectal cancer (CRC) is the third most frequent neoplasm in the world and its average lifetime risk in the general population is ~5%.[1] There is some degree of familial aggregation in up to 35% of CRC patients, but the majority of the underlying germline predisposition factors remain still unidentified. The Mendelian CRC syndromes, with Lynch syndrome and familial adenomatous polyposis being the most common, correspond to 5% of total CRC cases and are mainly due to germline mutations in *APC*, *MUTYH* and the mismatch repair genes (ie, *MLH1*, *MSH2*, *MSH6*, *PMS2*). Recently, next-generation sequencing efforts in familial CRC have identified additional causative germline mutations in genes such as *POLE*, *POLD1* and *NTHL1*.[2–4]

Fanconi anemia (FA) is a very rare genetic disease with an incidence of 1–3 per 500 000 births and it causes bone marrow failure, congenital growth abnormalities and cancer predisposition. FA patients have chromosome fragility and hypersensitivity to drugs that induce DNA interstrand crosslinks (ICLs).[5] It corresponds to an autosomal recessive condition and it has been associated with germline mutations in 18 FA genes. Among them, monoallelic mutations in *FANCD1/BRCA2*, *FANCJ/BRIP1*, *FANCN/PALB2* and *FANCC* have also been linked to breast and ovarian cancer genetic predisposition.[6,7] The comprehensive FA DNA damage repair pathway requires the collaboration of 53 proteins and it is necessary to restore genome integrity by efficiently repairing damaged DNA, especially ICLs (Figure 1). ICLs affect both DNA strands impeding transcription and replication-fork progression and also complicating correct DNA repair as there is no unaffected template available.[8] Besides the link between FA genes and breast and ovarian cancer, some other genes not contributing to FA but part of the FA DNA damage repair pathway have additionally been involved in the same cancer predisposition and include *BRCA1*, *RAD51C* and *FANCM*.[9] Very recently, mutations in some FA DNA damage repair pathway genes have also been postulated to be the germline triggers in familial CRC cases, including *BRCA2*,[10,11] *FAN1*[12] and *BLM*.[13]

## MATERIALS AND METHODS

We selected 74 CRC probands from 40 unrelated Spanish families (4 affected relatives from 1 family, 3 affected relatives from 8 families, 2 affected relatives

[1]Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain; [2]Department of Pathology, Hospital Clinic, Barcelona, Spain; [3]Bioinformatics Platform, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Barcelona, Spain; [4]Centre Nacional d'Anàlisi Genòmica (CNAG), Parc Científic de Barcelona, Barcelona, Spain; [5]Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, University of Santiago de Compostela, Santiago de Compostela, Spain; [6]Gastroenterology Department, Hospital Donostia–Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Country University (UPV/EHU), San Sebastián, Spain; [7]Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain; [8]Gastroenterology Department, Complexo Hospitalario Universitario de Ourense, Instituto de Investigación Biomédica Ourense, Pontevedra y Vigo, Ourense, Spain
*Correspondence: Dr S Castellví-Bel, Genetic Predisposition to Colorectal Cancer Group, IDIBAPS, Centre Esther Koplowitz (CEK), Rosselló 153 planta 4, 08036 Barcelona, Catalonia, Spain. Tel: +34 93 2275400 ext. 4183/2915; E-mail: sbel@clinic.ub.es
Received 16 November 2015; revised 29 March 2016; accepted 7 April 2016

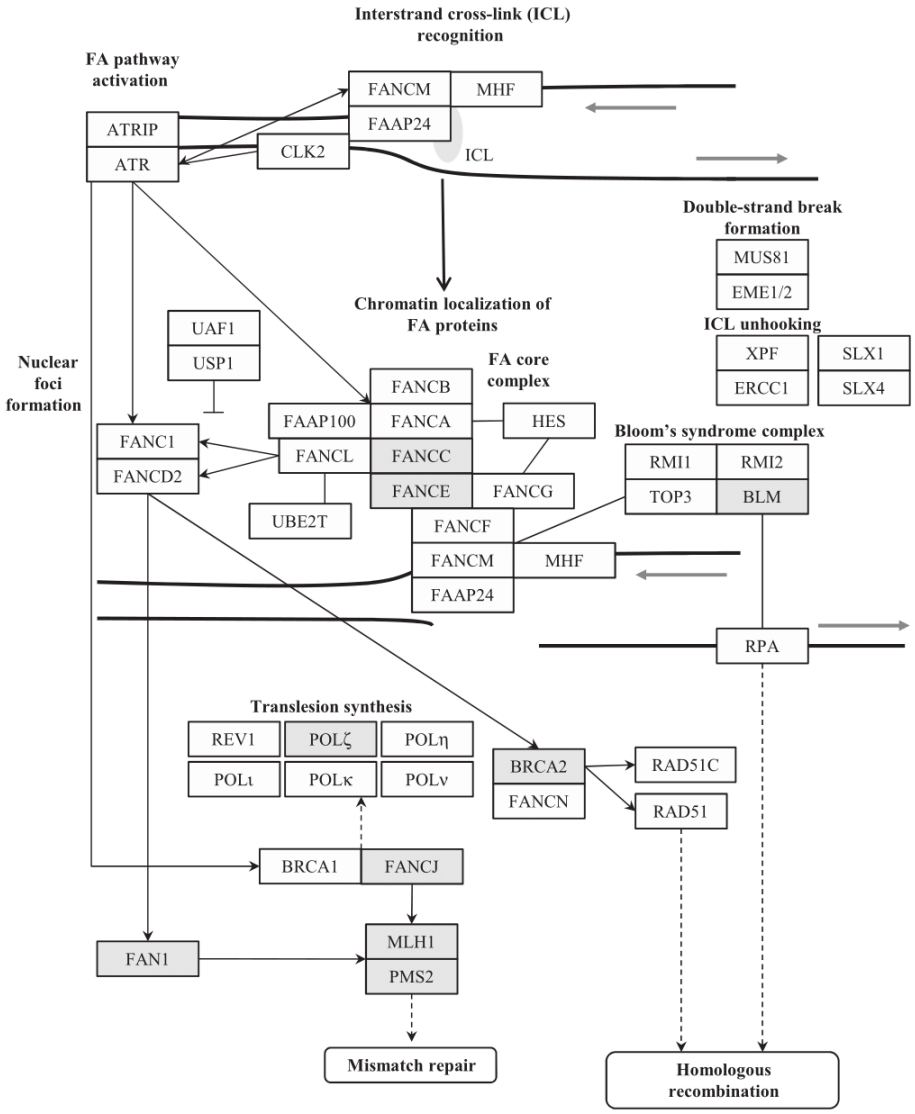



**Figure 1** The Fanconi anemia (FA) DNA damage repair pathway. Genes linked to colorectal cancer (CRC) predisposition by the present report and previous evidence are shaded in gray (adapted from KEGG database, http://www.genome.jp/kegg/pathway.html).

from 15 families and 16 CRC unrelated patients) with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance and without point mutations or large rearrangements in the most common hereditary CRC genes (*APC*, *MUTYH* and mismatch repair genes). Families were selected based on the following criteria: $\geq 3$ relatives with CRC, $\geq 2$ consecutive affected generations and at least one CRC diagnosed before the age of 60 years. This study was approved by the institutional ethics committee of each participating hospital and written informed consent was obtained at CRC diagnosis.

Sequencing, raw data analysis and variant filtering was performed as previously described for a subset of 42 patients.[14] In this regard, it should be noted that this previous cohort was completed with 31 additional new CRC patients, corresponding to 11 new families and 5 new cases in previously analyzed families, totaling 74 CRC probands from 40 families. Briefly, germline DNA whole-exome sequencing (WES) used the HiSeq2000 platform (Illumina, San Diego, CA, USA) and SureSelectXT Human All Exon for exon enrichment V4 (Agilent, Santa Clara, CA, USA). Mean coverage was $>95\times$ in all samples and 51 Megabases was the target size that required $\sim 4$ Gigabytes of sequencing per sample. After sequencing and data analysis, variant prioritization selected only those very rare alterations (0–0.1%), shared by individuals sequenced from the same family, producing a putative loss of function and located in genes with a role compatible with cancer (Supplementary Table 1). Variants were validated by Sanger sequencing (GATC Biotech, Germany) and segregation analysis of the prioritized variants was performed in additional affected family members (CRC and advanced adenoma) when constitutive DNA was available. Genetic variants have been submitted to the ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/; accession numbers SCV000262600, SCV000262601, SCV000262602, SCV000262603, SCV000262604 and SCV000262605). In addition, somatic loss of heterozygosity (LOH) was studied by Sanger sequencing or microsatellites in tumor DNA of patients (one per family) carrying the selected variants when possible. DNA was extracted from a percentage of tumor cells of 70–80% in most cases.

## RESULTS AND DISCUSSION

The aim of our study was to find candidate germline predisposition variants by performing exome sequencing in a cohort of familial CRC

patients compatible with an autosomal dominant inheritance and without an alteration in the previously known hereditary CRC genes in order to facilitate genetic counseling and correctly address prevention strategies. Our preliminary results for a subset of CRC patients were previously published[14] but additional family members and new families were whole-exome sequenced more recently and the present report corresponds to the analysis of the complete cohort of 74 samples.

Interestingly, our data revealed heterozygous, rare, potentially pathogenic variants in six families located in genes belonging to the FA DNA damage repair pathway including *BRCA2*/*FANCD1*, *BRIP1*/*FANCJ*, *FANCC*, *FANCE* and *REV3L*/*POLZ* after data analysis and variant prioritization (Table 1 and Figure 2). All six prioritized variants were validated by Sanger sequencing (Supplementary Figure 1) and segregation analysis was correct in family members (CRC and advanced adenoma). It is also interesting to point out that among the 1–57 variants identified by WES that remained after filtering in each family, the genetic variants finally prioritized corresponded to the best candidate for genetic predisposition to CRC (not reported or with a very low frequency $<0.01\%$ in external control exome data sets including a Spanish database, potential loss-of-function variant affecting residues highly conserved in evolution, previous implication with cancer predisposition and correct segregation). It is important to highlight that three of the variants corresponded to frameshift alterations and three to missense variants. The former are expected to truncate the protein and, therefore, are likely pathogenic, whereas for the latter pathogenicity is plausible but has not been proven with functional studies. On the other hand, the identified missense variants comply with several criteria to be potentially deleterious and two of them (those in *BRCA2* and *FANCE*) that fall in protein domains were pathogenic mutations that have been previously reported in FA or familial breast and ovarian cancer patients in the ClinVar database or the FA mutation database (http://www.rockefeller.edu/fanconi/). In addition, when comparing with a publically available genetic variation control data set (Exome Aggregation Consortium), we detected a clear enrichment for putative loss-of-function variants in these FA DNA damage repair pathway genes in our familial CRC cohort, when considering nonsense, canonical splice site, frameshift and missense variants with a likely pathogenicity prediction (CADD $>15$, Combined Annotation Dependent Depletion, CADD, http://cadd.gs.washington.edu/) and a genotype frequency $<0.01\%$ (6/40, 15% *vs* 2546/60 706, 4.2%; $\chi^2$ test, uncorrected for multiple testing $P$-value $=0.0025$), or only nonsense, canonical splice site and frameshift with a genotype frequency $<0.01\%$ (3/40, 7.5% *vs* 306-/60 706, 0.5%; $\chi^2$ test, uncorrected for multiple testing $P$-value $<0.0001$). It is also worth mentioning that in 5 of the 6 families carrying the reported variants, other neoplasms besides CRC were present with an age of onset $<60$ years, including breast cancer, endometrial cancer, prostate cancer, lung cancer, leukemia and gastric cancer. No relevant clinical or histopathological characteristics were detected among variant carriers, although a small sample size could be precluding the detection of such correlation.

Taking into account our results and previous evidence,[10–13] our report draws the attention to the fact that the FA DNA damage

**Table 1 Description of the 6 genetic variants belonging to the Fanconi anemia DNA damage repair pathway detected in a cohort of 40 Spanish colorectal cancer families**

| *Family* | *Gene* | *RefSeq* | *Genetic variant* | *Genotype frequency (ExAC, EVS, CSVS)* | *In silico* | *Family phenotype (age <60)* | *Cancer-AA carriers* | *LOH* | *Domain/region* | *OMIM* |
|---|---|---|---|---|---|---|---|---|---|---|
| FAM6 | *BRCA2/FANCD1* | NM_000059.3 | c.7759C>T p.(L2587F) | 5/60676<br>0/6503<br>0/572 | 5/5 | CRC, gastric | 2/2 | INC[a] | Interaction with DSS1 | Breast and ovarian cancer, FA |
| FAM20 | *BRCA2/FANCD1* | NM_000059.3 | c.4963delT p.(Y1655fs*15) | 0/60706<br>0/6503<br>0/572 | FS | CRC, breast cancer, endometrial cancer, prostate cancer, lung cancer, leukemia | 2/2 | INC[a] | — | Breast and ovarian cancer, FA |
| H463 | *BRIP1/FANCJ* | NM_032043.2 | c.1702_1703delAA p.(N568fs*9)[b] | 0/60706<br>0/6503<br>0/572 | FS | CRC, gastric cancer | 1/1 | INC[a] | — | Breast and ovarian cancer, Spanish FA J family[c] |
| FAMN4 | *FANCC* | NM_000136.2 | c.591_592dupC p.(L199fs*12) | 0/60706<br>0/6503<br>0/572 | FS | CRC, adenomas | 2/2 | N | — | Breast cancer, pancreatic cancer, FA |
| FAM40 | *FANCE* | NM_021922.2 | c.598C>T p.(R200C) | 0/60706<br>1/6503<br>0/572 | 5/5 | CRC, breast cancer | 2/2 | NA | Interaction with FANCC | FA, esophageal and gastric cancer |
| FAM3 | *REV3L/POLZ* | NM_002912.4 | c.559A>T p.(R187W) | 0/60706<br>0/6503<br>0/572 | 5/5 | CRC, prostate cancer, adenomas | 4/4 | Y | — | Lung cancer, chromosomal instability |

Abbreviations: AA, advanced adenoma; CRC, colorectal cancer; FA, Fanconi anemia; FS, frameshift; INC, inconclusive; LD, linkage disequilibrium; LOH, loss of heterozygosity; N, no; NA, not available; RefSeq, reference sequence; Seg, segregation; Y, yes.
Genotype frequency: presence or absence in external control exome data sets (ExAC (exome aggregation consortium), EVS (exome variant server) and CSVS (CIBERER Spanish variant server)).
*In silico*: number of deleterious predictions by bioinformatics tools used (CADD, PolyPhen, SIFT, PhyloP and LRT).
Cancer-AA carriers: number of cancer/advanced adenoma cases within the family that carry the variant.
LOH: depletion of the wild-type allele in tumor DNA in comparison with the germline.
Domain/region: the protein domain or region where the variant is located.
OMIM: OMIM database information including previous hereditary cancer involvement.
[a]Minimal LOH was observed.
[b]Previously reported in Esteban-Jurado *et al*.[14]
[c]Same variant in our CRC cohort and in the Spanish FA J family.

repair pathway may play an important role in the inherited predisposition to CRC. It is also important to highlight that pleiotropy is becoming important in germline predisposition to cancer as a higher number of genes may be involved in the genetic predisposition to a broader spectrum of neoplasms, as evidenced by our results and others.[11] Finally, it could be hypothesized that defects in the FA DNA

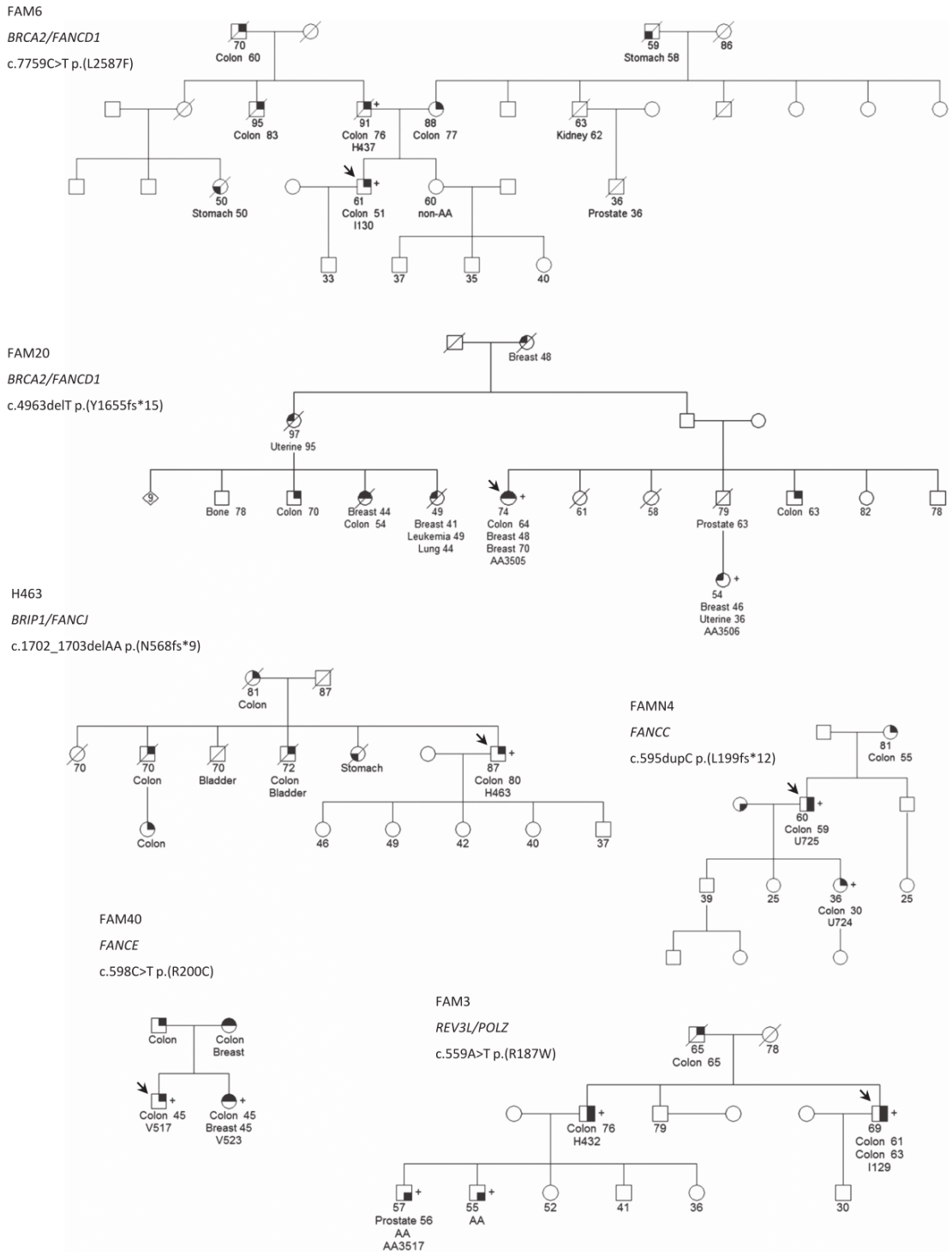



**Figure 2** Pedigrees from families FAM3, FAMN4, FAM6, FAM20, FAM40 and H463 are shown. Filled symbol indicate affected for CRC (upper right quarter) adenoma/s (lower right quarter), stomach cancer (lower left quarter) and breast cancer (upper left quarter). Colon, breast, stomach, lung, prostate, kidney, uterine, leukemia and bladder refer to the type of cancer. AA/non-AA, advanced adenoma/nonadvanced adenoma; +, variant carrier. Index cases are indicated with an arrow.

damage repair pathway would affect correct homologous recombination and contribute to genome instability. However, the contribution to CRC predisposition of genetic variants in this pathway needs further investigation and collaborative efforts should be made in order to fully characterize it. If this involvement is further confirmed in additional familial CRC cohorts, it would become very relevant regarding the molecular genetic diagnosis for the hereditary forms of this neoplasm.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

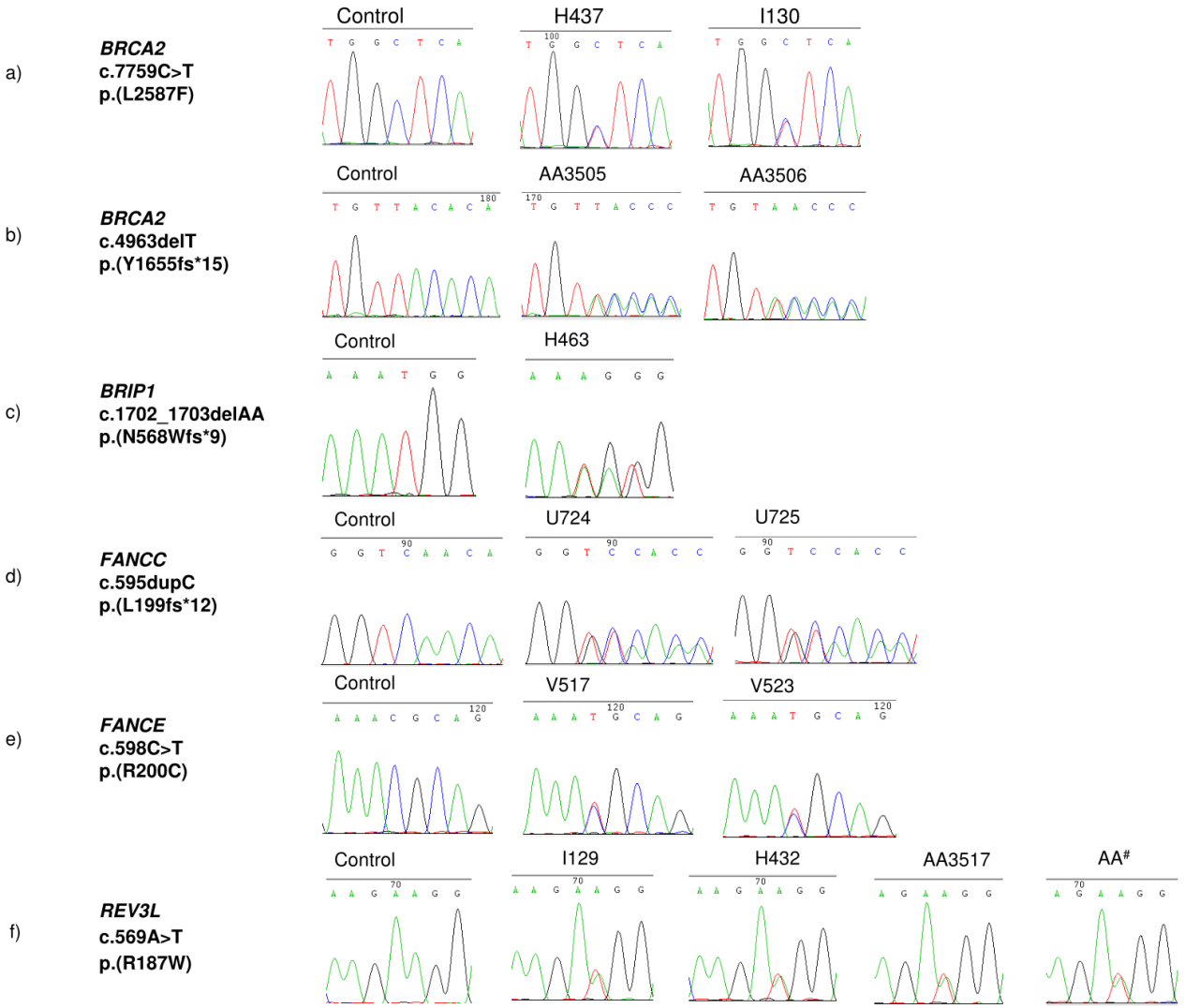
## ACKNOWLEDGEMENTS

We are sincerely grateful to the patients and their families for their participation. CE-J and JM are supported by a contract from CIBERehd. MV-C is supported by Ministerio de Educación, Cultura y Deporte (FPU12/05138). PG is supported by a contract from the Fondo de Investigación Sanitaria (JR13/00013). CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III. This work was supported by grants from the Fondo de Investigación Sanitaria/FEDER (13/02588, 14/00173, 14/00230), the Ministerio de Economía y Competitividad (SAF2014-54453-R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), COST Action BM1206, Beca Grupo de Trabajo 'Oncología' AEG (Asociación Española de Gastroenterología) and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255, 2014SGR135). We thank the Centre Nacional d'Anàlisi Genòmica and the Biobank of Hospital Clínic–IDIBAPS, Barcelona, for technical help. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

1 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**: 2893–2917.
2 Short E, Thomas LE, Hurley J, Jose S, Sampson JR: Inherited predisposition to colorectal cancer: towards a more complete picture. *J Med Genet* 2015; **52**: 791–796.
3 Palles C, Cazier JB, Howarth KM *et al*: Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013; **45**: 136–144.
4 Weren RD, Ligtenberg MJ, Kets CM *et al*: A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat Genet* 2015; **47**: 668–671.
5 Kitao H, Takata M: Fanconi anemia: a disorder defective in the DNA damage response. *Int J Hematol* 2011; **93**: 417–424.
6 Thompson ER, Doyle MA, Ryland GL *et al*: Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet* 2012; **8**: e1002894.
7 Bogliolo M, Surrallés J: Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Curr Opin Genet Dev* 2015; **33**: 32–40.
8 Walden H, Deans AJ: The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu Rev Biophys* 2014; **43**: 257–278.
9 Economopoulou P, Dimitriadis G, Psyrri A: Beyond BRCA: new hereditary breast cancer susceptibility genes. *Cancer Treat Rev* 2015; **41**: 1–8.
10 Garre P, Martín L, Sanz J *et al*: BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. *Clin Genet* 2015; **87**: 582–587.
11 Yurgelun MB, Allen B, Kaldate RR *et al*: Identification of a variety of mutations in cancer predisposition genes in patients with suspected Lynch syndrome. *Gastroenterology* 2015; **149**: 604–613.
12 Seguí N, Mina LB, Lázaro C *et al*: Germline mutations in FAN1 cause hereditary colorectal cancer by impairing DNA repair. *Gastroenterology* 2015; **149**: 563–566.
13 de Voer RM, Hahn MM, Mensenkamp AR *et al*: Deleterious germline BLM mutations and the risk for early-onset colorectal cancer. *Sci Rep* 2015; **5**: 14060.
14 Esteban-Jurado C, Vila-Casadesús M, Garre P *et al*: Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet Med* 2015; **17**: 131–142.



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




**Supplementary Figure 1.** Sanger sequencing chromatograms of the 6 genetic variants belonging to the Fanconi anemia DNA damage repair detected in our sequenced patients and indicated with their codes. a) H437 and I130 from FAM6. b) AA3505 and AA3506 from FAM20. c) H463 patient. d) U724 and U725 from FAMN4. For this variant the reverse sequence is shown. e) V517 and V523 from FAM40. f) I129, H432, AA3517 from FAM3. #AA:family member from FAM3 pedigree with advanced adenomas was tested for segregation. Sequences were compared to a healthy control.

**Supplementay Table 1.** Key terms used to prioritize genetic variants with functional annotations and/or bibliography related to cancer. Functional annotations listed corresponded to those selected from Gene Ontology, KEGG and REACTOME, whereas the Bibliography category referred to terms from NCBI gene summary and Gene Reference Into Function (GeneRIF). Noteworthy, other annotations containing listed key terms were also selected (e.g. “breast cancer” since it contains “cancer”).

| **Gene ontology biological process terms** |
|---|
| Apoptosis and cell death |
| Apoptotic process |
| Execution phase of apoptosis |
| Apoptotic signaling pathway |
| Mitochondrial outer membrane permeabilization |
| Release of cytochrome c from mitochondria |
| Apoptosome assembly |
| Cell death |
| Programmed cell death |
| Autophagic cell death |
| Macroautophagy |
| Microautophagy |
| Regulation of autophagy |
| Cytolysis |
| Necrotic cell death |
| Cell adhesion |
| Cell-cell adhesion |
| Homophilic cell adhesion |
| Cell-cell junction assembly |
| Cell-substrate junction assembly |
| Regulation of cell junction assembly |
| Adherens junction assembly |
| Cell adhesion mediated by integrin |
| Cell-substrate adhesion |
| Regulation of cell adhesion |
| Tight junction assembly |
| Cell cycle |
| Cell cycle |
| Cell cycle phase |
| Cell cycle phase transition |
| Mitotic cell cycle |
| Abortive mitotic cell cycle |
| Mitotic spindle organization |
| Mitotic nuclear envelope disassembly |
| Mitotic nuclear envelope reassembly |
| Establishment of mitotic spindle localization |
| Mitosis |
| Mitotic prophase |
| Mitotic metaphase |
| Mitotic anaphase |
| Mitotic telophase |
| Mitotic prometaphase |
| Mitotic sister chromatid segregation |
| Spindle assembly involved in mitosis |
| Mitotic spindle assembly checkpoint |
| Mitotic centrosome separation |
| Cell cycle arrest |
| Cell cycle checkpoint |
| DNA integrity checkpoint |
| Spindle assembly checkpoint |
| Cell differentiation |
| Cell differentiation |
| Epithelial cell differentiation |
| G1 to G0 transition involved in cell differentiation |
| Mesodermal cell differentiation |
| Stem cell differentiation |
| Cell development |
| Cell fate commitment |
| Negative regulation of cell differentiation |
| Positive regulation of cell differentiation |
| Regulation of epithelial cell differentiation |
| Cell growth |
| Cell growth |

| |
|---|
| Developmental cell growth |
| Regulation of cell proliferation |
| Regulation of epithelial cell proliferation |
| **Cellular response to DNA damage stimulus** |
| DNA repair |
| Single strand break repair |
| Mismatch repair |
| Postreplication repair |
| Double-strand break repair |
| DNA replication proofreading |
| Non-recombinational repair |
| Base-excision repair |
| Nucleotide-excision repair |
| UV-damage excision repair |
| Interstrand cross-link repair |
| DNA damage checkpoint |
| DNA damage response, detection of DNA damage |
| DNA damage induced protein phosphorylation |
| Replication fork processing |
| Telomere maintenance in response to DNA damage |
| Intrinsic apoptotic signaling pathway in response to DNA damage |
| Signal transduction in response to DNA damage |
| **Chromatin** |
| Chromatin organization |
| Helicase activity |
| Maintenance of DNA methylation |
| Chromosome organization |
| Kinetochore organization |
| Chromosome condensation |
| Chromosome decondensation |
| Chromosome breakage |
| Telomere organization |
| Maintenance of DNA repeat elements |
| Sister chromatid cohesion |
| Regulation of cohesin localization to chromatin |
| Centromere complex assembly |
| Regulation of chromosome organization |
| Maintenance of chromatin silencing |
| Telomere maintenance |
| Maintenance of fidelity involved in DNA-dependent DNA replication |
| DNA protection |
| Chromatin remodeling |
| **Pathways** |
| Wnt receptor signaling pathway |
| BMP signaling pathway |
| SMAD protein signal transduction |
| SMAD protein complex assembly |
| Negative regulation of MAPKKK cascade |
| Negative regulation of MAP kinase activity |
| Regulation of JUN kinase activity |
| Negative regulation of JNK cascade |
| Insulin receptor signaling pathway |
| Cellular response to TGF-beta stimulus |
| Cellular response to TGF-beta stimulus |
| Transforming growth factor beta receptor signaling pathway |
| Response to transforming growth factor beta stimulus |
| Epidermal growth factor receptor signaling pathway |
| Notch signaling pathway |
| Positive regulation of transcription of Notch receptor target |
| Notch receptor processing, ligand-dependent |
| Negative regulation of protein kinase B signaling cascade |
| Anterior/posterior pattern formation |
| Netrin-activated signaling pathway |
| Regulation of TOR signaling cascade |
| Target of rapamycin signaling pathway |
| Regulation of phosphatidylinositol 3-kinase cascade |
| NF-kappab import into nucleus |
| Regulation of NIK/NF-kappab cascade |
| Regulation of p52-dependent NF-kappab signaling |

**KEGG pathway terms**

| |
|---|
| Base excision repair |
| Nucleotide excision repair |
| Mismatch repair |
| Homologous recombination |
| Non-homologous end-joining |
| Cell cycle checkpoint |
| Metabolism of xenobiotics by cytochrome P450 |
| Benzoate degradation |
| Aminobenzoate degradation |
| Fluorobenzoate degradation |
| Chloroalkane and chloroalkene degradation |
| Chlorocyclohexane and chlorobenzene degradation |
| Toluene degradation |
| Xylene degradation |
| Nitrotoluene degradation |
| Ethylbenzene degradation |
| Styrene degradation |
| Atrazine degradation |
| Caprolactam degradation |
| DDT degradation |
| Bisphenol degradation |
| Dioxin degradation |

| |
|---|
| Naphthalene degradation |
| Polycyclic aromatic hydrocarbon degradation |
| Furfural degradation |
| Steroid degradation |
| Metabolism of xenobiotics by cytochrome P450 |
| Focal adhesion |
| Adherens junction |
| Tight junction |
| Gap junction |
| Purine metabolism |
| Pyrimidine metabolism |
| Wnt signaling pathway |
| Notch signaling pathway |
| TGF-beta signaling pathway |
| Insulin signaling pathway |
| PI3K-Akt signaling pathway |
| Mucin type O-Glycan biosynthesis |

**Reactome terms**

| |
|---|
| DNA repair |
| Base excision repair |
| Cleavage of the damaged purine |
| Depurination |
| Removal of DNA patch containing abasic residue |
| Resolution of AP sites via the single-nucleotide replacement pathway |
| Resolution of Abasic Sites (AP sites), organism-specific biosystem |
| Gap-filling DNA repair synthesis and ligation in GG-NER |
| Gap-filling DNA repair synthesis and ligation in TC-NER |
| Transcription-coupled NER (TC-NER) |
| Repair synthesis for gap-filling by DNA polymerase in TC-NER |
| Apoptosis |
| Apoptotic execution phase |
| Apoptotic cleavage of cellular proteins |
| Mitotic anaphase |
| M phase |
| G1/S transition |
| DNA replication |
| G2/M transition |

| |
|---|
| Integrin cell surface interactions |
| Adherens junctions interactions |
| Cell-Cell communication |
| Telomere maintenance |
| Extension of Telomeres |
| Chromosome maintenance |
| Signaling by BMP |
| Signaling by TGF beta |
| Signaling by Wnt |
| Beta-catenin phosphorylation cascade |
| Degradation of beta-catenin by the destruction complex |
| PKB-mediated events |
| TGF-beta receptor signaling |
| Constitutive PI3K/AKT Signaling in Cancer |
| Chromosome maintenance |
| O-linked glycosylation of mucins |

**Bibliography terms**

| |
|---|
| Cancer |
| Colon cancer |
| Rectal cancer |
| Tumor |
| Gastrointestinal tumor |
| Tumor suppressor |
| Oncogene |
| Carcinoma |
| Sarcoma |
| Adenoma |
| Polyposis |
| Polyps |
| Cowden |
| Inflammatory bowel disease |
| Crohn |
| Osteosarcoma |
| Glioma |
| Melanoma |
| Leukemia |
| Lymphoma |

## ARTÍCULO 3

### *POLE* and *POLD1* screening in 155 patients with multiple polyps and early-onset colorectal cancer

Esteban-Jurado C, Giménez-Zaragoza D, Muñoz J, Franch-Expósito S, Álvarez-Barona M, Ocaña T, Cuatrecasas M, Carballal S, López-Cerón, MMarti-Solano M, Díaz-Gay M, Castells A, Bujanda L, Balmaña J, Gonzalo V, Llort G, Ruiz-Ponte C, Cubiella J, Balaguer F, Aligué R, Castellví-Bel S. *POLE* and *POLD1* screening in 155 Spanish multiple polyposis and early-onset colorectal cancer patients (Enviado a Scientific Reports).

## *POLE* and *POLD1* screening in 155 patients with multiple polyps and early-onset colorectal cancer

Clara Esteban-Jurado[1], David Giménez-Zaragoza[2], Jenifer Muñoz[1], Sebastià Franch-Expósito[1], Miriam Álvarez-Barona[3], Alejandro Brea-Fernández[3], Teresa Ocaña[1], Miriam Cuatrecasas[4], Sabela Carballal[1], María López-Cerón[1], Maria Marti-Solano[5], Marcos Díaz-Gay[1], Antoni Castells[1], Luis Bujanda[6], Judith Balmaña[7], Victoria Gonzalo[8], Gemma Llort[9], Clara Ruiz-Ponte[3], Joaquín Cubiella[10], Francesc Balaguer[1], Rosa Aligué[2], Sergi Castellví-Bel[1,*].

1. Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, 08036 Barcelona, Catalonia, Spain.
2. Biomedical Sciences Department, School of Medicine, University de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Catalonia, Spain.
3. Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, 15706 Santiago de Compostela, University of Santiago de Compostela, Galicia, Spain.
4. Department of Pathology, Hospital Clinic, Biobanc Clinic-IDIBAPS, Villarroel 170, 08036 Barcelona, Catalonia, Spain.
5. Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6, 35032 Marburg, Germany.

6. Gastroenterology Department, Hospital Donostia – Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Basque Country University (UPV/EHU), 20080 San Sebastián, Spain.

7. High Risk and Cancer Prevention Unit, Medical Oncology Department, University Hospital Vall d'Hebron and Vall d'Hebron Institute of Oncology, 08035 Barcelona, Spain.

8. Gastroenterology Department, Hospital Universitari Mútua de Terrassa, 08221 Terrassa, Barcelona, Spain.

9. Clinical Oncology Department, Corporacio Parc Tauli, 08208 Sabadell, Barcelona, Spain.

10. Gastroenterology Department, Complexo Hospitalario Universitario de Ourense, Instituto de Investigación Biomédica Ourense, Pontevedra y Vigo, 32005 Ourense, Spain.

*Corresponding author: Dr. Sergi Castellví-Bel, Genetic Predisposition to Colorectal Cancer Group, IDIBAPS, Centre Esther Koplowitz (CEK), Rosselló 153 planta 4, 08036 Barcelona, Catalonia, Spain. Tel. +34-93-2275400 ext.4183/2915, email: sbel@clinic.ub.es

Keywords: Colorectal neoplasm; colorectal adenoma; genetic predisposition to disease; hereditary disease; *POLE*; *POLD1*; genetic variant.

**ABSTRACT**

Germline mutations in *POLE* and *POLD1* have been shown to cause predisposition to colorectal multiple polyposis and a wide range of neoplasms, being early-onset colorectal cancer the most prevalent. In order to find additional mutations affecting the proofreading activity of these polymerases, we sequenced its exonuclease domain in 155 patients with multiple polyps or an early-onset colorectal cancer phenotype without alterations in the known hereditary colorectal cancer genes. Interestingly, none of the previously reported mutations in *POLE* and *POLD1* were found. On the other hand, among the genetic variants detected, only two of them stood out as putative pathogenic in the *POLE* gene, c.1359+46del71 and c.1420G>A (p.Val474Ile). The first variant, detected in two families, was not proven to alter correct RNA splicing. Contrarily, c.1420G>A (p.Val474Ile) was detected in one early-onset colorectal cancer patient and located right next to the exonuclease domain. The pathogenicity of this change was suggested by its rarity and bioinformatics predictions, and it was further confirmed by functional assays in *Schizosaccharomyces pombe*. This is the first study to functionally analyse a *POLE* mutation outside the exonuclease domain and widens the spectrum of genetic changes in this DNA polymerase that could lead to colorectal cancer predisposition.

Keywords: Colorectal neoplasm, polymerase epsilon, polymerase delta, mutational screening, hereditary disease, genetic variant.

## INTRODUCTION

Colorectal cancer (CRC) is one of the most common tumours and an important cause of mortality in the developed world[1]. It is caused by environmental and genetic factors, with 35% of the variation in CRC susceptibility probably explained by inherited causes[2]. The best known examples of inherited CRC predisposition are Mendelian forms such as Lynch syndrome and familial adenomatous polyposis. They account for ~5% of all cases, and are due to germline mutations in *APC*, *MUTYH* and the mismatch repair (MMR) genes, which confer a high risk of developing this disease[3]. However, there is still a considerable number of cases with strong familial CRC aggregation and early disease onset with an unknown inherited genetic basis. An example could correspond to familial CRC type X, where Amsterdam clinical criteria used for Lynch Syndrome identification are fulfilled but there are no alterations in the MMR system[4].

In order to identify new inherited risk factors, whole-genome sequencing combined with linkage disequilibrium studies were recently conducted in families affected by multiple colorectal adenomas and early-onset CRC[5]. By doing so, p.Leu424Val and p.Ser478Asn mutations in *POLE* and *POLD1* DNA polymerases, respectively, were established as a new high-penetrance cause of germline CRC predisposition with an autosomal dominant pattern of inheritance. *POLD1* p.Ser478Asn was also found to be involved in germline predisposition to endometrial cancer[5,6]. These mutations are located in the exonuclease domain of the protein, which has proofreading activity by removing misincorporated nucleotides during DNA replication[7]. Therefore, mutations in

this protein domain will disrupt the fidelity of DNA replication, which will lead to a mutator phenotype, resulting in tumourigenesis. The pathogenicity of these first identified variants was confirmed by functional studies in the orthologous genes in yeast. Regarding somatic studies, tumours from *POLE* and *POLD1* mutation carriers showed a hypermutated phenotype with an excess of G>T/C>A and C>T/G>A transversions, especially in the context TCT>TAT and TCG>TTG[5,8-10].

Since the discovery of these two new hereditary CRC genes, some additional efforts have been made to characterize the mutational spectrum and the clinical features associated to this new syndrome, which was accordingly named polymerase proofreading-associated polyposis[6]. To this date, *POLE* mutations have been found to be the germline predisposition factor in families with multiple adenomas and early-onset CRC[11-16], as well as in other neoplasms such as endometrial, ovarian, brain, pancreas, small intestine and cutaneous melanoma[15,17-19]. On the other hand, *POLD1* mutations have been found to predispose carriers to multiple adenomas, CRC, endometrial and breast cancer[11,14,15]. Recently, some CRC patients with deficient MMR system caused by tumour biallelic inactivation were reported to also carry germline mutations in *POLE*, being these MMR somatic mutations the consequence of the *POLE* hypermutator tumour phenotype[12,20].

Regarding the prevalent mutations reported up to now, *POLE* p.L424V has been detected in 24 independent families[5,11-15,17], whereas *POLD1* p.S478N has been found in 4 independent families[5,14]. Additionally, new rare variants located in the active exonuclease domain of these two proteins have been reported. Among them, alteration of the proofreading activity as evidence of their pathogenicity was confirmed for some variants by functional assays in

yeast, T4 bacteriophage or *E. coli* orthologous polymerases. These functional results were previously available for *POLE* p.D368V*,* p.Y458F[14,18], and *POLD1* p.D316H, p.D316G, p.P327L and p.L474P[5,11,15], or they were specifically produced for *POLE* p.W347C[19].

Finally, the phenotype associated to this new hereditary CRC syndrome is not well defined yet and a better definition of the clinical characteristics will most likely help to detect potential mutation carriers in the general population. Accordingly, the aim of our study was to screen the exonuclease domain of *POLE* and *POLD1* in 155 patients with multiple polyps and early-onset CRC in order to find mutations affecting the replication fidelity of these proteins and shed light on this matter. Our final goal was to facilitate genetic counselling in order to correctly implement preventive strategies in those families.

## RESULTS

Patient characteristics

Clinical characteristics are summarized in **Table 1**. Our patient cohort included predominantly cases with multiple polyps (MP) phenotype (83 cases, 53.5%), most of them presenting adenomas (67.5%) with a mean polyp onset age around 57 y.o. The early-onset CRC group corresponded to 59 cases (38.1%) and distal location of the tumour was predominant. The less numerous group was the MMR-defective CRC without germline alterations in the known genes (13 patients, 8.4%). Noteworthy, adenomas were also commonly present in both selected CRC groups (59.3% and 66.7%).

Variant detection

After screening the exonuclease domain of *POLE* and *POLD1* in our cohort of patients, we detected several genetic variants that are listed in **Table 2**. Firstly, it is important to highlight that none of the two previously described recurrent mutations in *POLE* (p.Leu424Val) and *POLD1* (p.Ser478Asn) were present in our cohort. Among the detected variants, some of them corresponded to polymorphisms with an allelic frequency >10%, they were also present in accessed human genetic variant databases with a similar frequency and they were not considered for further analysis. We proceeded to select those genetic variants with an allelic frequency <10% or not present in any of the checked human genetic variant databases. Among them, some variants were located in intronic sequences or corresponded to coding synonymous variants. Their putative involvement in abnormal splicing processing was evaluated by using

several bioinformatics tools. Results are summarized in **Supplementary Table S1**. None of the analysed variants showed a prediction of altered splicing by 2 or more tools. However, one of the variants corresponded to a deletion of 71 base pairs in intron 13 of the *POLE* gene. This variant was not present in any of the human genetic variation databases but it was found in 2 early-onset CRC patients in our cohort, so we decided to further characterize it. In order to do so, we analysed its putative splicing alteration in a carrier at the RNA level by using RT-PCR. When compared to a control RNA not carrying this variant, we did not detect any additional amplification band in the deletion carrier being indicative of no splicing alteration, which was also confirmed by Sanger sequencing (**Supplementary Fig. S1**).

Novel *POLE* missense variant

Besides the previously described variants, we detected an interesting *POLE* missense variant in an early-onset CRC patient that corresponded to c.1420G>A (p.Val474Ile) (**Table 2**). The carrier proband presented a distal MMR-proficient CRC at the age of 47 and 2 non-advanced adenomas. Her father and maternal aunt also had CRC at the age of 58 and 47, respectively, and one of her sisters presented non-advanced adenomas (**Fig. 1**). This family fulfilled Amsterdam criteria and it could be considered an example of familial CRC type X. Unfortunately, variant segregation analysis was not feasible since DNA was unobtainable from additional family members.

The *POLE* c.1420G>A (p.Val474Ile) was suggestive of being a new putative pathogenic variant. It was not present in any of the accessed human genetic variation databases including a Spanish repository and the affected

amino acid was found conserved in 100 vertebrates as well as in *D. melanogaster*, *C. elegans* and yeast. Regarding its protein location, Val474 is located in the N-terminal domain but very close to the C terminus end of the exonuclease domain (only 3 amino acids away). In order to assess the potential functional impact of the *POLE* Val474Ile variant, we analysed further its position in the polymerase structure. To do so, we took advantage of the crystallized structure of the *Saccharomyces cerevisiae* DNA polymerase epsilon (PDB IB 4M8O). Indeed, the acceptable amount of sequence identity between human and yeast proteins (57%) had already been exploited for the construction of a homology model for human POLE (available at the ModBase database, UP Q07864). Analysis of this model allowed us to locate the Val474Ile variant in the N-terminal domain but in very close proximity to the exonuclease domain (**Fig. 2A**). Superposition with the template structure (**Fig. 2B**) in complex with DNA showed that the variant may not affect DNA binding directly. Conversely, it could have an indirect effect on the helical packing of the exonuclease and N-terminal domains, which could distort the physiological conformation needed for a correct polymerase activity. Furthermore, besides the structural analysis, additional bioinformatics assessment of this variant (using PolyPhen and CADD) predicted a deleterious effect for this amino acid change and most protein stability predictions were also in favour of a damaging effect (Eris, PoPMuSiC, I-Mutant).

Functional studies in yeast for POLE Val474Ile

In order to further test the putative functional effect of this variant, we proceeded to analyse it in yeast. *POLE* V474 residue is highly conserved in eukaryotes,

including *Schizosaccharomyces pombe* (*S. pombe*). We constructed the equivalent substitution in this organism, *pol2* p.V475I, and another strain carrying the equivalent change to the previously reported *POLE* p.L424V mutation (*pol2* p.L425V) as a positive control. We compared the *ade6*-485 allele reversion rates in these mutant strains with the wild-type *pol2* strand as negative control (**Fig. 3**). When comparing *Pol2-L425V* substitution with the wild-type strain, 72 times more revertants were observed (*P*-value<0.01). Regarding *Pol2* p.V475I, the mutation rate was 31 times higher when compared to the wild-type strain (*P*-value <0.005).

Somatic studies for *POLE* Val474Ile

When tested on the corresponding tumour, loss of heterozygosity (LOH) could not be detected by Sanger sequencing in the carrier when comparing to her germline DNA. Whole-exome sequencing (WES) was also performed in the tumour DNA of the patient carrying the *POLE* p.Val474Ile variant to study the number and spectrum of somatic mutations. A second mutational event in the *POLE* gene was not found. We had available WES data obtained from four other MMR proficient CRC tumours that did not present germline or tumour alterations in *POLE* or *POLD1* to compare their number of substitutions and mutational spectrum with our *POLE* mutant. Regarding tumour WES results in the five samples analysed, mean coverage was >90X and >79% of DNA in each tumour was sequenced with ≥30X coverage. First, a tumour profile for each sample was generated by eliminating variants present in a germline exome dataset. Mutation density plots suggested that tumour profiles were correctly generated and germline variants were mostly eliminated

(**Supplementary Fig. S2**). Additionally, WES data was normalized by selecting only those sequenced regions with coverage ≥10X in all five tumour samples. By doing so, a slight increase of the total number of substitutions in the tumour DNA from the p.Val474Ile germline carrier was detected when compared with the *POLE* wild-type MMR proficient tumours (**Supplementary Table S2**). However, when mutation spectrum was analysed, the tumour DNA from the p.Val474Ile germline carrier did not show an increase in G>T/C>A or C>T/G>A transversions as suggested by previous studies.

## DISCUSSION

Our molecular screening of the *POLE* and *POLD1* exonuclease domains in a cohort of 155 patients with MP or early-onset CRC identified a novel *POLE* mutation in one family. We also detected several intronic variants most likely polymorphic and without pathogenic involvement.

Importantly, we did not find any of the previously described mutations for *POLE* (p.Leu424Val) and *POLD1* (p.Ser478Asn) in our cohort. As reported by previous studies, the *POLE* p.Leu424Val mutation frequency in multiple colorectal adenomas or familial CRC cohorts seems to be typically ≤0.3%[5,11,12,14,15], whereas the *POLD1* p.Ser478Asn mutation seems to be even less frequent (<0.1%)[5,14]. Taking our results into account, we can conclude that these mutations have indeed a very low frequency. On the other hand, we can also hypothesize that our sample size was probably not large enough to be able to detect any carrier of these mutations (334 samples needed to be screened to detect one carrier at a 0.3% frequency). Additionally, it could be also possible that the frequency of these mutations may be even lower in the Spanish population since only one carrier for the *POLE* p.Leu424Val mutation and no *POLD1* p.Ser478Asn mutation carriers have been reported so far[11].

Leaving aside these potentially recurrent mutations, other different variants have already been reported in *POLE* and *POLD1*. Most of them are located in the protein exonuclease domain and include p.Trp347Cys, p.Asn363Lys, p.Asp386Val, p.Lys425Arg, p.Pro436Ser, p.Tyr458Phe in *POLE*[13,14,17-20], and p.Asp316His, p.Asp316Gly, p.Pro327Leu, p.Arg409Trp, p.Leu474Pro for *POLD1*[5,11,15]. These previously reported variants and our newly

identified variant are indicative that the entire coding region for *POLE* and *POLD1* should be screened instead of focusing only in a few variants.

It should also be noted that the phenotype selection criteria in our screened cohort included MP with at least one affected first-degree relative, early-onset CRC or MMR-defective CRC without germline alterations in the known genes. Previous studies have either used similar[11] or more permissive selection criteria[12-14] with similar results. It could be argued that including only MP with family history may have reduced the chances of detecting carriers. Considering previously reported known *POLE* or *POLD1* mutation carriers, the phenotypic spectrum included MP and early-onset CRC, as well as family history and would reinforce therefore the phenotype selection criteria used in our cohort. Regarding our molecular screening approach, we used PCR amplification of genomic DNA and subsequent Sanger sequencing corresponding to the entire exonuclease domain and adjacent intronic sequences. This approach is not biased to detect only *POLE* p.Leu424Val and *POLD1* p.Ser478Asn mutations as it was the case for some previous studies[11,12] and permitted to detect additional mutations located in this region.

We were able to detect a new mutation in the *POLE* gene corresponding to c.1420G>A (p.Val474Ile). The heterozygous carrier was recruited in the early-onset CRC group and belonged to an Amsterdam I family without alterations in the MMR repair system. Its rarity, amino acid species conservation and location in the POLE protein already predicted a plausible pathogenic role. Our results in yeast suggest that the human *POLE* p.L424V and *POLE* p.V474I variants will cause an increased mutation rate due to faulty proofreading activity of this protein, although with an attenuated phenotype when compared to *POLE*

p.L424V. Exome sequencing of the tumour corresponding to the p.V474I variant carrier revealed a slightly higher number of substitutions compared with four *POLE* wild-type tumours, although the increment was not as high as could be expected taking into account previously reported data[5,10]. Besides, the percentage of G>T/C>A transversions was not increased compared with the other *POLE* wild-type tumours.

It is also worth mentioning that the same variant was previously reported as somatic in the COSMIC database in a gastric cancer with microsatellite instability (TCGA-BR-6452-01)[21]. This gastric tumour presented, as defined by the authors, “an ultramutated profile”, with 11,375 substitutions, and a mutation rate of approximately 283 mutations per Megabase. The percentage of G>T/C>A transversions in this sample was 10.29%. Notably, this ultramutated tumour presented another three somatic missense variants in *POLE* located far away from the exonuclease domain (p.R1111Q, p.S681R and Y1889C) that could be promoting even a stronger effect. In this regard, it is also worth mentioning that previous reports regarding *POLE* tumour mutation profile were generated mostly by using gene panel sequencing and with much higher coverage than the used in the present study. We can hypothesize that our tumour profiling may have failed to show a clear distinctive profile for our novel *POLE* mutation due to a suboptimal filtering of germline variants and a limited sequencing coverage, as well as a milder mutator effect as shown by the functional assays in yeast.

In conclusion, we detected a new *POLE* mutation, p.V474I, in an early-onset CRC patient. This variant is located right next to the exonuclease domain and

affects protein function, leading to a proofreading activity defect as shown by yeast studies. The pathogenicity of this change was also suggested by bioinformatics and protein structure predictions. When checking its tumour profile, it showed an increase in the number of variants but not as strong as in the p.L424V POLE mutation in agreement with the yeast functional results. It is also worth mentioning that this is the first study to functionally analyse a *POLE* mutation outside the exonuclease domain in *POLE* and widens the spectrum of genetic changes in this DNA polymerase that could lead to CRC predisposition.

## METHODS

### Patients

Three subgroups of patients were studied including those presenting MP, early-onset CRC or MMR-defective CRC without germline alterations in the known hereditary CRC genes. MP patients presented 10-100 polyps, being the main precursor lesion adenomatous, serrated or a combination of adenomatous and serrated polyps with an age of onset <70 and no alterations in the *APC* or *MUTYH* genes. Another selection criteria for the MP group included having at least one first-degree relative with MP, CRC, advanced adenomas or endometrial cancer diagnosed before the age of 70. Early-onset CRC patients were selected with an age of onset <50 and no alterations in *MUTYH* or the MMR genes. CRC patients with MMR deficiency presented loss of MMR protein (MLH1, MSH2, MSH6, PMS2) expression by immunohistochemistry with neither detected germline mutation in the MMR genes, nor somatic *MLH1* hypermethylation.

We selected 155 patients (83 MP, 59 early-onset CRC and 13 MMR-defective CRC) from seven Spanish hospitals (Hospital Clínic de Barcelona, 61 patients; Complexo Hospitalario de Ourense, 28 patients; Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, 18 patients; Corporació Sanitaria Parc Taulí, 18 patients; Hospital Mútua de Terrasa, 14 patients; Hospital Vall d'Hebrón, 13 patients; Hospital de Donostia, 3 patients). This project respected the fundamental principles established in the Helsinki Declaration, including any later update, by the World Medical Association, by the European Council declaration related to human rights and biomedicine, in

the UNESCO Universal Declaration about the human genome and human rights, as well as fulfilling the terms established by the Spanish legislation in the area of biomedical research, personal data protection and bioethics, and especially the Biomedicine Research law and Biobanking regulations (14/2007 3rd July and RD 1716/2011). This study was approved by the institutional ethics committee of each participating hospital and written informed consent was obtained at diagnosis on a systematic basis.

Germline DNA samples used for Sanger sequencing were obtained from peripheral blood, whereas in one case showing a relevant POLE variant, formalin-fixed, paraffin-embedded (FFPE) tumour DNA was also isolated for LOH and tumour profiling studies using the QIAamp DNA Blood Kit or QIAamp Tissue Kit, respectively (QIAGEN, Redwood City, USA) and following manufacturers' instructions. RNA for splicing analysis was obtained from peripheral blood collected in a PAXgene Blood RNA tube in one patient and isolated using the PAXgene blood RNA kit (PreAnalytiX, Hombrechtikon, Switzerland) following the manufacturer's protocol.

Variant detection

The entire exonuclease domain of *POLE* and *POLD1* was screened for mutations by PCR amplification using custom primers designed with Primer3Plus (**Supplementary Table S3**) and subsequent Sanger sequencing (GATC Biotech, Germany). For the *POLE* gene, 2,095 nucleotides were sequenced including exons 9-14 and covering protein residues 267-491. The sequenced region for the *POLD1* gene covered 2,538 nucleotides, included exons 8-13 and comprised protein residues 304-562. Besides including the

coding sequence for the entire exonuclease domain of *POLE* and *POLD1*, additional flanking intronic sequences were also screened, including introns 9, 11 and 13 entirely and fragments of introns 8, 10, 12 and 14 for *POLE*, as well as introns 7, 8, 9 and 11 entirely and introns 10, 12 and 13 partially for *POLD1*. Resulting sequences were compared to the human reference genome hg19 and visualized using Chromas (http://technelysium.com.au/) and ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) All variants that differed from the reference genome were reported and their frequency was checked in several human genetic variation databases including dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), 1000Genomes (http://www.1000genomes.org/), Exome Variant Server (http://evs.gs.washington.edu/EVS/), Exome Aggregation Consortium (http://exac.broadinstitute.org/) and CIBERER Spanish variant server (http://csvs.babelomics.org/).

Variant evaluation

Synonymous and intronic variants with an allelic frequency <10% were analysed with Human Splicing Finder, (http://www.umd.be/HSF/) Berkeley Drosophila Genome Project (http://www.fruitfly.org/seq_tools/splice.html) and SPANR (http://tools.genes.toronto.edu/) in order to predict plausible splice site alterations.

One rare intronic variant in the *POLE* gene was studied at the RNA level by RT-PCR and PCR amplification using custom primers located two exons upstream and downstream from the variant (**Supplementary Table S3**) and Sanger sequencing to verify for correct exon splicing. A nonsynonymous variant

with an allelic frequency <10% was evaluated with bioinformatics tools (Polyphen, http://genetics.bwh.harvard.edu/pph2/ and CADD_phred, http://cadd.gs.washington.edu/score) in order to predict its possible effect on the protein function. This variant has been submitted to the ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/; accession number SUB1552845). Amino acid conservation in 100 vertebrates was also checked in the UCSC genome browser (http://genome.ucsc.edu/).

Structural analysis

The sequence of DNA polymerase epsilon catalytic subunit A (POLE) was retrieved from Uniprot (ID Q07864) and used to perform a protein blast against the Protein Data Bank. That revealed that the closest available crystal structure was the *Saccharomyces cerevisiae* DNA polymerase epsilon (PDB IB 4M8O, http://www.rcsb.org/pdb/explore/explore.do?structureId=4M8O)[22]. Indeed, a model of human POLE (identity of 57% with 4M8O) is deposited in the ModBase database (annotated as UP Q07864, http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi)[23]. This model was visualized using the VMD 1.9.1 software[24] and the Stamp plugin was used to superpose the model with the original template to assess the relative position of DNA. Protein stability predictions were performed using Eris (http://troll.med.unc.edu/eris/), PoPMuSiC (http://dezyme.com/en/Software), I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) and CUPSAT (http://cupsat.tu-bs.de/).

Functional assays

Both *pol2*-L425V and *pol2*-V475I *S. pombe* mutant strains (encoding variants equivalent to human *POLE* L424V and V474I respectively) were constructed by cloning the wild type *pol2* gene fragment into the pFA6a-kanMX6 vector, followed by point mutation and insertion into an *ade6*-485 strain by recombination. The *pol2* wild-type segment was amplified using primers Pol2 fw (including *Bgl*II restriction site) and Pol2 rev (including *Asc*I restriction site) (**Supplementary Table S3**). The PCR product and the pFA6a-*kanMX6* plasmid were digested with *Asc*I/*Bgl*II and *Asc*I/*Bam*HI respectively, and *Bgl*I and *Bam*HI sites were abolished after ligation. Point mutations were performed with Quikchange lightning site-directed mutagenesis kit (Agilent, Santa Clara, CA, USA) using primers Pol2-L425V fw + Pol2-L425V rev for *pol2*-L425V and Pol2-V475I fw + Pol2-V475I rev for *pol2*-V475I (**Supplementary Table S3**). The created pFA6a-*kanMX6* plasmids carrying wild-type *pol2* version, *pol2*-L425V or *pol2*-V475I mutants were integrated into *pol2* locus of the ade6-485 strain after linearization with *Bam*HI and selecting for G418/Geneticin resistance. Constructs were verified by Sanger sequencing. Mutation rates of the *ade6*-485 allele in the different strain backgrounds (wild-type *pol2* as negative control, *pol2*-L425V positive control and *pol2*-V475I as the assayed genetic variant) were determined by fluctuation analysis as previously described[25]. Two cultures were used for each construct, culturing $0.6x10^7$ cells/plate. Plates were scored after 12 days.

Somatic analysis

LOH was studied in FFPE tumour DNA comparing to the germline DNA of a patient carrying a missense variant in *POLE* by PCR amplification using specific

primers in the region surrounding the variant designed (**Supplementary Table S3**) and subsequent Sanger sequencing (GATC Biotech, Germany).

WES was performed in the tumour DNA of a patient carrying a missense variant detected in *POLE* and four other MMR proficient CRC tumours that did not present germline or tumour alterations in *POLE* or *POLD1* in order to study the number and spectrum of somatic mutations. WES was characterized by using the HiSeq2000 platform (Illumina, San Diego, USA) and SureSelectXT Human All Exon for exon enrichment V4 (Agilent, Santa Clara, USA). Since paired germline WES data was not available for the analysed samples and to select only for tumour-only variants, we first selected only heterozygous substitutions not present in a germline WES dataset of 65 individuals to generate a somatic profile for the five samples analysed. Results took into account only those regions with coverage >10X in all samples in order to normalize for DNA and sequencing quality.

**REFERENCES**


1. Ferlay, J., et al. GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available at: http://globocan.iarc.fr (Accessed: 20th March 2016).
2. Lichtenstein, P., et al. 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 (2000).
3. Short, E., Thomas, L. E., Hurley, J., Jose, S., Sampson, J. R. Inherited predisposition to colorectal cancer: towards a more complete picture. J. Med. Genet. **52**, 791-6 (2015).
4. Lindor, N. M., et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA* **293**, 1979–1985 (2005).
5. Palles, C., et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45**, 136-44 (2013).
6. Briggs, S., Tomlinson, I. Germline and somatic polymerase ε and δ mutations define a new class of hypermutated colorectal and endometrial cancers. *J Pathol* **230**, 148-53 (2013).
7. Shevelev, I. V., Hübscher, U. The 3' 5' exonucleases. *Nat. Rev. Mol. Cell Biol.* **3**, 364-76 (2002).
8. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-7

(2012).

9. Shinbrot, E., et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res* **24**, 1740-50 (2014).
10. Rayner, E., et al. A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat. Rev. Cancer* **16**, 71-81 (2016).
11. Valle, L., et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. *Hum. Mol. Genet.* **23**, 3506-12 (2014).
12. Elsayed, F. A., et al. Germline variants in POLE are associated with early-onset mismatch repair deficient colorectal cancer. *Eur. J. Hum. Genet.* **23**, 1080-4 (2015).
13. Spier, I., et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. *Int. J. Cancer* **137**, 320-31 (2015).
14. Chubb, D., et al. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. *J. Clin. Oncol.* **33**, 426-32 (2015).
15. Bellido, F., et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. *Genet. Med.* **18**, 325-32 (2016).
16. Rohlin, A., et al. GREM1 and POLE variants in hereditary colorectal cancer syndromes. *Genes Chromosomes Cancer* **55**, 95-106 (2016).
17. Rohlin, A., et al. A mutation in POLE predisposing to a multi-tumour

phenotype. *Int. J. Oncol.* **45**, 77-81 (2014).

18. Hansen, M. F., et al. A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine. *Fam. Cancer* **14**, 437-48 (2015).
19. Aoude, L. G., et al. POLE mutations in families predisposed to cutaneous melanoma. *Fam. Cancer* **14**, 621-8 (2015).
20. Jansen, A. M., et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. *Eur. J. Hum. Genet.* 2015 Dec 9. doi: 10.1038/ejhg.2015.252..
21. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-9 (2014).
22. Hogg, M., et al. Structural basis for processive DNA synthesis by yeast DNA polymerase ε. *Nat. Struct. Mol. Biol.* **21**, 49–55 (2013).
23. Pieper, U., Eswar, N., Stuart, A. C., Ilyin, V. A., Sali, A. MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* **30,** 255–9 (2002).
24. Humphrey, W., Dalke, A., Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–8 (1996).
25. Mansour, A. A., Tornier, C., Lehmann, E., Darmon, M., Fleck, O. Control of GT repeat stability in Schizosaccharomyces pombe by mismatch repair factors. *Genetics* **158**, 77-85 (2001).
26. Desmet, F. O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., Béroud, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
27. Reese, M. G., Eeckman, F. H., Kulp, D., Haussler, D. Improved splice site

detection in Genie. *J. Comput. Biol.* **4**, 311-23 (1997).

28. Xiong, H. Y., et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).

**ACKNOWLEDGEMENTS**

We are sincerely thankful to the patients and their families for their participation. We are sincerely grateful to the Centre Nacional d'Anàlisi Genòmica and the Biobank of Hospital Clínic–IDIBAPS, Barcelona, for technical help. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona. CEJ, SFE and JM are supported by a contract from CIBERehd. CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III. This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (13/00719, 14/00173, 14/00230), Ministerio de Economía y Competitividad (SAF2014-54453-R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), COST Action BM1206, Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya) and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255, 2014SGR135).

## AUTHOR CONTRIBUTIONS

S.C.-B., R.A., and F.B. conceived and designed the study. C.E.-J., D.G.-Z., J.M., S.F.-E., M.M.-S. and M.D.-G. performed the experiments. S.C.-B., R.A. and C.E.-J. analysed the data. M.A.-B., A.B.-F., T.O., M.C., S.C., M.L.-C., A.C., L.B., J.B., V.G., G.L., C.R.-P., J.C. and F.B. contributed to the collection of samples, reagents, materials, or analyses tools. C.E.-J. and S.C.-B. wrote the paper. All authors revised the manuscript.

## COMPETING INTERESTS

The authors declare no competing financial interests.

## FIGURE LEGENDS

**Figure 1.** Pedigree for the proband carrying the *POLE* c.1420G>A (p.Val474Ile) variant.

**Figure 2.** Structural analysis of the Val474Ile variant in the human POLE protein. **A)** Representation of the homology model. Polymerase domains are represented as follows: N-terminal (yellow), exonuclease (violet), palm (lime), thumb (orange) and fingers (silver). Position 474 is represented in magenta in van der Waals surface. **B)** Superposition of the model and the template structure of the *Saccharomyces cerevisiae* DNA polymerase epsilon allows determining the relative position of the DNA chain (shown in black).

**Figure 3.** Mutation rates of *Schizosaccharomyces pombe* strains expressing pol2 p.L425V and p.V475I variants corresponding to human POLE p.L424V and p.V474I, respectively. Mutation rates for Ade+ reversion of the *ade6*-485 allele are shown. Numbers were calculated from two experiments with two independently generated strains per $0{,}6 \times 10^7$ cell divisions. In the graphic, vertical bars correspond to standard deviation (SD) of the fold-change increase in the number of revertants.

**Table 1.** Clinical characteristics of the 155 patients with multiple polyps, early-onset CRC or mismatch repair-defective CRC.

| **Multiple polyps patients** | | |
|---|---|---|
| Number of cases | 83 | |
| Females | 26 (31.3%) | |
| *Type of lesion/polyps* | | |
| AFAP* | 56 (67.5%) | |
| Serrated* | 4 (4.8%) | |
| Mixed polyps* | 6 (7.2%) | |
| Unknown[#] | 17 (20.5%) | |
| Polyp onset, range (mean) | 32-70 y.o. (56.8) | |
| Adenomas, range (mean) | 4-91 (28) | |
| Serrated polyps, range (mean) | 0-63 (7.8) | |
| CRC | 21 (24.7%) | |
| Other neoplasms^ | 5 | |
| Advanced adenoma | 54 | |
| FDR with CRC | 50 (60.2%) | |
| FDR with endometrial cancer | 9 | |
| FDR with advanced adenoma | 15 | |
| **CRC patients** | | |
| | **Early-onset CRC** | **MMR-defective CRC** |
| Number of cases | 59 | 13 |
| Females | 33 (55.9%) | 6 (46.15%) |
| CRC onset, range (mean) | 22-50 y.o. (41.9) | 31-68 y.o. (53.7) |
| *CRC location* | | |
| Proximal to the splenic flexure | 15 (25.4%) | 7 (53.8%) |
| Distal | 36 (61%) | 3 (23.1%) |
| Unknown[#] | 8 (13.5%) | 3 (23.1 %) |
| Patients with adenomas | 35 (59.3%) | 8 (66.7%) |
| Adenomas, range (mean) | 0-25 (3.1) | 0-18 (4) |
| Patients with other neoplasms^ | 2 (3.3%) | 2 (15.4%) |
| Familial CRC type X | 11 (19.3%) | -- |
| FDR with CRC | 33 (57.9%) | 1 (7.4%) |
| FDR with EC | 4 (7%) | 3 (25%) |
| FDR with AA | 14 (24.6%) | 1 (7.7%) |

AFAP, attenuated familial adenomatous polyposis; y.o., years old; CRC, colorectal cancer; FDR, first degree relative.
*The main precursor lesions were 10-100 adenomatous, serrated or mixed polyps.
[#]Clinical data was partially available for 17 multiple polyposis and 3 MMR-defective CRC cases.
^Include those located in thyroid gland, bladder, prostate, pancreas, breast, uterus, lymphatic cells, kidney and larynx.

**Table 2.** Genetic variants detected in screening of the exonuclease domain of the *POLE* and *POLD1* genes.

*POLE*

| Variant | Location | rs ID | GMAF | EVS | ExAC | CSVS | Inner allelic frequency |
|---|---|---|---|---|---|---|---|
| c.910-157A>G | intronic | rs5744759 | 0.282 | NA | NA | NA | 0.374 |
| c.910-6G>C | intronic | rs4077170 | 0.443 | 0.2957 | 0.6307 | 0.406 | 0.571 |
| c.1020+29C>T | intronic | rs369332806 | 0.0002 | NA | 0.0001591 | NA | 0.0032 |
| c.1020+46C>A | intronic | rs375701878 | 0.0002 | NA | 0.000135 | NA | 0.0032 |
| c.1226+13G>A | intronic | rs577646338 | NA | NA | 0.0001171 | NA | 0.0032 |
| c.1226+44G>A | intronic | rs79883120 | 0.0014 | NA | 0.0001274 | NA | 0.0097 |
| c.1226+45C>T | intronic | rs5744761 | 0.0655 | 0.0464 | 0.05914 | 0.043 | 0.0161 |
| c.1359+43G>A | intronic | rs4883555 | 0.4235 | 0.4302 | 0.4957 | 0.313 | 0.365 |
| c.1359+46del71 | intronic | NA | NA | NA | NA | NA | 0.0065 |
| c.1359+144G>T | intronic | rs5744776 | 0.0569 | NA | NA | NA | 0.0581 |
| c.1420G>A (p.Val474Ile) | missense | NA | NA | NA | NA | NA | 0.0032 |

*POLD1*

| Variant | Location | rs ID | GMAF | EVS | ExAC | CSVS | Inner allelic frequency |
|---|---|---|---|---|---|---|---|
| c.970+79G>A | intronic | rs559071730 | 0.0002 | NA | NA | NA | 0.0032 |
| c.971-93G>C | intronic | NA | NA | NA | NA | NA | 0.0032 |
| c.1137+19C>G | intronic | rs572449832 | 0.0002 | NA | NA | NA | 0.0032 |
| c.1137+53G>A | intronic | rs1673043 | 0.2925 | NA | NA | 0.054 | 0.0613 |
| c.1137+69G>A | intronic | NA | NA | NA | NA | NA | 0.0032 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| c.1138-8A>G | intronic | rs41544624 | 0.0002 | 0.001 | 0.0006018 | 0.004 | 0.0032 |
| c.1173C>T (p.Asp391=) | synonymous | rs2230244 | 0.0274 | 0.0003 | 0.008229 | 0.001 | 0.0032 |
| c.1182C>T (p.Thr394=) | synonymous | rs377462923 | 0.0002 | 0.0001 | 0.0001565 | NA | 0.0032 |
| c.1485C>T (p.Thr495=) | synonymous | rs2230245 | 0.077 | 0.1147 | 0.124 | 0.112 | 0.1419 |
| c.1494+198T>A | intronic | NA | NA | NA | NA | NA | 0.0032 |
| c.1495-109A>C | intronic | rs3219395 | 0.011 | NA | NA | NA | 0.0226 |
| c.1495-107C>T | intronic | NA | NA | NA | NA | NA | 0.0032 |
| c.1495-105C>T | intronic | rs559203182 | 0.0006 | NA | NA | NA | 0.0065 |
| c.1495-44A>G | intronic | rs3219396 | 0.018 | 0.0212 | 0.02073 | NA | 0.029 |
| c.1687-49G>A | intronic | rs368086982 | NA | NA | NA | NA | 0.0032 |
| c.1687-38C>T | intronic | rs373705242 | NA | 0.0008 | 0.0007298 | 0.002 | 0.0032 |

GMAF, global minor allele frequency; EVS, exome variant server; ExAC, exome aggregation consortium; NA, not available or not present in databases.
rs ID: variant identifier reported in dbSNP for this variant (http://www.ncbi.nlm.nih.gov/SNP/).
GMAF: Minor Allele Frequency in the 1000Genomes database (http://www.1000genomes.org/).
EVS: allelic frequency of the variant in European Americans in Exome Variant Server database (http://evs.gs.washington.edu/EVS/).
ExAC: allelic frequency of the variant in Exome aggregation consortium browser (http://exac.broadinstitute.org/).
CSVS: allelic frequency of the variant in CIBERER Spanish Variant Server (http://csvs.babelomics.org/).
Inner allelic frequency: Allelic frequency in our cohort of 155 patients (310 alleles).

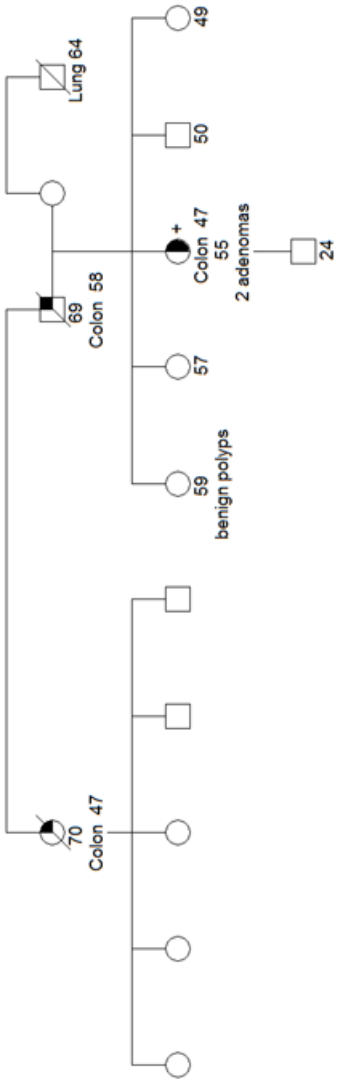

49
Lung 64
50
+
Colon 47
55
2 adenomas
24
69
Colon 58
57
59
benign polyps
70
Colon 47

Figure 1

Figure 2

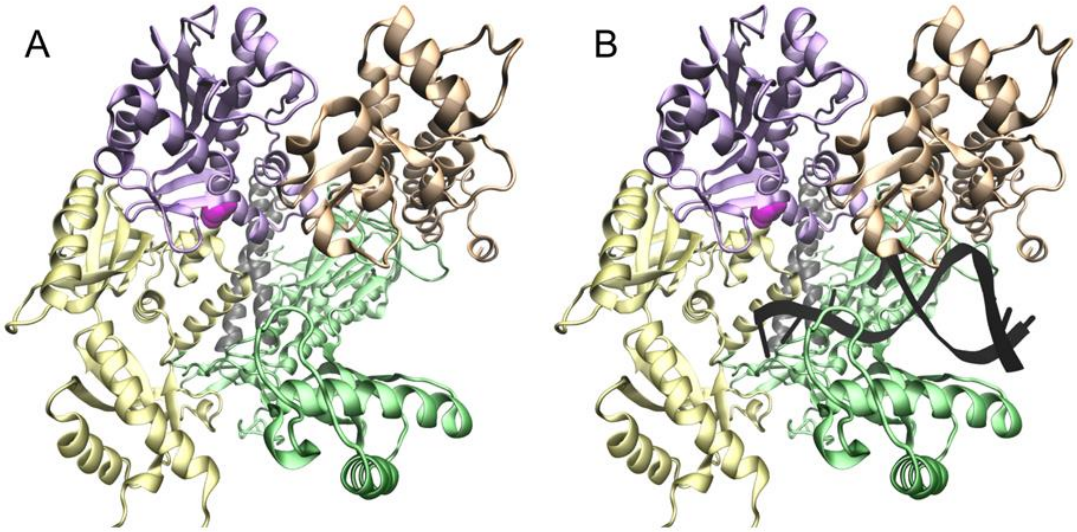
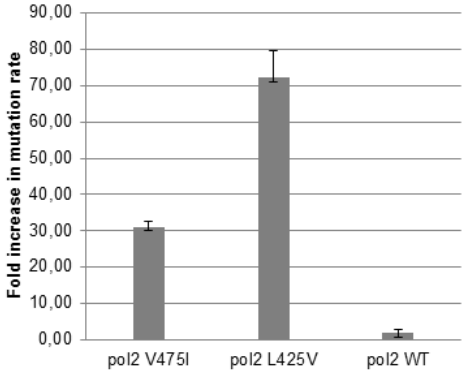


Figure 3



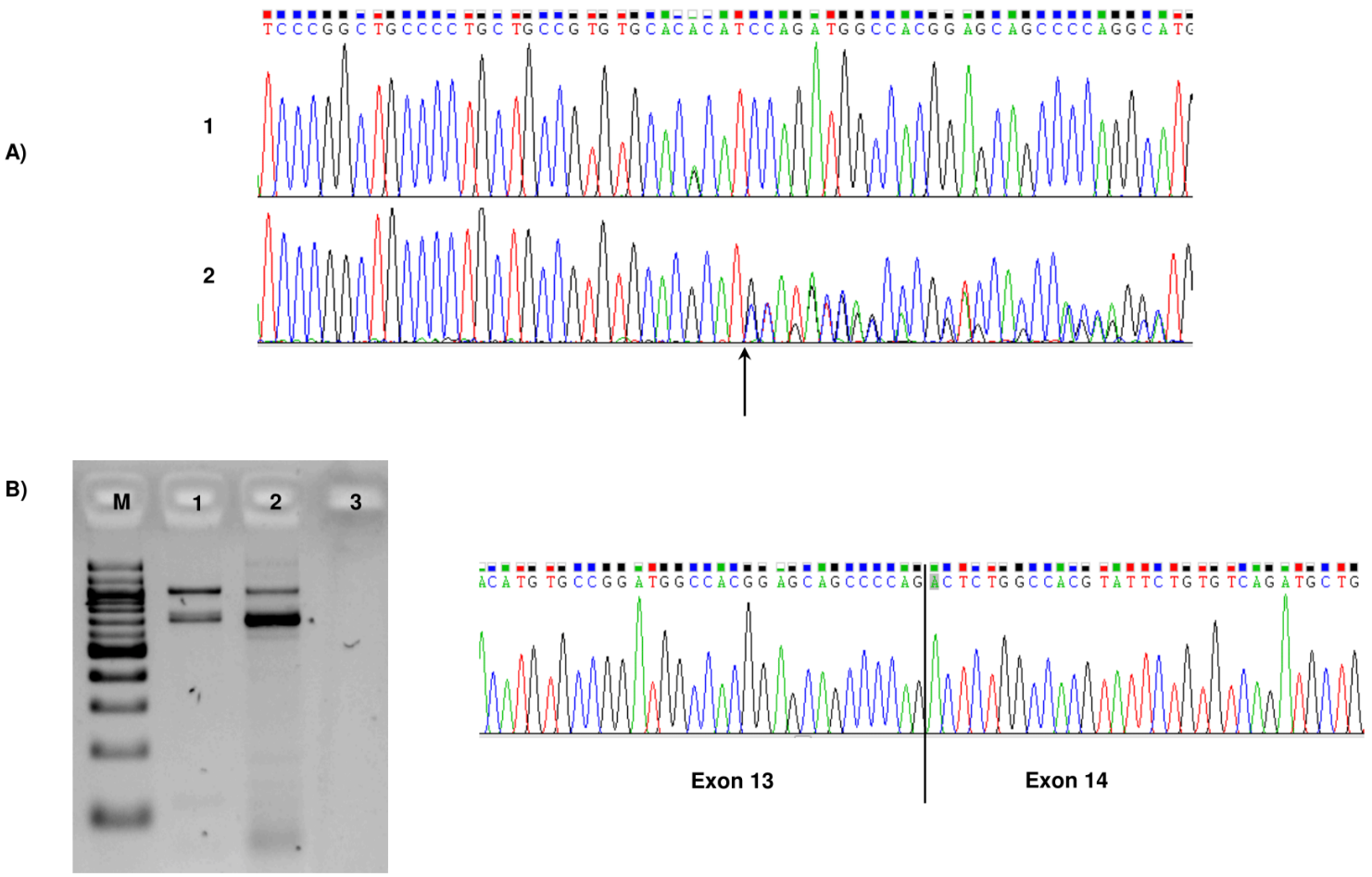

| Strain | Number of revertants (SD) | Fold change increase (SD) |
|---|---|---|
| *pol2-V475I* | 279.5 (13.44) | 31.06 (1.49) |
| *pol2-L425V* | 649.5 (65.76) | 72.17 (7.31) |
| *pol2* WT | 16 (9.90) | 1.78 (1.10) |

**Supplementary Table S1.** Rare intronic genetic variants tested for their putative involvement in abnormal splicing by using several bioinformatics tools.

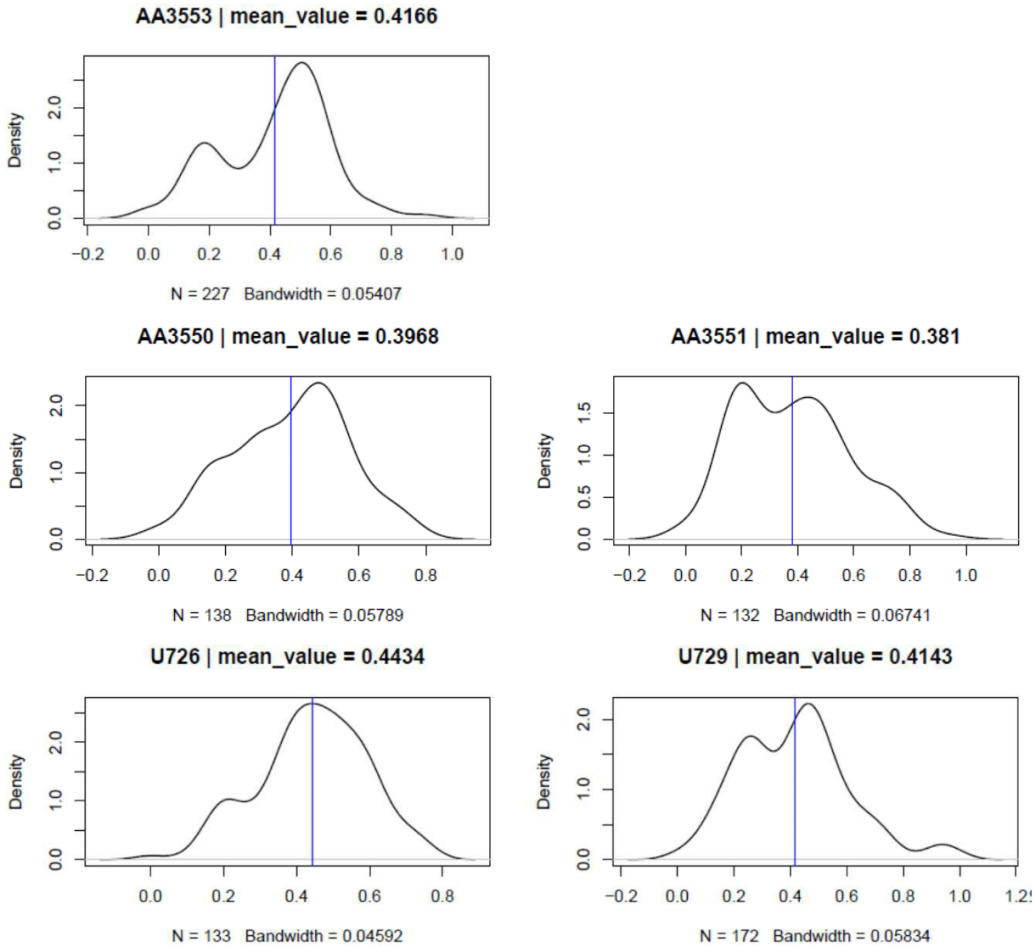
| Gene | Variant | HSF | BDGP | SPANR |
|---|---|---|---|---|
| *POLE* | c.1020+29C>T | no effect (0% variation) | no effect (0% variation) | no effect (exon11 dPSI -0.21, exon10 dPSI 0.63) |
| *POLE* | c.1020+46C>A | no effect (0% variation) | no effect (0% variation) | slight effect (exon11 dPSI 0.14, exon10 dPSI-5.73) |
| *POLE* | c.1226+13G>A | no effect (0% variation) | NA | no effect (exon12 dPSI 2.60) |
| *POLE* | c.1226+44G>A | no effect (0% variation) | no effect (0% variation) | no effect (exon 12 dPSI -0.56) |
| *POLE* | c.1226+45C>T | no effect (0% variation) | NA | no effect (exon12 dPSI 2.85) |
| *POLE* | c.1359+46del71 | no effect (0% variation) | no effect (0% variation) | NA |
| *POLE* | c.1359+144G>T | no effect (0% variation) | no effect (0% variation) | no effect (exon13 dPSI 2.49) |
| *POLD1* | c.970+79G>A | no effect (0.23% variation) | no effect (0% variation) | no effect (exon 8 dPSI -0.67, exon 9 dPSI 0.11) |
| *POLD1* | c.971-93G>C | no effect (0% variation) | no effect (0% variation) | no effect (exon 8 dPSI -0.55, exon 9 dPSI -0.12) |
| *POLD1* | c.1137+19C>G | no effect (0% variation) | no effect (0% variation) | no effect (exon 9 dPSI -0.14, exon 10 dPSI 0.07) |
| *POLD1* | c.1137+69G>A | no effect (0% variation) | no effect (0% variation) | no effect (exon 9 dPSI 0.14, exon 10 dPSI -0.36) |
| *POLD1* | c.1138-8A>G | no effect (0.1 %variation) | no effect (0% variation) | no effect (exon12 dPSI 0.07) |
| *POLD1* | c.1173C>T; (p.Asp391=) | no effect (0% variation) | no effect (0% variation) | no effect (exon10 dPSI -0.30) |
| *POLD1* | c.1182C>T; (p.Thr394=) | no effect (0% variation) | no effect (0% variation) | no effect (exon10 dPSI -0.20) |
| *POLD1* | c.1485C>T; (p.Thr495=) | no effect (0.7 % variation) | no effect (0% variation) | no effect (exon12 dPSI 1.87) |

| | | | | |
|---|---|---|---|---|
| *POLD1* | c.1494+198T>A | no effect (0% variation) | no effect (0% variation) | no effect (exon 12 dPSI -0.17, exon 13 dPSI 0.21) |
| *POLD1* | c.1495-109A>C | no effect (2.2% variation) | no effect (3% variation) | no effect (exon12 dPSI 0.17) |
| *POLD1* | c.1495-107C>T | no effect (0% variation) | no effect (0% variation) | no effect (exon13 dPSI -0.20) |
| *POLD1* | c.1495-105C>T | no effect (0.97%variation) | no effect (4% variation) | no effect (exon13 dPSI -0.20) |
| *POLD1* | c.1495-44A>G | no effect (0% variation) | no effect (0% variation) | no effect (exon13 dPSI -1.88) |
| *POLD1* | c.1687-49G>A | no effect (0% variation) | no effect (0% variation) | no effect (exon13 dPSI -2.44, exon14 dPSI -1.90) |
| *POLD1* | c.1687-38C>T | no effect (0% variation) | no effect (0% variation) | no effect (exon13 dPSI-0.84, exon14 dPSI 0.20) |

HSF, Human Splicing Finder; BDGP, Berkeley Drosophila Genome Project; SPANR, Splicing-based Analysis of Variants; dPSI, maximum difference across tissues in percentage of transcripts with the exon spliced; NA, not available.
Human Splicing Finder: Variants affecting splicing have a 10% of variation[26].
BDGP: Berkeley Drosophila Genome Project. Variants considered to affect splicing have a percentage of variation of 10%[27].
SPANR: Variants predicted to affect splicing have a dPSI above 5 or below -5[28].




**Supplementary Figure S1.** DNA and RNA analysis of the c.1359+46del71 variant in *POLE* intron 13. **A)** Sequencing result for POLE-3 PCR fragment in a control sample (1) and a carrier (2). The arrow indicates the presence of the deletion. **B)** RNA analysis using the POLE DEL primers located in exon 11 and exon 16 in a control sample (1), a carrier (2) and negative control (3). The main PCR amplification corresponds to the expected size (697 base pairs). Sequencing of the main band yielded no alterations and showed intact exons 12 and 13 in the carrier as shown. An additional band of a bigger size was also present in both control and carrier corresponded to an alternative splicing including intron 15. (M) 100-base pair molecular weight marker.




**Supplementary Figure S2.** Density plots for variants detected by WES in the five FFPE tumor DNA samples analyzed after filtering those present in a database of 65 germline exomes and normalizing by selecting only those sequenced regions with coverage above 10X in all samples. The X axis shows the proportion of all reads that present the alternative allele for all variants. Mean value has usually a proportion around 0.4 in tumor tissue (when tumor purity is 80%), whereas it corresponds 0.5 in germline tissue. These results will be in agreement of an appropriate germline filtering in our tumor dataset.

**Supplementary Table S2.** Normalized number of heterozygous variants not present in a germline exome dataset that were found in FFPE tumor DNA from p.Val474Ile variant carrier (AA3553) and four colorectal tumor samples without alterations in *POLE* or the MMR system (AA3550, AA3551, U726, U729). The normalization was applied counting only those variants situated in genomic regions with coverage above 10X in all samples. Mutation spectrum is expressed in percentages of each type of substitution divided by the total number of substitutions. WT=wild-type.

| Number of variants | AA3553 (*POLE* p.Val474Ile) | AA3550 (*POLE* WT) | AA3551 (*POLE* WT) | U726 (POLE WT) | U729 (*POLE* WT) |
|---|---|---|---|---|---|
| C>T | 47 | 30 | 22 | 26 | 24 |
| G>A | 50 | 32 | 34 | 31 | 41 |
| C>T / G>A | 97 | 62 | 56 | 57 | 65 |
| A>G | 22 | 13 | 18 | 15 | 23 |
| T>C | 31 | 17 | 13 | 18 | 23 |
| A>G / T>C | 51 | 30 | 31 | 33 | 46 |
| C>G | 6 | 8 | 7 | 10 | 14 |
| G>C | 10 | 7 | 4 | 8 | 7 |
| C>G / G>C | 16 | 15 | 11 | 18 | 21 |
| G>T | 11 | 3 | 4 | 2 | 6 |
| C>A | 5 | 3 | 6 | 4 | 11 |
| G>T / C>A | 16 | 6 | 10 | 6 | 17 |
| A>C | 10 | 12 | 6 | 4 | 6 |
| T>G | 6 | 2 | 0 | 3 | 6 |
| A>C / T>G | 16 | 14 | 6 | 7 | 12 |
| A>T | 12 | 7 | 7 | 8 | 5 |
| T>A | 17 | 4 | 11 | 4 | 6 |
| A>T / T>A | 29 | 11 | 18 | 12 | 11 |
| Total number of substitutions | 227 | 138 | 132 | 133 | 172 |

| Mutation spectrum (%) | AA3553 (*POLE* p.Val474Ile) | AA3550 (*POLE* WT) | AA3551 (*POLE* WT) | U726 (POLE WT) | U729 (*POLE* WT) |
|---|---|---|---|---|---|
| C>T / G>A | 42.73 | 44.93 | 42.42 | 42.86 | 37.79 |
| A>G / T>C | 23.35 | 21.74 | 23.48 | 24.81 | 26.74 |
| C>G / G>C | 7.05 | 10.87 | 8.33 | 13.53 | 12.21 |
| G>T / C>A | 7.05 | 4.35 | 7.58 | 4.51 | 9.88 |
| A>C / T>G | 7.05 | 10.14 | 4.55 | 5.26 | 6.98 |
| A>T / T>A | 12.78 | 7.97 | 13.64 | 9.02 | 6.40 |

**Supplementary Table S3.** Custom DNA oligonucleotides.

Sequencing primers for the exonuclease domain of *POLE* and *POLD1*

| GENE | FRAGMENT | Forward primer sequence 5'-3' | Reverse primer sequence 5'-3' | PCR size (bp) |
|---|---|---|---|---|
| *POLE* | POLE-1 | ATGGGGAGTTTAGAGCTTGG | ACGGTCATACCCTGAGAACA | 736 |
| | POLE-2 | CCATGAGCTTTGTTCTCAGG | CTGCCATACTCTTGGGTGAC | 738 |
| | POLE-3 | TCCTGATGAAGGGTTTTCAC | CCGACAGGACAGATAATGCT | 742 |
| *POLD1* | POLD1-1 | CTCACTTCTCCGGCCTCTAT | TTCTCGTAGCTCTGCACCTT | 814 |
| | POLD1-2 | CCTGTCATCCAGATCTGCTC | GCCTCTGATAAAAGCCACAG | 718 |
| | POLD1-3 | CTCAATCTCCGTTCTTCAGG | ACAGAGGTCACAGGAGTGGA | 602 |
| | POLD1-4 | CCACGTCTGACCTCACTCTT | GAGAAGGTGGGAAATGGAGT | 716 |

Primers for LOH analysis in FFPE tumor DNA of the p.V474I variant carrier

| GENE | FRAGMENT | Forward primer sequence 5'-3' | Reverse primer sequence 5'-3' | PCR size (bp) |
|---|---|---|---|---|
| *POLE* | POLE-3 FFPE | TCTGTGCTTCACACTTGACC | CCGACAGGACAGATAATGCT | 206 |

Primers for cDNA sequencing in *POLE* (exons 11-16)

| GENE | FRAGMENT | Forward primer sequence 5'-3' | Reverse primer sequence 5'-3' | PCR size (bp) |
|---|---|---|---|---|
| *POLE* | POLE DEL | CATCATGGTCACCTACAACG | AGGCACTTTCTCCTCTTCCT | 697 |

Primers for functional assessment in *S.pombe*

| GENE | FRAGMENT | Forward primer sequence 5'-3' | Reverse primer sequence 5'-3' | PCR size (bp) |
|---|---|---|---|---|
| *Pol2* | POL2 WT | CGCGCGAGATCTCTGATTCGTTTAAGCTTTTTCAGTTAATGGTGG | CTAAAGGCGCGCCGGTACATTGAGGCGACATCAAGATGG | 2593 |
| | POL2-L425V | GGGTAAAGAGAGATAGTTATTTACCTCAAGGAAGTCAAGGTGTCAAAGCTGTCACTGTCAGTAAATTAGGTTATAATCC | GGATTATAACCTAATTTACTGACAGTGACAGCTTTGACACCTTGACTTCCTTGAGGTAAATAACTATCTCTCTTTACCC | 2593 |
| | POL2-V475I | TTTCAGATGCTGTTGCTACTTATTTTCTTTACATGAAATATATCCATCCTTTCATTTTTTCTCTTTGTAACATTATCCC | GGGATAATGTTACAAAGAGAAAAAATGAAAGGATGGATATATTTCATGTAAAGAAAATAAGTAGCAACAGCATCTGAAA | 2593 |

bp, base pairs

# DISCUSIÓN

El objetivo general de esta tesis doctoral es la búsqueda de la causa de predisposición genética germinal en pacientes con cáncer colorrectal familiar. El primer apartado de esta tesis, consistente en los dos primeros trabajos publicados, está enfocado en la identificación y caracterización de nuevos genes de predisposición al CCR mediante secuenciación del exoma completo de familias con una fuerte agregación para la enfermedad y sin alteraciones en *APC*, *MUTYH* o los genes del síndrome Lynch. Por otro lado el segundo bloque de esta tesis constituye el tercer artículo presentado, que reporta el cribado de los recientemente descubiertos genes de predisposición *POLE* y *POLD1* en una cohorte española de 155 pacientes con poliposis y CCR de aparición temprana, incluyendo la caracterización funcional de las variantes encontradas.

## 1- Identificación de nuevos genes de predisposición mediante secuenciación del exoma

### 1.1- Tecnología utilizada

En la primera parte de esta tesis se han buscado variantes raras de alta penetrancia en nuevos genes candidatos de predisposición a CCR mediante la secuenciación del exoma completo en familias con una fuerte agregación para esta neoplasia pero sin causa genética conocida. La secuenciación del exoma está dirigida a secuenciar las regiones codificantes del genoma. Otra opción para detectar la causa genética de predisposición a CCR en estas familias sería secuenciar el genoma completo. No obstante, esta elección implica un coste mayor, lo que limita el número de muestras a secuenciar. La ventaja de secuenciar todo el genoma frente al exoma es que permite la detección de las variantes situadas en zonas reguladoras alejadas de la región codificante. No obstante, al ser el objetivo la búsqueda de variantes raras causantes de CCR familiar, la secuenciación del exoma es una estrategia efectiva, pues se ha demostrado que el genoma codificante está enriquecido en variantes causantes de enfermedades mendelianas (Botstein 2003).

### 1.2- Selección de pacientes

A la hora de realizar un estudio de secuenciación de nueva generación con el fin de detectar nuevos genes implicados en una enfermedad, un punto crítico a discutir del diseño experimental es la selección de pacientes. Para la búsqueda de variantes de alta penetrancia en genes implicados en CCR familiar, se puede optar entre diferentes estrategias, entre ellas secuenciar familias con una fuerte agregación para la enfermedad (Gylfe 2013, DeRycke 2013, Nieminen 2014, Schulz 2014, Seguí 2015, Thutkawkorapin 2016) o individuos no emparentados con CCR de aparición temprana (Smith 2013, Taskanen 2015, De Voer 2016), pues ambas

opciones serían reflejo de la existencia de predisposición genética germinal. En nuestro primer estudio de secuenciación se seleccionaron 29 familias con una fuerte agregación para CCR, secuenciando uno o dos pacientes por familia, llegando a un total de 42 individuos secuenciados. El motivo por el cual se decidió secuenciar varios miembros de una misma familia es la alta heterogeneidad genética del CCR hereditario y de la población española. Esto significa que varios genes distintos causan la enfermedad, lo cual se ve reflejado en que en la mayoría de estudios realizados hasta la fecha no existe solapamiento de los genes candidatos encontrados (Smith 2013, DeRycke 2013, Gylfe 2013, Tanskanen 2015, Spier 2016, De Voer 2016, Thutkawkorapin 2016). Otro tipo de heterogeneidad genética es la variabilidad que existe dentro de una población (McClellan J 2010). La población española es bastante heterogénea (Chillón 1994), por tanto será difícil encontrar variantes candidatas en común en familias distintas así como validar los resultados en cohortes caso-control.

Por otra parte, la astringencia de los criterios de selección puede causar que se elija un número mayor o menor de pacientes a secuenciar dentro de cada familia. Con el fin de reducir el número de variantes candidatas, es mejor secuenciar el máximo de individuos afectados por familia disponible. No obstante, secuenciar muchos familiares afectos aumenta la probabilidad de incluir fenocopias, que alterarían el análisis de datos. Esta probabilidad se reduce eligiendo individuos afectos con una edad de aparición lo más temprana posible. En nuestro caso el promedio de edad de los pacientes secuenciados era de 60,32 años (rango 30-81 años). Un 35 % de los pacientes secuenciados es mayor de 65 años en nuestra cohorte, aunque algunos de estos tienen descendencia afectada a edades más jóvenes, por lo que se asume que son portadores obligados de la variante de predisposición. La mayoría de estudios de secuenciación (con excepción de Palles 2013 y Thutkawkorapin 2016) no han combinado una edad de aparición temprana y una fuerte agregación familiar en los pacientes secuenciados (**Tabla 9**).

La cohorte de estudio actualmente está formada por 75 individuos pertenecientes a 40 familias (4 pacientes afectos de 2 familias, 3 pacientes afectos de 8 familias, 2 pacientes afectos de 15 familias y 16 pacientes con CCR no emparentados). En todos los pedigrees encontramos una fuerte agregación familiar para CCR, pero en varias familias, especialmente en las que se ha secuenciado un solo individuo, no se han encontrado variantes candidatas de predisposición, debido a que el número de variantes después de filtrar era demasiado alto. No obstante, contando ambos estudios, hemos encontrado variantes candidatas en nuevos genes de predisposición en 13 de 40 familias.

**Tabla 9. Criterios de selección de pacientes en estudios de secuenciación para descubrir las causas genéticas de predisposición a CCR familiar no polipósico.** La columna de agregación familiar hace referencia a si los individuos secuenciados pertenecen a familias con agregación para CCR o bien son casos esporádicos. La columna "Familias/Individuos" describe si dentro de un núcleo familiar se secuencia un solo paciente o por el contrario se secuencian dos o más pacientes.

| Autores | Número de pacientes | Edad pacientes | Agregación familiar | Familias/ Individuos | Genes identificados |
|---|---|---|---|---|---|
| Palles y col 2013 | 20 pacientes (15 familias) | CCR aparición temprana <50 años | si | Familias e individuos | *POLE* y *POLD1* |
| Smith y col 2013 | 50 | 18/50 pacientes con <35 años | no | Individuos | *FANCM, LAMB4, PTCHD3, LAMC3* y *TREX2* |
| DeRycke y col 2013 | 40 (16 familias) | Promedio <70 años (rango 25-79) | si | Familias | *CENPE* y *KIF23* |
| Gylfe y col 2013 | 96 | Promedio 70 años, rango 32-90 años | Si (uno o más familiares de primer grado afectados) | Individuos | *UACA, SFXN4, TWSG1, PSPH, NUDT7, ZNF490, PRSS37, CCDC18, PRADC1, MRPL3,* y *AKR1C4.* |
| Nieminen y col 2014 | 4 (1 familia) | edad promedio 52,3 años | si | Una familia con 4 pacientes secuenciados | *RPS20* |
| Schulz y col 2014 | 4 (1 familia) | (CCR con 55, 62, 71 y desconocido) | si | Una familia | *SEMA4A* |
| Weren 2015b | 51 individuos con poliposis (48 familias) | No especificada | En el 53% de los casos | Individuos sueltos menos 2 familias | *NTHL1* |
| Tanskanen y col 2015 | 22 | CCR aparición temprana <40 años | Sólo en 14% de los casos | Individuos | *ADAMTS4, CYTL1, SYNE1, MCTP2, ARHGAP12, ATM, DONSON* y *ROS1* |
| Seguí y col 2015 | 3 (1 familia) | 3 individuos con CCR a los 46, 76 y 82 años | si | Una familia | *FAN1* |
| De Voer y col 2016 A | 55 | CCR aparición temprana <45 años | No especificado | Individuos | *PTPN12* y *LRP6* |
| De Voer y col 2016 B | 55 | CCR aparición temprana <45 años | No especificado | Individuos | *BLM* |
| Thutkawkorapin y col 2016 | 3 (1 familia) | CCR aparición temprana (edad no especificada) | si | Una familia con 3 pacientes secuenciados | *DZIP1L, PCOLCE2, IGSF10, SUCNR1, OR13C8, EPB41L4B, SEC16A, NOTCH1, TAS2R7, SF3A1, GAL3ST1* y *TRIOBP* |

Finalmente, secuenciar personas con neoplasias extracolónicas o adenomas avanzados en familias con agregación para CCR puede ser una buena estrategia siempre que la edad del diagnóstico sea temprana y exista herencia mendeliana que sugiera una causa de predisposición genética común en los distintos tumores, ya que un mismo gen de

predisposición puede presentar pleiotropía en cuanto a las neoplasias que causa. Por ejemplo, mutaciones en el gen *POLE* predisponen a adenomas múltiples, CCR, endometrio, ovario, cerebro, páncreas, intestino delgado y melanoma cutáneo (Rholin 2014, Hansen 2015, Bellido 2015, Aoude 2015). Por otro lado, es conveniente no secuenciar individuos sanos debido a que se ha observado penetrancia incompleta en una fracción importante de genes de predisposición, como es el caso de *PMS2* que presenta una penetrancia del 15-20% para CCR (Lynch 2009).

**1.3- Implementación de la metodología de anotación y filtrado de variantes.**

El diseño de una metodología de filtrado automática mediante un programa bioinformático permite manejar los datos de secuenciación de nueva generación de forma más objetiva, permitiendo que los resultados no cambien en función de la persona que realiza el análisis. Esto tiene mayor probabilidad de ocurrir si una persona realiza los filtrados de forma manual, ya cada persona puede mostrar preferencia por seleccionar distintas variantes, aplicando inconscientemente filtros más laxos en las variantes de su elección. Otra ventaja es que se elimina el error humano causado por tener que tratar con grandes cantidades de datos. Además es una forma rápida y reproducible de obtener los resultados deseados, ya que se pueden añadir nuevas muestras (nuevas familias o individuos adicionales de familias ya secuenciadas) y se puede repetir el análisis siempre que se necesite. Otro punto a favor del diseño de una metodología de filtrado automática es que se trata de una herramienta flexible, en la cual se pueden introducir y quitar filtros según de forma que se puede adaptar a diferentes cohortes de pacientes (por ejemplo con herencia dominante o recesiva para la enfermedad).

Pese a estas ventajas, el decidir qué filtros establecer será crítico para seleccionar la variante de predisposición entre los miles de falsos positivos mostrados junto a esta candidata final. La mayoría de estos filtros son comunes en cualquier estudio de exoma germinal en familias con agregación para una enfermedad rara con herencia mendeliana. Estos filtros serían la cobertura mínima de secuenciación, el genotipo heterocigoto u homocigoto (según el patrón de herencia), la baja frecuencia en bases de datos de variación genética, el estar compartidas entre los miembros afectos de la misma familia y la eliminación de variantes sinónimas y alejadas de la región codificante (Gilissen 2012).

Uno de los filtrados aplicados que es exclusivo de nuestra metodología consiste en la selección de variantes situadas en genes cuya función biológica coincida con una lista de términos clave que incluye funciones biológicas y términos bibliográficos relacionados con cáncer. Esta

selección no se restringió únicamente a genes previamente implicados en cáncer sino que también incluyó términos describiendo genes con una función biológica que al afectarse pudiera explicar predisposición a cáncer. Es una selección menos limitada y sesgada que en el caso de otros estudios donde se utilizó para filtrar una lista de genes candidatos (Smith 2013). No obstante, en la selección con la lista de términos usada, si la mutación causal estuviera en un gen sin absolutamente ninguna información disponible, esta mutación se perdería al filtrar. Se podría decir que es un riesgo asumible, ya que para la mayoría de genes de predisposición recientemente reportados ya existía cierta información de su función biológica previa a su descubrimiento como gen de CCR hereditario. Es decir, algunas de las funciones biológicas de estos genes podían relacionarse con la tumorigenesis aún sin existir evidencia de que estuvieran efectivamente implicados en esta enfermedad (**Tabla 10**).

**Tabla 10. Tabla de términos clave de nuestra lista anotados para los nuevos genes de predisposición a CCR hereditario/familiar.** Todos los genes excepto *SEMA4A* habrían pasado el filtrado usando nuestra metodología. Los términos marcados con asterisco se añadieron a la lista de términos clave posteriormente a la publicación del trabajo, debido a que en ese momento no poseíamos un conocimiento suficientemente amplio de la nomenclatura de los distintos tipos de tumores existentes y algunos términos fueron pasados por alto.

| Gen | Términos clave de función biológica | Términos clave bibliográficos (previos a su descubrimiento como gen de CCR hereditario) |
|---|---|---|
| *GREM1* | Apoptosis, regulación de la vía de BMP, regulación del crecimiento celular, regulación de la vía de Wnt, | Carcinoma, cáncer |
| *POLE/ POLD1* | Mitosis, reparación por escisión de bases, mantenimiento de los telómeros, reparación de las brechas de doble hebra, | Cáncer, meningioma* |
| *RPS20* | -- | Meduloblastoma* |
| *NTHL1* | Reparación por escisión de bases, reparación por escisión de nucleótidos, | Cáncer |
| *SEMA4A* | -- | -- |
| *FAN1* | Reparación del ADN, reparación por escisión de nucleótidos, | Cáncer |
| *BLM* | reparación de las brechas de doble hebra, punto de control de la mitosis, ciclo celular | Cáncer |

La especificidad de esta lista de términos clave para detectar variantes implicadas en CCR se comprobó parcialmente al aplicarse en un set de exomas no afectados de CCR secuenciados y filtrados con la misma metodología, detectándose un promedio de variantes por individuo

significativamente superior en los individuos con CCR familiar que en los individuos del set de exomas externo (promedio cohorte CCR = 41,87 variantes por familia; promedio set de exomas externo = 34,05 variantes por familia; $P = 3{,}75 \times 10^{-10}$).

Por último cabe destacar que esta metodología de filtrado automática se ha utilizado en ambos estudios de secuenciación reportados en esta tesis, variando únicamente el hecho de que en el trabajo más reciente se ha añadido nueva información disponible para realizar los filtrados como la herramienta de predicción bioinformática CADD (referenciar) y la base de datos de variación genética ExAC (referenciar).

**1.4- Priorización de las variantes**

En ambos estudios de secuenciación aquí presentados al terminar el proceso consistente en el filtrado de variantes, el número de cambios genéticos seleccionados era todavía demasiado alto en la mayoría de familias para distinguir la posible causa de predisposición genética a CCR. Para seleccionar la variante o variantes con más probabilidad de ser la causante de la enfermedad es necesaria entonces una priorización basada en un análisis detallado de las variantes. Algunas familias tuvieron hasta cuatro variantes candidatas priorizadas mientras que en otras no se encontró ninguna. En nuestro estudio hemos usado los criterios de priorización (frecuencia, predicciones *in silico*, posición del residuo afectado en la estructura proteica, interacciones proteicas, implicación previa en cáncer) de forma uniforme en todas las variantes candidatas, para reducir el riesgo de elegir de manera subjetiva el gen que nos parezca más atractivo. No obstante, en este paso existe un riesgo importante de descartar la verdadera variante causante de predisposición por no cumplir estrictamente todas las condiciones mencionadas anteriormente. Este riesgo se puede reducir en cierto modo si existen miembros adicionales afectados para comprobar si las variantes segregan correctamente, lo que aumenta la probabilidad de descartar falsos positivos. La evidencia de haber seleccionado un verdadero gen candidato aumenta cuando se encuentra un enriquecimiento de mutaciones deletéreas en individuos afectos adicionales en otras cohortes en comparación con controles (Lu 2015). En última instancia para establecer el papel de una variante candidata en el mecanismo de predisposición a CCR, siempre será necesario realizar estudios funcionales.

**1.5- Nuevos genes candidatos de predisposición a cáncer colorrectal**

En el primer estudio de esta tesis al terminar el proceso de filtrado y priorización, 29 variantes candidatas finales fueron seleccionadas, de las cuales 28 fueron validadas mediante

secuenciación Sanger. Después de estudiar la segregación en familiares afectos adicionales con CCR y adenomas avanzados, se descartaron aquellas variantes que no segregaban bien en con la enfermedad en esas familias. Finalmente, siete variantes fueron seleccionadas por estar presentes en dos o más individuos afectos dentro de cada familia: *CDKN1B* p.Q65H, *XRCC4 p.V166Efs*3, EPHX1 p.R98Q, NFKBIZ* p.*719Ifs*10, *SMARCA4* p.R99W y *BARD1* c.1811-2A>G. Cabe mencionar que las tres primeras variantes mencionadas mostraron LOH del alelo normal en el tumor.

En primer lugar, la proteína CDKN1B (p27, Kip1) se une los complejos ciclina E/A-CDK2 y ciclina D-CDK4 y previene su activación, controlando así la progresión del ciclo celular. La variante c.195G>T (p.Q65H está situada en un barril beta que interacciona con CDK2 (*Russo 1996)*. Por lo tanto esta variante es susceptible de afectar la interacción entre CDKN1B y CDK2, desregulando la progresión del ciclo celular. Mutaciones germinales en este gen se han implicado previamente en neoplasia endocrina múltiple (Pellegata 2006).

Por otro lado, el gen *XRCC4* participa junto con *LIG4* y *XLF* en la reparación de las brechas de doble hebra en el ADN. La variante c.497_498delTG (p.V166Efs*3), se predice que abolirá la función proteica y provocará así la inestabilidad del genoma como mecanismo oncogénico. Recientemente se ha descubierto que mutaciones bialélicas en este gen provocan un fenotipo de estatura baja, microcefalia y alteraciones del sistema endocrino (Shaheen 2014, de Bruin 2015, Murray 2015, Bee L 2015, Rosin 2015, Guo 2015). En una pequeña fracción de los casos además del fenotipo mencionado se presentaron neoplasias a edades tempranas, incluyendo un tumor gastrointestinal a los 28 años (de Bruin 2015) y un glioma a los 19 años (Guo 2015), por lo que se podría sugerir una penetrancia moderada para *XRCC4*.

El tercer gen candidato a discutir es *EPHX1, que* codifica para una enzima que transforma los epóxidos producidos por la degradación de los compuestos aromáticos a otros compuestos que puedan ser conjugados y excretados del cuerpo (Fretland 2000). Debido el efecto diferencial de *EPHX1* en la detoxificación de compuestos cancerígenos se ha propuesto que mutaciones en este gen pueden tener un efecto sobre el riesgo de padecer cáncer, debido a que una detoxificación deficiente de los substratos de EPHX1 puede producir daños en el ADN celular (Liu 2012). Un estudio posterior al aquí discutido reportó un polimorfismo en *EPHX1* que confería una mayor actividad enzimática y asociado a una mayor protección frente al daño en el ADN de las células, con una OR igual a 0,67 (0,45-0,98, P valor 0,08) (Hemminki 2015).

Por otra parte, el gen *NFKBIZ* está implicado en la respuesta inflamatoria a través de la regulación de *NFκB* (Totzke 2006). La variante c.2153_2154dupAT (p.*719Ifs*10) elimina un

codón de stop, produciendo un extremo C-terminal anormalmente largo. Esto puede afectar la interacción con los complejos de NFκB que se unen a esa región, alterando la regulación transcripcional de sus genes diana y contribuyendo a la aparición del cáncer.

El quinto gen a discutir de los seis candidatos es *SMARCA4*, que codifica para un componente del complejo de remodelación de la cromatina SNF/SWI, necesario para activar la transcripción de genes que normalmente están epigenéticamente reprimidos (Wilson 2011). La variante c.295C>T p.R99W está localizada en la región que interacciona con SS18L1, la cual inhibe la transcripción de *c-FOS* y es requerida para el crecimiento dendrítico y la ramificación en las neuronas corticales. Puede hipotetizarse que esta variante puede causar predisposición a CCR mediante la desregulación de la vía de señalización de c-*FOS* causando una proliferación celular anormal. Mutaciones germinales en este gen pueden causar síndrome de predisposición a los tumores rabdoides (Schneppenheim 2010) y cáncer las células pequeñas del ovario de tipo hipercalcémico (Witkowski 2014). Además se ha descrito que ratones mutantes heterocigotos para *SMARCA4* desarrollan tumores mamarios (Bultman 2007).

Finalmente, BARD1 interacciona con BRCA1 y participa en distintas vías de señalización implicadas en la reparación del daño en el ADN, la ubiquitinización y la regulación de la transcripción, con el fin de preservar la integridad del genoma. La variante c.1811-2A>G se predice que causa la eliminación del exón 9, destruyendo parte del dominio BRCT1. Este dominio presenta una gran homología con el dominio BRCT1 de *BRCA1*, que se encuentra mutado germinalmente en cáncer, sugiriendo que el dominio BRCT otorga la función de supresor tumoral (Williams y Glover 2003) (**Figura 27**). Mutaciones germinales en el gen *BARD1* predisponen a cáncer de mama y ovario (Ratajska 2012). Estudios posteriores refuerzan el papel de este gen como causante de predisposición a CCR, ya que describen mutaciones adicionales en pacientes afectos de CCR familiar (Yurgelun 2015).

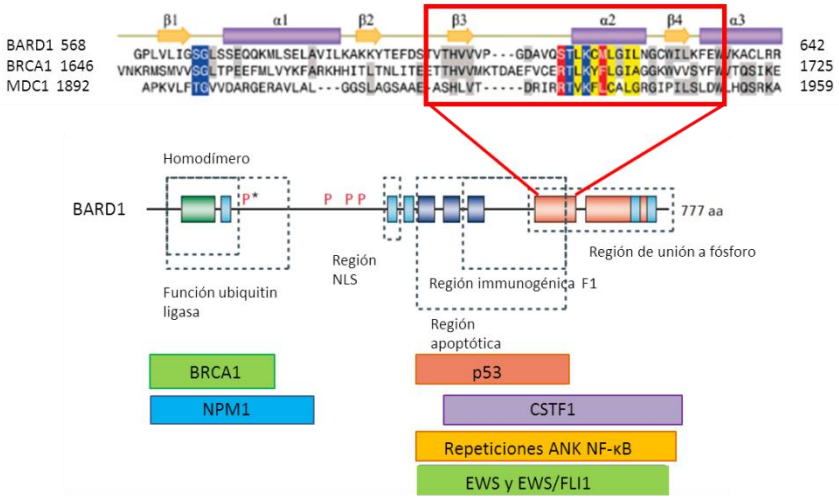



**Figura 27.** Regiones de la proteína BARD1. Los dominios RING (verde), ankirina (ANK, azul) y BRCA1 carboxi-terminal (BRCT, rojo) son mostrados. La posición de serinas o treoninas fosforiladas se indica con la letra P. Los residuos fosforilados que compiten por la unión con BRCA1 y la función ubiquitin ligasa son representados como P*. Las proteínas que interaccionan con BARD1 se muestran en forma de barras cubriendo de forma aproximada la región de BARD1 participante en la interacción. La estructura primaria del dominio BRCT1 deleccionado por la variante c.1811-2A>G se muestra arriba. Esta región presenta una alta homología con el dominio BRCT1 de BRCA1 y MDC1 y es necesaria para la interacción con p53, CSTF1, proteínas del complejo NFκB y EWS. Figura adaptada de Irminger-Finger 2006. CSTF1, cleavage stimulation factor subunit 1 50kDa; EWS, Ewing sarcoma gene product; IkappaBa, NF-kappaB inhibitor, alpha.

Aparte del descubrimiento de los seis nuevos genes candidatos es interesante mencionar que se ha encontrado en un paciente con CCR sin pólipos la misma mutación en el gen *BMPR1A* c.1327C>T (p.R443C) que en dos casos anteriormente reportados de poliposis juvenil (Sayed 2002, Greenman 2007). El gen *BMPR1A* ya había sido implicado anteriormente en CCR hereditario no polipósico por Nieminen y col., que encontraron una mutación alteradora del splicing en 2 pacientes con CCR familiar de tipo X (Nieminen 2011). Este sería el segundo estudio donde se encuentra una variante patogénica en *BMPR1A* en una familia de CCR no polipósico. Por último, este sería el tercer estudio en el que se detectan variantes germinales de pérdida de función en el gen *AKR1C4,* reforzando el papel de este gen como candidato de predisposición a CCR familiar (Smith 2013, Gylfe 2013).

En el segundo artículo de esta tesis se aumentó la cohorte a estudiar hasta 74 individuos de 40 familias. Esto permitió reportar que en seis pedigrees después de filtrar y priorizar las variantes el mejor gen candidato pertenecía a la vía de reparación del daño en el ADN de la anemia de Fanconi. Esta vía de señalización es necesaria para restaurar la integridad del genoma mediante la reparación de los entrecruzamientos entre hebras de ADN *(interstrand cross-links*) (**Figura 28**).

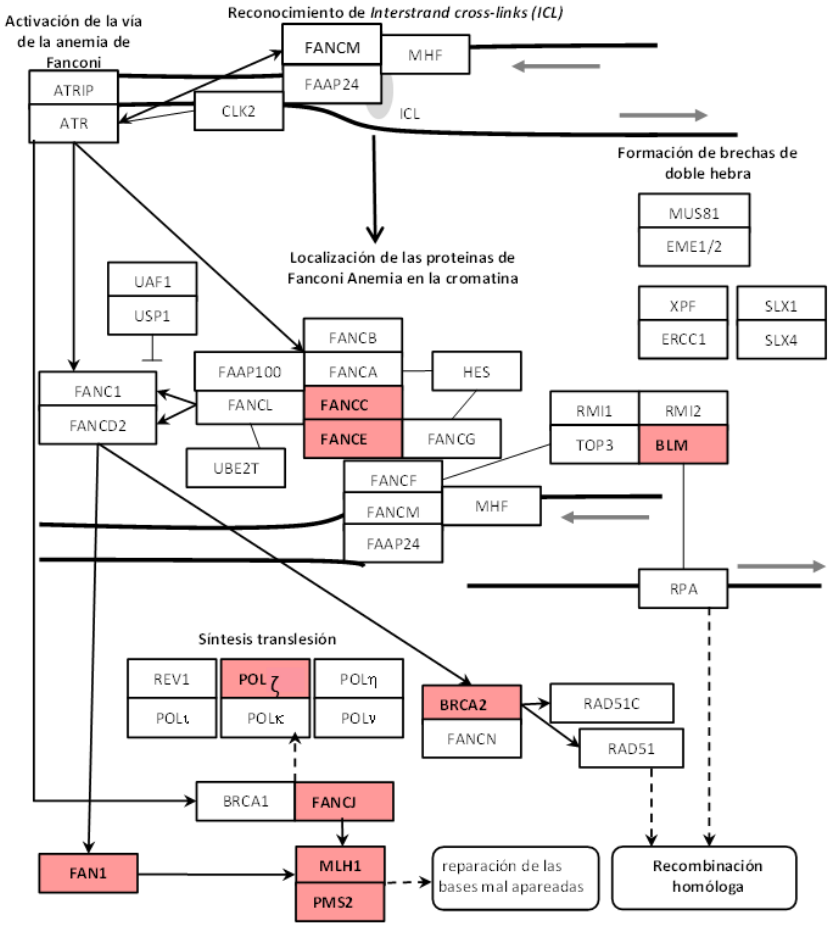



**Figura 28. Vía de reparación de daños en el ADN de la anemia de Fanconi.** Los genes de esta vía de señalización implicados en CCR por este estudio y otros (Papadopoulos 1994, Nicolaides 1994, Garre 2015, Seguí 2015, de Voer 2015) están marcados en rojo.

Encontramos tres variantes frameshift: *BRCA2/FANCD1* c.4963delT p.(Y1655fs*15), *BRIP1/FANCJ* c.1702_1703delAA p.(N568fs*9) y *FANCC* c.591_592dupC p.(L199fs*12), así como tres variantes missense: *BRCA2/FANCD1* c.7759C>T p.(L2587F), *FANCE* c.598C>T p.(R200C) y *REV3L/POLZ* c.559A>T p.(R187W). Al comparar nuestra cohorte de 40 familias con una base de datos de exomas de controles *(Exome Aggregation Consortium)* encontramos un enriquecimiento en el número de variantes raras con predicciones deletéreas en estos genes. Las seis familias portadoras en nuestra cohorte presentan agregación para un amplio espectro de neoplasias aparte de CCR, incluyendo estómago, mama, útero, próstata, pulmón, vejiga y leucemia, así como adenomas avanzados. Un estudio reciente demostró que variantes germinales de pérdida de función en *BRCA2* y *BRIP1* se encontraban sobrerrepresentadas respecto a controles en muchas neoplasias distintas, lo cual refuerza la evidencia de que para estos genes existe pleiotropía en el espectro tumoral que provocan (Lu 2015). El modelo tradicional de cáncer hereditario engloba la identificación de individuos cuyas historias

cumplen criterios clínicos para un síndrome específico, seguidas del cribado mutacional dirigido a los genes asociados con ese síndrome (Domcheck 2013). El uso de paneles de genes implicados en cáncer germinal puede ser la solución a este cambio de paradigma en que un que un número creciente de genes puede implicarse en un amplio espectro de neoplasias. A este respecto, un estudio consistente en la secuenciación de un panel germinal de 25 genes en 1.260 pacientes con CCR hereditario permitió la identificación de 182 pacientes con mutaciones patogénicas. Además de los genes de reparación del ADN y de los síndromes de CCR hereditario polipósico, se encontraron mutados *BRCA1/2, CDKN2A, CDK4, CDH1, TP53, ATM, BARD1, BRIP1, CHEK2, NBN, PALB2, RAD51C* y *RAD51D,* que estaban previamente implicados en otras neoplasias extracolónicas (Yurgelun 2015).

**1.6- Futuros estudios**

En conjunto en ambos estudios de secuenciación del exoma se han detectado 13 variantes germinales en 12 nuevos genes candidatos de predisposición a CCR. Para poder establecerlos como genes hereditarios para el diagnóstico clínico se tendrían que realizar estudios funcionales que prueben que la función proteica está afectada debido a la variante y que la pérdida o alteración de la función proteica activa algún mecanismo desencadenante de la tumorigenesis. En la actualidad no se ha demostrado aún de manera definitiva la causalidad de la predisposición a CCR con las variantes candidatas. Esto se ha debido a que gran parte del trabajo realizado durante la tesis ha consistido en desarrollar las etapas anteriores del estudio, como la selección de pacientes, la secuenciación, el análisis de datos, la segregación y el estudio somático, habiendo sido estos estudios una nueva línea de investigación dentro del grupo. Estas variantes candidatas serán validadas próximamente mediante estudios funcionales.

A continuación se detallan algunos estudios adicionales que se podrían realizar en el futuro con el fin de relacionar las variantes candidatas detectadas con la tumorigenesis. La variante *CDKN1B* p.Q65H parece alterar la interacción con CDK2, lo cual podría probarse mediante una inmunoprecipitación de proteínas seguida por Western blot. Mediante mutagénesis dirigida se podría crear un mutante homocigoto y calcular la proliferación celular mediante un ensayo de viabilidad celular por luminiscencia. En el caso de la variante en *XRCC4*, Murray y col. irradiaron fibroblastos mutantes para *XRCC4* con luz ultravioleta y contaron el número de focos de γ-H2AX, indicadores de las brechas de doble hebra, para ver si estaba aumentado en comparación con los fibroblastos *wild-type* (Murray 2015). De esta forma comprobaron que los portadores de estas mutaciones tenían afectado el mecanismo de reparación del ADN por

recombinación no homóloga. Este mismo ensayo se podría realizar en los linfocitos de los portadores de la mutación *XRCC4* p.V166Efs*3. Por otro lado para validar el efecto de la variante en *EPHX1* se podría estudiar in vivo el tiempo de eliminación de los epóxidos dentro de las células con la mutación frente a células con el alelo normal de *EPHX1*, y también se podría cuantificar si hay más lesiones en el ADN de estas células provocadas por agentes citotóxicos mediante cultivo de linfocitos y extracción de mitosis utilizando el protocolo descrito previamente (Vodicka 2010). Respecto a la mutación en *NFKBIZ* se podría calcular si la interacción con NFκB está alterada mediante inmunoprecipitación de proteínas seguida por Western blot, o comprobar si hay mayor unión de NFκB a su promotor mediante Inmunoprecipitación de la cromatina. En el caso del gen *SMARCA4* se podría estudiar si la interacción con la proteína SSL18 está afectada por la variante mediante inmunoprecipitación de proteínas más Western blot. También se podría estudiar si la activación de c-Fos esta desregulada comparando la unión de c-Fos e su promotor en células *wild-type* para *SMARCA4* y células con la mutación mediante inmunoprecipitación de la cromatina. Estudios previos demostraron el papel supresor tumoral de este gen clonando el gen *SMARCA4* en líneas celulares que no lo expresaban, con el resultado de que el ciclo celular quedaba arrestado (Jelinic 2014). De la misma manera se podría clonar el gen mutado e introducirlo en estas células para ver si se produce también arresto del ciclo celular o no debido a la variante. El gen *SMARCA4* también se ha implicado en la reparación de las brechas de doble hebra (Qi 2014), así que podríamos tratar estas células mutadas con bleomicina (droga inductora de este tipo de lesiones) para ver si son más sensibles que las células con la proteína normal. Por último, la variante en *BARD1* se predice que elimina parte del dominio BRCT, que está implicado en interacción con diversos substratos. Esto también se podría comprobar mediante inmunoprecipitación y Western blot de los substratos de BARD1.

Respecto al segundo estudio, una buena forma de demostrar que los genes de la vía de reparación del ADN de la anemia de Fanconi están alterados por las variantes candidatas es comprobando que la capacidad de reparar brechas de doble hebra en el ADN se encuentra reducida en los portadores de mutaciones heterocigotas en estos genes. Para ello se debería extraer linfocitos de los pacientes y tratarlos con mitomicina C. También se tratarían estos linfocitos con afidicolina para estudiar si presentan más inestabilidad cromosómica después de someterlos a estrés replicativo, ya que un estudio previo demostró que mutantes *BLM, FANCA, FANCD2 y FANCC* presentaban más micronúcleos (indicadores de daño en el ADN post-estrés replicativo) que las células *wild-type* después del tratamiento con este fármaco (Naim y Rosselli 2009). Además sería interesante estudiar los perfiles tumorales para observar si

presentan recombinación mitótica que se refleja en una mayor pérdida de heterocigosidad, como se sugiere en el estudio de Luo y colaboradores (Luo 2000).

**2- Screening mutacional en *POLE* y *POLD1***

El tercer trabajo presentado en esta tesis conforma un estudio independiente de los dos anteriormente mencionados y consiste en el cribado mutacional del dominio exonucleasa de los nuevos genes hereditarios *POLE* y *POLD1,* en una cohorte española de 155 pacientes con múltiples pólipos y CCR de aparición temprana.

**2.1- Tecnología utilizada**

La técnica utilizada para el cribado mutacional en la cohorte de estudio fue la secuenciación Sanger de la zona codificante del dominio exonucleasa de los genes *POLE* y *POLD1*. Esta técnica no está enfocada en la detección de los cambios prevalentes *POLE* L424V y *POLD1* V478N, como otros estudios (Valle 2014, Elsayed 2015) sino que permite la detección de nuevas variantes situadas en la zona. Sin embargo, su mayor coste en comparación a la detección de cambios en un solo nucleótido limita el número de pacientes a analizar. Otro método para ampliar el espectro mutacional conocido es la secuenciación de nueva generación, como es el caso del estudio de Spier y col., donde se utilizó un panel de genes que cubría *POLE, POLD1* y otros 7 genes de las polimerasas en 266 pacientes (Spier 2014). Otro ejemplo de cribado por NGS sería el estudio de Chubb y col., donde se secuenció el exoma de 626 pacientes con CCR de aparición temprana (Chubb 2015). Sin embargo, el coste de aplicar secuenciación de nueva generación a una gran cohorte de pacientes para el cribado mutacional no es económicamente asequible para la mayoría de grupos de investigación y hay que optar por una estrategia coste-efectiva, como el genotipado o la secuenciación Sanger, siendo esta última la mejor para detectar nuevas mutaciones.

Por otro lado, para comparar los perfiles somáticos en un portador germinal de una mutación en *POLE* con pacientes de CCR sin alteración en este gen, se secuenció el exoma tumoral de estos individuos. El análisis de estos datos tuvo en cuenta tres particularidades. No se tuvo acceso a la secuenciación del exoma germinal de las muestras apareadas ni la cobertura de secuenciación del exoma es tan alta como en los paneles de genes. Además, la calidad de las distintas muestras de DNA tumoral secuenciado es variable debido a su degradación. Para evitar la variabilidad debido a la degradación del DNA en los distintos bloques de parafina sólo se tuvieron en cuenta aquellas variantes situadas en regiones dónde la cobertura fuera mayor a 10 en todos los individuos. Por otro lado, al no disponer de DNA germinal apareado con el

tumoral para poder filtrar las variantes germinales y así contar únicamente las variantes somáticas, una aproximación fue la utilización de una base de datos de 65 exomas germinales de individuos con CCR familiar, eliminándose los cambios presentes en esta base de datos. Para poder saber si se habían eliminado la mayor parte de variantes germinales después de estos filtrados se construyeron gráficos de densidad de probabilidad de la proporción del alelo alternativo en las variantes, que nos permite diferenciar las variantes germinales (con un porcentaje del alelo alternativo del 50%) y las somáticas (con un porcentaje generalmente menor al 50% de alelo alternativo, siendo alrededor del 40% cuando la pureza del tumor es de un 80%) Estos gráficos mostraron que la mayoría de variantes germinales habían sido filtradas correctamente.

**2.2- Características de los pacientes secuenciados**

Dentro de la cohorte de 155 casos a cribar, 83 pacientes correspondieron al grupo de poliposis múltiple (adenomatosa, serrada o mixta) y con agregación familiar para poliposis, CCR de aparición temprana, adenomas avanzados o cáncer de endometrio. Cincuenta y seis casos se englobaron en la categoría de CCR de aparición temprana (edad de aparición menor a 50 años), y 13 casos pertenecieron al grupo de CCR con MSI. Este último grupo engloba pacientes sin mutaciones germinales en los genes de Lynch ni metilación del promotor de *MLH1* pero con MSI e inmunohistoquímica alterada para alguna de las proteínas reparadoras. Otros estudios han usado criterios de selección de pacientes similares (Valle 2014) o menos restrictivos (Elsayed 2015, Spier 2015, Chubb 2015), con resultados similares en cuanto a la frecuencia de detección de las variantes *POLE* p.L424V y *POLD1* p.S478N. Teniendo en cuenta las características clínicas de los portadores de mutaciones conocidas en estos genes, su espectro fenotípico parece incluir poliposis múltiple y CCR de aparición temprana, así como historia familiar, lo cual reforzaría la elección de nuestros criterios de inclusión de pacientes. Concretamente, Bellido y col. crearon unos criterios clínicos para el diagnóstico genético de estos genes, basados en las características clínicas de los portadores. Estas pautas también señalan la presencia de pólipos y el cumplimiento de los criterios de Ámsterdam de agregación familiar para CCR para *POLE* y CCR y endometrio para *POLD1* (Bellido 2015).

**2.3- Mutaciones prevalentes en *POLE* y *POLD1***

Las variantes prevalentes en *POLE* (p.L424V) y *POLD1* (p.S478N) no fueron detectadas en nuestra cohorte de 155 pacientes. En estudios anteriores se ha observado que la frecuencia de la variante *POLE* p.L424V es inferior al 0,3% y la mutación p.S478N en *POLD1* es aún más rara con una frecuencia alélica menor al 0,1%. Nuestros resultados están en concordancia con la

rareza de estas mutaciones y evidencian que el tamaño de nuestra cohorte no es lo suficientemente grande como para detectarlas, pues hubiéramos necesitado unas 334 muestras para encontrar una variante con una frecuencia del 0,3%.

**2.4- Variantes detectadas en nuestro estudio**

Además de las mutaciones recurrentes, otras variantes han sido reportadas para *POLE* y *POLD1* en previos estudios. La mayoría de ellas están localizadas en el dominio exonucleasa e incluyen p.Trp347Cys, p.Asn363Lys, p.Asp386Val, p.Lys425Arg, p.Pro436Ser, p.Tyr458Phe en *POLE* (Hansen 2014, Rohlin 2014, Chubb 2015, Spier 2015, Aoude 2015, Hansen 2015), y p.Asp316His, p.Asp316Gly, p.Pro327Leu, p.Arg409Trp, p.Leu474Pro para *POLD1* (Palles 2013, Valle 2014, Bellido 2015).

Nuestro estudio no ha detectado ninguna de estas variantes. Sin embargo, hemos encontrado dos variantes no reportadas anteriormente, *POLE* c.1359+46del71 y *POLE* p.V474I. La primera variante es intrónica y está localizada entre los exones 13 y 14, detectándose en dos casos independientes de nuestra cohorte pero no en controles en las bases de datos de variación genética. Se realizaron estudios en el ARN y no se detectaron alteraciones en el splicing respecto a un individuo control, por lo cual esta variante intrónica es poco susceptible de causar predisposición a CCR.

**2.5- Variante p.V474I en *POLE***

Por otro lado, la variante *POLE* p.V474I se encontró en un individuo con CCR de aparición temprana y estaba localizada en el dominio N-terminal de la proteína a tan solo 3 residuos del dominio exonucleasa. El aminoácido afectado estaba altamente conservado en vertebrados así como en levaduras. El cambio no estaba presente en bases de datos de controles y todas las herramientas bioinformáticas de patogeneicidad predecían que resultaba deletéreo para la proteína; además tres de cuatro programas de predicción de la estabilidad de la estructura tridimensional indicaban que el cambio tenía un efecto desestabilizador.

Para demostrar que la variante afectaba la actividad correctora de la polimerasa, se realizaron estudios funcionales en la levadura *Schizosaccharomyces pombe* con el fenotipo *ade6*-485. Cuando comparamos tasa de reversión de las diferentes cepas en ausencia de adenina, la cepa con la variante equivalente a p.L424V creada como control positivo presentó 72,17 veces más revertientes que la cepa *wild-type*. Respecto la cepa creada con la mutación equivalente a p.V474I, se observaron 31 veces más mutantes que en el *wild-type*. Estos resultados sugerirían que la mutación p.V474I afecta la función reparadora de *POLE*, posiblemente de una manera

más atenuada que la mutación prevalente p. L424V. Esta sería la primera vez que se demuestra mediante estudios funcionales que una mutación germinal fuera del dominio exonucleasa puede afectar la función correctora de *POLE*. Por otro lado, Bellido y col. identificaron la variante en *POLD1* p.V295M, situada también fuera del dominio exonucleasa y con predicciones bioinformáticas deletéreas, aunque no se ha verificado funcionalmente (Bellido 2015).

Es interesante mencionar que la variante germinal *POLE* p.V474I, identificada en nuestro estudio, se ha encontrado de forma somática en una muestra de cáncer gástrico con MSI. Esta muestra tenía un número anormalmente alto de mutaciones somáticas (283 mutaciones por megabase, sumando 11.375 substituciones en total). Este tumor presentaba, además de la variante p.V474I, otras 3 mutaciones en *POLE* muy alejadas del dominio exonucleasa (p.S681R, p.R1111Q y p.Y1889C). Al secuenciar el exoma tumoral de nuestro paciente con la variante germinal POLE V474I, se observó un número mayor de substituciones en comparación con otros cuatro tumores secuenciados sin alteraciones germinales ni somáticas en el gen *POLE*. Sin embargo, este incremento fue sutil en comparación con la muestra de cáncer gástrico con la variante detectada somáticamente. Esto puede ser a que el número real de variantes somáticas se encuentre subestimado en nuestra muestra debido a que la secuenciación del exoma no tiene una cobertura tan alta como los paneles de genes y esto pueda provocar que al normalizar por cobertura mayor a 10 muchas mutaciones somáticas sean eliminadas del contaje. Por otro lado el número de variantes germinales no ha sido eliminado completamente del análisis y también puede estar enmascarando los resultados. Dejando de lado el aspecto técnico, el hecho de no encontrar un número tan alto de mutaciones somáticas en el exoma tumoral podría significar que esta variante pese a tener un efecto en la capacidad correctora de la polimerasa, provoca un fenotipo más atenuado que la variante p.L424V situada en la región de interacción con el ADN (Palles 2013). Esta hipótesis se ve apoyada por los ensayos funcionales realizados en levadura, donde la tasa de reversión de los mutantes ade685 a ade+ es significativamente mayor en las cepas *pol2* V475I que en las cepas *pol2* wild-type, pero menor que en las cepas pol2 L425V. Por otro lado, las otras mutaciones somáticas encontradas en el gen *POLE* en la muestra de cáncer gástrico podrían tener un efecto aditivo con la variante V474I, provocando una mayor gravedad del fenotipo hipermutador. En cuanto al espectro mutacional, mutaciones en el dominio exonucleasa de *POLE* y *POLD1* provocan un espectro mutacional caracterizado por un incremento del porcentaje de substituciones G>T/C>A (35,2%) y C>T/G>A (51,1%) (Palles 2013, Rayner 2016). En cambio, no se detecta una mayor frecuencia de variantes G>T/C>A (7,05 %) ni de C>T/G>A (42,73%) en nuestra muestra mutada

germinalmente en comparación con los otros cuatro tumores secuenciados. En la muestra de cáncer gástrico tampoco se observa un aumento importante de este tipo de substituciones, presentando el 10,29 % de substituciones G>T/C>A y el 42,38 % de cambios C>T/G>A. Esto puede significar que pese a que esta mutación fuera del dominio exonucleasa de *POLE* aumenta el número de mutaciones, estas no presentan el espectro mutacional característico de las mutaciones del dominio exonucleasa.

# CONCLUSIONES

- Se ha implementado una metodología automática para el análisis de los datos de secuenciación del exoma germinal en familias de cáncer colorrectal familiar, permitiendo que el análisis se realice de forma más rápida y objetiva.

- En el primer estudio de esta tesis doctoral, la secuenciación del exoma completo de 42 pacientes con cáncer colorrectal familiar pertenecientes a 29 familias ha permitido identificar nuevas variantes candidatas de predisposición destacando las localizadas en los genes *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4* y *BARD1*.

- Entre los genes candidatos identificados en este primer estudio, cabe resaltar que la mayoría de ellos se habían implicado previamente en la reparación del ADN o el control del ciclo celular, así como en predisposición germinal a otras neoplasias.

- En el segundo estudio de esta tesis doctoral, la secuenciación del exoma completo de 74 pacientes con cáncer colorrectal familiar pertenecientes a 40 familias ha permitido identificar nuevas variantes candidatas de predisposición destacando las localizadas en los genes *BRCA2*/*FANCD1*, *BRIP1*/*FANCJ*, *FANCC*, *FANCE* y *REV3L*/*POLZ*.

- Entre los genes candidatos identificados en este segundo estudio, cabe resaltar que forman parte de la vía de reparación del daño en el ADN de la anemia de Fanconi. Por tanto, este estudio concluye que esta vía puede jugar un papel importante en la predisposición germinal al cáncer colorrectal.

- Este estudio también respalda que genes implicados tradicionalmente en cáncer de mama y ovario como *BRCA2*, *BARD1* y *BRIP1* también se pueden encontrar mutados de forma germinal en cáncer colorrectal familiar.

- En el segundo estudio se evidencia que el fenómeno de la pleiotropía es una característica importante en cáncer hereditario ya que un número creciente de genes puede estar implicado en un amplio espectro de neoplasias.

- El tercer estudio permitió el cribado mutacional del dominio exonucleasa de los genes *POLE* y *POLD1* previamente relacionados con predisposición germinal a pólipos múltiples y cáncer colorrectal en una cohorte de 155 pacientes, identificando dos variantes genéticas potencialmente patogénicas.

- Este tercer estudio no detectó ninguna de las variantes genéticas reportadas previamente como prevalentes en *POLE* (p.Leu424Val) o *POLD1* (p.Ser478Asn) en nuestra cohorte de pacientes.

- La variante de cambio de aminoácido c.1420G>A (p.Val474Ile) constituye una nueva mutación en *POLE* identificada en una paciente con cáncer colorrectal familiar de tipo X. La patogenicidad de esta variante está demostrada por su ausencia en las bases de datos de variación genética humana, la conservación del aminoácido afectado, las predicciones funcionales y, finalmente, los ensayos funcionales en levadura.

- La mutación c.1420G>A (p.Val474Ile) en *POLE* sería la primera mutación germinal fuera del dominio exonucleasa que alteraría la función correctora de la polimerasa *POLE*.

- Se observó un aumento en el número de mutaciones generadas en los estudios en levadura tanto para la mutación p.L424V en *POLE* como para la nueva mutación p.V474I, aunque de una forma menos pronunciada en esta última, pudiendo indicar que mutaciones fuera del dominio exonucleasa pudieran ir asociadas a un fenotipo más atenuado.

# REFERENCIAS

# REFERENCIAS

Abdel-Rahman, W. M., M. Ollikainen, R. Kariola, H. J. Jarvinen, J. P. Mecklin, M. Nystrom-Lahti, S.Knuutila and P. Peltomaki. Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. Oncogene 2005;24(9):1542-51.

Abdus Sattar AK, Lin TC, Jones C, Konigsberg WH. Functional consequences and exonuclease kinetic parameters of point mutations in bacteriophage T4 DNA polymerase. Biochemistry 1996;35(51):16621-9.

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;Chapter 7:Unit7.20.

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(2): 227-32.

Al-Tassan NA, Whiffin N, Hosking FJ, Palles C, Farrington SM, Dobbins SE, Harris R, Gorman M, Tenesa A, Meyer BF, Wakil SM, Kinnersley B, Campbell H, Martin L, Smith CG, Idziaszczyk S, Barclay E, Maughan TS, Kaplan R, Kerr R, Kerr D, Buchanan DD, Ko Win A, Hopper J, Jenkins M, Lindor NM, Newcomb PA, Gallinger S, Conti D, Schumacher F, Casey G, Dunlop MG, Tomlinson IP, Cheadle JP, Houlston RS. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep 2015;5:10442.

Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, Cuatrecasas M, BujandaL, Giraldez MD, Takahashi M, Cabanne A, Barugel ME, Arnold M, Roca EL, Andreu M, Castellvi-Bel S, Llor X, Jover R, Castells A, Boland CR, Goel A. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. PLoS One 2012;7(9):e45357.

Aoude LG, Heitzer E, Johansson P, Gartside M, Wadt K, Pritchard AL, Palmer JM, Symmons J, Gerdes AM, Montgomery GW, Martin NG, Tomlinson I, Kearsey S, Hayward NK. POLE mutations in families predisposed to cutaneous melanoma. Fam Cancer 2015;14(4):621-8.

Baudhuin LM, Burgart LJ, Leontovich O, Thibodeau SN. Use of microsatellite instability and immunohistochemistry testing for the identification of individuals at risk for Lynch syndrome. Fam Cancer 2005;4(3):255-65. Review.

Beach R, Chan AO, Wu TT, White JA, Morris JS, Lunagomez S, Broaddus RR, Issa JP, Hamilton SR, Rashid A. BRAF mutations in aberrant crypt foci and hyperplastic polyposis. Am J Pathol 2005;166(4):1069-75.

Bee L, Nasca A, Zanolini A, Cendron F, d'Adamo P, Costa R, Lamperti C, Celotti L, Ghezzi D, Zeviani M. A nonsense mutation of human XRCC4 is associated with adult-onset progressive encephalocardiomyopathy. EMBO Mol Med 2015;7(7):918-29

Bellido F, Pineda M, Aiza G, Valdés-Mas R, Navarro M, Puente DA, Pons T, González S, Iglesias S, Darder E, Piñol V, Soto JL, Valencia A, Blanco I, Urioste M, Brunet J, Lázaro C, Capellá G,

Puente XS, Valle L. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med 2016;18(4):325-32.

Bingham SA, Day NE, Luben R, Ferrari P, Slimani N, Norat T, Clavel-Chapelon F,Kesse E, Nieters A, Boeing H, Tjønneland A, Overvad K, Martinez C, Dorronsoro M, Gonzalez CA, Key TJ, Trichopoulou A, Naska A, Vineis P, Tumino R, Krogh V, Bueno-de-Mesquita HB, Peeters PH, Berglund G, Hallmans G, Lund E, Skeie G, Kaaks R, Riboli E; European Prospective Investigation into Cancer and Nutrition. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. Lancet 2003;361(9368):1496-501.

Bland JM, Altman DG. Statistics notes. The odds ratio. BMJ 2000;320(7247):1468. Review.

Boparai KS, Dekker E, Van Eeden S, Polak MM, Bartelsman JF, Mathus-Vliegen EM,Keller JJ, van Noesel CJ. Hyperplastic polyps and sessile serrated adenomas as a phenotypic expression of MYH-associated polyposis. Gastroenterology 2008;135(6):2014-8

Boparai KS, Mathus-Vliegen EM, Koornstra JJ, Nagengast FM, van Leerdam M, van Noesel CJ, Houben M, Cats A, van Hest LP, Fockens P, Dekker E. Increased colorectal cancer risk during follow-up in patients with hyperplastic polyposis syndrome: a multicentre cohort study. Gut 2010;59(8):1094-100.

Bosserhoff AK, Grussendorf-Conen EI, Rübben A, Rudnik-Schöneborn S, Zerres K, Buettner R, Merkelbach-Bruse S. Multiple colon carcinomas in a patient with Cowden syndrome. Int J Mol Med 2006;18(4):643-7.

Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003;33:228-37. Review.

Botteri E, Iodice S, Raimondi S, Maisonneuve P, Lowenfels AB. Cigarette smoking and adenomatous polyps: a meta-analysis. Gastroenterology 2008;134(2):388-95.

Boveri T. Zur Frage der Entstehung Maligner Tumoren. Jena, Germany: Gustav Fisher; 1914. pp. 1–64.

Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 2013;14(10):681-91. Review.

Boyle T, Keegel T, Bull F, Heyworth J, Fritschi L. Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. J Natl Cancer Inst 2012;104(20):1548-61.

Brazil DP, Church RH, Surae S, Godson C, Martin F. BMP signalling: agony and antagony in the family. Trends Cell Biol 2015;25(5):249-64. Review.

Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet 2014;383(9927):1490-502.

Briggs S, Tomlinson I. Germline and somatic polymerase ε and δ mutations define a new class of hypermutated colorectal and endometrial cancers. J Pathol 2013;230(2):148-53.

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

Bülow S, Björk J, Christensen IJ, Fausa O, Järvinen H, Moesgaard F, Vasen HF; DAF Study Group. Duodenal adenomatosis in familial adenomatous polyposis. Gut 2004;53(3):381-6.

Bultman SJ, Herschkowitz JI, Godfrey V, Gebuhr TC, Yaniv M, Perou CM, Magnuson T. Characterization of mammary tumors from Brg1 heterozygous mice. Oncogene 2008;27(4):460-8

Burt RW. Hereditary and familial colon cancer. Gastroenterology 2010;138(6):2044-58. Review.

Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487(7407):330-7.

Cao X, Eu KW, Kumarasinghe MP, Li HH, Loi C, Cheah PY. Mapping of hereditary mixed polyposis syndrome (HMPS) to chromosome 10q23 by genome wide high-density single nucleotide polymorphism (SNP) scan and identification of BMPR1A loss of function. J Med Genet 2006;43(3):e13.

Casey G, Conti D, Haile R, Duggan D. Next generation sequencing and a new era of medicine. Gut 2013;62(6):920-32. Review.

Chan AO, Issa JP, Morris JS, Hamilton SR, Rashid A. Concordant CpG island methylation in hyperplastic polyposis. Am J Pathol 2002 ;160(2):529-36.

Chan AT, Giovannucci EL. Primary prevention of colorectal cancer. Gastroenterology 2010;138(6):2029-2043.e10.

Cheah PY, Wong YH, Chau YP, Loi C, Lim KH, Lim JF, Koh PK, Eu KW. Germline bone morphogenesis protein receptor 1A mutation causes colorectal tumorigenesis in hereditary mixed polyposis syndrome. Am J Gastroenterol 2009;104(12):3027-33.

Chen JR, Chiang JM, Changchien CR, Chen JS, Tang RP, Wang JY. Mismatch repair protein expression in Amsterdam II criteria-positive patients in Taiwan. Br J Surg 2008;95(1):102-10. Koh PK, Kalady M,

Cheng TH, Thompson D, Painter J, O'Mara T, Gorman M, Martin L, Palles C, Jones A, Buchanan DD, Ko Win A, Hopper J, Jenkins M, Lindor NM, Newcomb PA, Gallinger S, Conti D, Schumacher F, Casey G, Giles GG, Pharoah P, Peto J, Cox A, Swerdlow A, Couch F, Cunningham JM, Goode EL, Winham SJ, Lambrechts D, Fasching P, Burwinkel B, Brenner H, Brauch H, Chang-Claude J, Salvesen HB, Kristensen V, Darabi H, Li J, Liu T, Lindblom A, Hall P, de Polanco ME, Sans M, Carracedo A, Castellvi-Bel S, Rojas-Martinez A, Aguiar Jnr S, Teixeira MR, Dunning AM, Dennis J, Otton G, Proietto T, Holliday E, Attia J, Ashton K, Scott RJ, McEvoy M, Dowdy SC, Fridley BL, Werner HM, Trovik J, Njolstad TS, Tham E, Mints M, Runnebaum I, Hillemanns P, Dörk T, Amant F, Schrauwen S, Hein A, Beckmann MW, Ekici A, Czene K, Meindl A, Bolla MK, Michailidou K, Tyrer JP, Wang Q, Ahmed S, Healey CS, Shah M, Annibali D, Depreeuw J, Al-Tassan NA, Harris R, Meyer BF, Whiffin N, Hosking FJ, Kinnersley B, Farrington SM, Timofeeva

M, Tenesa A, Campbell H, Haile RW, Hodgson S, Carvajal-Carmona L, Cheadle JP, Easton D, Dunlop M, Houlston R, Spurdle A, Tomlinson I. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. Sci Rep 2015;5:17369.

Chillón M, Casals T, Giménez J, Ramos MD, Palacio A, Morral N, Estivill X, Nunes V. Analysis of the CFTR gene confirms the high genetic heterogeneity of the Spanish population: 43 mutations account for only 78% of CF chromosomes. Hum Genet 1994;93(4):447-51.

Chubb D, Broderick P, Frampton M, Kinnersley B, Sherborne A, Penegar S, Lloyd A, Ma YP, Dobbins SE, Houlston RS. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. J Clin Oncol 2015;33(5):426-32.

Cicek MS, Cunningham JM, Fridley BL, Serie DJ, Bamlet WR, Diergaarde B, Haile RW, Le Marchand L, Krontiris TG, Younghusband HB, Gallinger S, Newcomb PA, Hopper JL, Jenkins MA, Casey G, Schumacher F, Chen Z, DeRycke MS, Templeton AS, Winship I, Green RC, Green JS, Macrae FA, Parry S, Young GP, Young JP, Buchanan D, Thomas DC, Bishop DT, Lindor NM, Thibodeau SN, Potter JD, Goode EL; Colon CFR. Colorectal cancer linkage on chromosomes 4q21, 8q13, 12q24, and 15q22. PLoS One 2012;7(5):e38175.

Cleary SP, Zhang W, Di Nicola N, Aronson M, Aube J, Steinman A, Haddad R, Redston M, Gallinger S, Narod SA, Gryfe R. Heterozygosity for the BLM(Ash) mutation and cancer risk. Cancer Res 2003;63(8):1769-71.

Cleary SP, Cotterchio M, Jenkins MA, Kim H, Bristow R, Green R, Haile R,Hopper JL, LeMarchand L, Lindor N, Parfrey P, Potter J, Young husband B, Gallinger S. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology 2009;136(4):1251-60.

COGENT Study, Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S; Colorectal Cancer Association Study Consortium, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS; CoRGI Consortium, Kerr D, Cazier JB, Niittymäki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IP, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PD, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Völzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40(12):1426-35.

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(6):2039-46.

Cook NR, Lee IM, Zhang SM, Moorthy MV, Buring JE. Alternate-day, low-dose aspirin and cancer risk: long-term observational follow-up of a randomized trial. Ann Intern Med 2013;159(2):77-85.

Corradetti MN, Inoki K, Bardeesy N, DePinho RA, Guan KL. Regulation of the TSC pathway by LKB1: evidence of a molecular link between tuberous sclerosis complex and Peutz-Jeghers syndrome. Genes Dev 2004;18(13):1533-8.

Daley D, Lewis S, Platzer P, MacMillen M, Willis J, Elston RC, Markowitz SD, Wiesner GL. Identification of susceptibility genes for cancer in a genome-wide scan: results from the colon neoplasia sibling study. Am J Hum Genet 2008;82(3):723-36.

Davies RJ, Miller R, Coleman N. Colorectal cancer screening: prospects formolecular stool analysis. Nat Rev Cancer 2005;5(3):199-209. Review.

de Bruin C, Mericq V, Andrew SF, van Duyvenvoorde HA, Verkaik NS, Losekoot M, Porollo A, Garcia H, Kuang Y, Hanson D, Clayton P, van Gent DC, Wit JM, Hwa V, Dauber A. An XRCC4 splice mutation associated with severe short stature, gonadal failure, and early-onset metabolic syndrome. J Clin Endocrinol Metab 2015;100(5):E789-98.

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(3):544-7.

de Voer RM, Hahn MM, Mensenkamp AR, Hoischen A, Gilissen C, Henkes A, Spruijt L, van Zelst-Stams WA, Kets CM, Verwiel ET, Nagtegaal ID, Schackert HK, van Kessel AG, Hoogerbrugge N, Ligtenberg MJ, Kuiper RP. Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. Sci Rep 2015;5:14060.

de Voer RM, Hahn MM, Weren RD, Mensenkamp AR, Gilissen C, van Zelst-Stams WA, Spruijt L, Kets CM, Zhang J, Venselaar H, Vreede L, Schubert N, Tychon M, DerksR, Schackert HK, Geurts van Kessel A, Hoogerbrugge N, Ligtenberg MJ, Kuiper RP. Identification of Novel Candidate Genes for Early-Onset Colorectal Cancer Susceptibility. PLoS Genet 2016;12(2):e1005880.

Derbyshire V, Grindley ND, Joyce CM. The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. EMBO J 1991;10(1):17-24.

DeRycke MS, Gunawardena SR, Middha S, Asmann YW, Schaid DJ, McDonnell SK, Riska SM, Eckloff BW, Cunningham JM, Fridley BL, Serie DJ, Bamlet WR, Cicek MS, Jenkins MA, Duggan DJ, Buchanan D, Clendenning M, Haile RW, Woods MO, Gallinger SN, Casey G, Potter JD, Newcomb PA, Le Marchand L, Lindor NM, Thibodeau SN, Goode EL. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. Cancer Epidemiol Biomarkers Prev 2013;22(7):1239-51.

Di Gregorio C, Losi L, Fante R, Modica S, Ghidoni M, Pedroni M, Tamassia MG,Gafà L, Ponz de Leon M, Roncucci L. Histology of aberrant crypt foci in the humancolon. Histopathology 1997;30(4):328-34.

Dickinson BT, Kisiel J, Ahlquist DA, Grady WM. Molecular markers for colorectal cancer screening. Gut 2015;64(9):1485-94. Review.

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(3):362-6.

Domchek SM, Bradbury A, Garber JE, Offit K, Robson ME. Multiplex genetic testing for cancer susceptibility: out on the high wire without a net? J Clin Oncol 2013;31(10):1267-70.

Dove-Edwin I, de Jong AE, Adams J, Mesher D, Lipton L, Sasieni P, Vasen HF, Thomas HJ. Prospective results of surveillance colonoscopy in dominant familial colorectal cancer with and without Lynch syndrome. Gastroenterology 2006;130(7):1995-2000.

Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi LY, Domingo E, Smillie C, Henrion M, Frampton M, Martin L, Grimes G, Gorman M, Semple C, Ma YP, Barclay E, Prendergast J, Cazier JB, Olver B, Penegar S, Lubbe S, Chander I, Carvajal-Carmona LG, Ballereau S, Lloyd A, Vijayakrishnan J, Zgaga L, Rudan I, Theodoratou E; Colorectal Tumour Gene Identification (CORGI) Consortium, Starr JM, Deary I, Kirac I, Kovacević D, Aaltonen LA, Renkonen-Sinisalo L, Mecklin JP, Matsuda K, Nakamura Y, Okada Y, Gallinger S, Duggan DJ, Conti D, Newcomb P, Hopper J, Jenkins MA, Schumacher F, Casey G, Easton D, Shah M, Pharoah P, Lindblom A, Liu T; Swedish Low-Risk Colorectal Cancer Study Group, Smith CG, West H, Cheadle JP; COIN Collaborative Group, Midgley R, Kerr DJ, Campbell H, Tomlinson IP, Houlston RS. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 2012;44(7):770-6.

Elsayed FA, Kets CM, Ruano D, van den Akker B, Mensenkamp AR, Schrumpf M, Nielsen M, Wijnen JT, Tops CM, Ligtenberg MJ, Vasen HF, Hes FJ, Morreau H, van Wezel T. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2015;23(8):1080-4.

Estrada R, Spjut HJ. Hamartomatous polyps in Peutz-Jeghers syndrome. A light-,histochemical, and electron-microscopic study. Am J Surg Pathol 1983;7(8):747-54.

Farrington SM, Tenesa A, Barnetson R, Wiltshire A, Prendergast J, Porteous M, Campbell H, Dunlop MG. Germline susceptibility to colorectal cancer due tobase-excision repair gene defects. Am J Hum Genet 2005;77(1):112-9.

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(12):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(6):1374-403.

Fernandez-Rozadilla C, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, Baiget M, López-Fernández LA, Brea-Fernández A, Abulí A, Bujanda L, Clofent J, Gonzalez D, Xicola R, Andreu M, Bessa X, Jover R, Llor X; EPICOLON Consortium, Moreno V, Castells A, Carracedo Á, Castellvi-Bel S, Ruiz-Ponte C. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. BMC Genomics 2013a;14:55.

Fernandez-Rozadilla C, Brea-Fernández A, Bessa X, Alvarez-Urturi C, Abulí A, Clofent J, Payá A; EPICOLON Consortium, Jover R, Xicola R, Llor X, Andreu M, Castells A, Carracedo A, Castellví-Bel S, Ruiz-Ponte C. BMPR1A mutations in early-onset colorectal cancer with mismatch repair proficiency. Clin Genet 2013b;84:94-6

Fishel, R y col. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell1993; 75:1027-1038.

Francisco I, Albuquerque C, Lage P, Belo H, Vitoriano I, Filipe B, Claro I, Ferreira S, Rodrigues P, Chaves P, Leitão CN, Pereira AD. Familial colorectal cancer type X syndrome: two distinct molecular entities? Fam Cancer 2011;10(4):623-31

Fretland AJ, Omiecinski CJ. Epoxide hydrolases: biochemistry and molecular biology. Chem Biol Interact 2000;129:41–59.

Friis S, Riis AH, Erichsen R, Baron JA, Sørensen HT. Low-Dose Aspirin or Nonsteroidal Anti-inflammatory Drug Use and Colorectal Cancer Risk: A Population-Based, Case-Control Study. Ann Intern Med 2015;163(5):347-55.

Gala MK, Mizukami Y, Le LP, Moriichi K, Austin T, Yamamoto M, Lauwers GY, Bardeesy N, Chung DC. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. Gastroenterology 2014 ; 146(2):520-9.

Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. Nature 2009;460(7252):278-82.

Garre P, Briceño V, Xicola RM, Doyle BJ, de la Hoya M, Sanz J, Llovet P, Pescador P, Puente J, Díaz-Rubio E, Llor X, Caldés 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(7):1701-12.

Garre P, Martín L, Sanz J, Romero A, Tosar A, Bando I, Llovet P, Diaque P, García-Paredes B, Díaz-Rubio E, de la Hoya M, Caldés T. BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. Clin Genet 2015;87(6):582-7.

German J. Bloom's syndrome. Dermatol Clin. 1995;13(1):7-18. Review.

Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, BookerSV, Cruz-Correa M, Offerhaus JA. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology 2000;119(6):1447-53.

Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet 2012;20(5):490-7. Review.

Gingras D, Béliveau R. Colorectal cancer prevention through dietary and lifestyle modifications. Cancer Microenviron 2011;4(2):133-9.

Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Kearney J,Willett WC. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. men. J Natl Cancer Inst 1994;86(3):183-91.

Giovannucci E. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2001;10(7):725-31.

Gismondi V, Meta M, Bonelli L, Radice P, Sala P, Bertario L, Viel A,Fornasarig M, Arrigoni A, Gentile M, Ponz de Leon M, Anselmi L, Mareni C, Bruzzi P, Varesco L. Prevalence of the Y165C, G382D and 1395delGGA germline mutations of the MYH gene in Italian patients with adenomatous polyposis coli and colorectal adenomas. Int J Cancer 2004;109(5):680-4.

Goldberg Y, Halpern N, Hubert A, Adler SN, Cohen S, Plesser-Duvdevani M, Pappo O, Shaag A, Meiner V. Mutated MCM9 is associated with predisposition to hereditary mixed polyposis and colorectal cancer in addition to primary ovarian failure. Cancer Genet 2015;208(12):621-4.

Gorlin RJ, Cohen MM Jr, Condon LM, Burke BA. Bannayan-Riley-Ruvalcaba syndrome. Am J Med Genet 1992;44(3):307-14.

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(13):5409-18.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR. Patterns of somatic mutation in human cancer genomes. Nature 2007;446(7132):153-8.

Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M, y col. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66(3):589-600. Jasperson KW, Tuohy TM, Neklason DW,

Gruber SB, Ellis NA, Scott KK, Almog R, Kolachana P, Bonner JD, Kirchhoff T, Tomsho LP, Nafa K, Pierce H, Low M, Satagopan J, Rennert H, Huang H, Greenson JK, Groden J, Rapaport B, Shia J,

Johnson S, Gregersen PK, Harris CC, Boyd J, Rennert G, Offit K. BLM heterozygosity and the risk of colorectal cancer. Science 2002;297(5589):2013.

Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, Parmigiani 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(31):12921-5.

Guo C, Nakazawa Y, Woodbine L, Björkman A, Shimada M, Fawcett H, Jia N, Ohyama K, Li TS, Nagayama Y, Mitsutake N, Pan-Hammarström Q, Gennery AR, Lehmann AR, Jeggo PA, Ogi T. XRCC4 deficiency in human subjects causes a marked neurological phenotype but no overt immunodeficiency. J Allergy Clin Immunol 2015;136(4):1007-17.

Guo J, Yu L, Turro NJ, Ju J. An integrated system for DNA sequencing by synthesis using novel nucleotide analogues. Acc Chem Res 2010;43(4):551-63.

Gylfe AE, Katainen R, Kondelin J, Tanskanen T, Cajuso T, Hänninen U, Taipale J, Taipale M, Renkonen-Sinisalo L, Järvinen H, Mecklin JP, Kilpivaara O, Pitkänen E, Vahteristo P, Tuupanen S, Karhu A, Aaltonen LA. Eleven candidate susceptibility genes for common familial colorectal cancer. PLoS Genet 2013;9(10):e1003876.

Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. Orphanet J Rare Dis 2009;4:22. Review.

Hamilton ML, Van Remmen H, Drake JA, Yang H, Guo ZM, Kewitt K, Walter CA,Richardson A. Does oxidative damage to DNA increase with age? Proc Natl Acad Sci U S A 2001;98(18):10469-74.

Hansen MF, Johansen J, Bjørnevoll I, Sylvander AE, Steinsbekk KS, Sætrom P, Sandvik AK, Drabløs F, Sjursen W. A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine. Fam Cancer 2015;14(3):437-48.

Hawkins NJ, Tomlinson I, Meagher A, Ward RL. Microsatellite-stable diploid carcinoma: a biologically distinct and aggressive subset of sporadic colorectal cancer. Br J Cancer 2001;84(2):232-6.

Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. Trends Biochem Sci 2014;39(4):199-218

Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G, Warren W, Aminoff M, Höglund P, Järvinen H, Kristo P, Pelin K, Ridanpää 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(6663):184-7.

Hemminki K, Chen B. Familial risk for colorectal cancers are mainly due to heritable causes. Cancer Epidemiol Biomarkers Prev 2004;13(7):1253-6.

Hemminki K, Frank C, Försti A, Musak L, Kazimirova A, Barancokova M, Horska A, Vymetalkova V, Smerhovsky Z, Naccarati A, Soucek P, Vodickova L, Buchancova J, Smolkova B, Dusinska M, Vodicka P. Metabolic gene variants associated with chromosomal aberrations in healthy humans. Genes Chromosomes Cancer 2015;54(4):260-6.

Hofstad, B. Colonoscopy: principles and practice. Massachusetts: Blackwell Publishing.

Holcomb and Trask (2007) Genetic variation. A laboratory manual. Cold Spring Harbour Laboratory Press. Edited by Weiner M.P, Gabriel S.B. and Stephen J.C.

Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, Spain SL, Broderick P, Domingo E, Farrington S, Prendergast JG, Pittman AM, Theodoratou E, Smith CG, Olver B, Walther A, Barnetson RA, Churchman M, Jaeger EE, Penegar S, Barclay E, Martin L, Gorman M, Mager R, Johnstone E, Midgley R, Niittymäki I, Tuupanen S, Colley J, Idziaszczyk S; COGENT Consortium, Thomas HJ, Lucassen AM, Evans DG, Maher ER; CORGI Consortium; COIN Collaborative Group; COINB Collaborative Group, Maughan T, Dimas A, Dermitzakis E, Cazier JB, Aaltonen LA, Pharoah P, Kerr DJ, Carvajal-Carmona LG, Campbell H, Dunlop MG, Tomlinson IP. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet 2010;42(11):973-7. .

Houlston RS; members of COGENT. COGENT (COlorectal cancer GENeTics) revisited. Mutagenesis 2012;27(2):143-51.

Howe JR, Roth S, Ringold JC, Summers RW, Järvinen HJ, Sistonen P, Tomlinson IP, Houlston RS, Bevan S, Mitros FA, Stone EM, Aaltonen LA. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science 1998;280(5366):1086-8.

Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, Petersen GM, Velculescu VE, Traverso G, Vogelstein B. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet 2001;28(2):184-7.

Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M.The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer 2009;125(1):171-80.

Hwang YT, Liu BY, Coen DM, Hwang CB. Effects of mutations in the Exo III motif of the herpes simplex virus DNA polymerase gene on enzyme activities, viral replication, and replication fidelity. J Virol 1997;71(10):7791-8.

IJspeert JE, Vermeulen L, Meijer GA, Dekker E. Serrated neoplasia-role in colorectal carcinogenesis and clinical implications. Nat Rev Gastroenterol Hepatol 2015;12(7):401-9.

Irminger-Finger I, Jefford CE. Is there more to BARD1 than BRCA1? Nat Rev Cancer. 2006 May;6(5):382-91. Review. PubMed PMID: 16633366.

Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinimann K, Howarth K, East J, Taylor J, Thomas H, Tomlinson I. 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(6):699-703.

Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology 2010;138(6):2044-58.Review.

Jass JR, Williams CB, Bussey HJ, Morson BC. Juvenile polyposis--a precancerous condition. Histopathology 1988;13(6):619-30.

Jelinic P, Mueller JJ, Olvera N, Dao F, Scott SN, Shah R, Gao J, Schultz N,Gonen M, Soslow RA, Berger MF, Levine DA. Recurrent SMARCA4 mutations in smallcell carcinoma of the ovary. Nat Genet 2014;46(5):424-6.

Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, Kim DH, Ren Z, Cai Q, Long J, Shi J, Wen W, Yang G, Delahanty RJ; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO); Colon Cancer Family Registry (CCFR), Ji BT, Pan ZZ, Matsuda F, Gao YT, Oh JH, Ahn YO, Park EJ, Li HL, Park JW, Jo J, Jeong JY, Hosono S, Casey G, Peters U, Shu XO, Zeng YX, Zheng W. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet 2013 ;45(2):191-6.

Johansson E, Dixon N. Replicative DNA polymerases. Cold Spring Harb Perspect Biol 2013;5(6). Review.

Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol 2001;96(10):2992-3003. Review.

Johnsen JM, Nickerson DA, Reiner AP. Massively parallel sequencing: the new frontier of hematologic genomics. Blood 2013;122(19):3268-75. Review.

Johnson RE, Klassen R, Prakash L, Prakash S. A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. Mol Cell 2015;59(2):163-75.

Johnson VL, Scott MI, Holt SV, Hussein D, Taylor SS. Bub1 is required for kinetochore localization of BubR1, Cenp-E, Cenp-F and Mad2, and chromosome congression. J Cell Sci 2004;117(Pt 8):1577-89.

Jones S, Emmerson P, Maynard J, Best JM, Jordan S, Williams GT, Sampson JR, Cheadle JP. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C-->T:A mutations. Hum Mol Genet 2002;11(23):2961-7.

Jones S, Lambert S, Williams GT, Best JM, Sampson JR, Cheadle JP. Increased frequency of the k-ras G12C mutation in MYH polyposis colorectal adenomas. Br J Cancer 2004;90(8):1591-3.

Kane DP, Shcherbakova PV. A common cancer-associated DNA polymerase ε mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. Cancer Res 2014;74(7):1895-901.

Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines.Cancer Res 1997 Mar 1;57(5):808-11.

Kastrinos F, Syngal S. Inherited colorectal cancer syndromes. Cancer J. 2011;17(6):405-15

Kazak L, Reyes A, Holt IJ. Minimizing the damage: repair pathways keep mitochondrial DNA intact. Nat Rev Mol Cell Biol 2012;13(10):659-71. Review.

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; ColoRectal tumour Gene Identification (CoRGI) Study Consortium. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet 2006;15(19):2903-10.

Kessel R, Vlachos A, Lipton JM. Ribosomopathy association with colorectal cancer. Gastroenterology 2015;148(1):258.

Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins SE, Casey G, Hopper J, Sieber O, Lipton L, Kerr DJ, Dunlop MG, Tomlinson IP, Houlston RS; Colon Cancer Family Registry. The TERT variant rs2736100 is associated with colorectal cancer risk. Br J Cancer 2012;107(6):10018.

Kinnersley B, Chubb D, Dobbins SE, Frampton M, Buch S, Timofeeva MN, Castellví-Bel S, Farrington SM, Forsti A, Hampe J, Hemminki K, Hofstra RM, Northwood E, Palles C, Pinheiro M, Ruiz-Ponte C, Schafmayer C, Teixeira MR, Westers H, van Wezel T, Timothy Bishop D, Tomlinson I, Dunlop MG, Houlston RS. Correspondence: SEMA4A variation and risk of colorectal cancer. Nat Commun 2016 Mar 10;7:10611. doi: 10.1038/ncomms10611.

Klarskov L, Holck S, Bernstein I, Nilbert M. Hereditary colorectal cancer diagnostics: morphological features of familial colorectal cancer type X versus Lynch syndrome. J Clin Pathol 2012;65(4):352-6.

Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68(4):820-3.

Koh PK, Kalady M, Skacel M, Fay S, McGannon E, Shenal J, Arroyo L, Toderick K, Church J. Familial colorectal cancer type X: polyp burden and cancer risk stratification via a family history score. ANZ J Surg 2011;81(7-8):537-42.

Korkaya H, Wicha MS. Cancer stem cells: nature versus nurture. Nat Cell Biol 2010;12(5):419-21.

Koushik A, Hunter DJ, Spiegelman D, Beeson WL, van den Brandt PA, Buring JE,Calle EE, Cho E, Fraser GE, Freudenheim JL, Fuchs CS, Giovannucci EL, Goldbohm RA, Harnack L, Jacobs DR Jr, Kato I, Krogh V, Larsson SC, Leitzmann MF, Marshall JR, McCullough ML, Miller AB, Pietinen P, Rohan TE, Schatzkin A, Sieri S,Virtanen MJ, Wolk A, Zeleniuch-Jacquotte A, Zhang SM, Smith-Warner SA. Fruits, vegetables, and colon cancer risk in a pooled analysis of 14 cohort studies. J Natl Cancer Inst 2007;99(19):1471-83.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073-81.

Küry S, Garrec C, Airaud F, Breheret F, Guibert V, Frenard C, Jiao S, Bonneau D, Berthet P, Bossard C, Ingster O, Cauchin E, Bezieau S. Evaluation of the colorectal cancer risk conferred by rare UNC5C alleles. World J Gastroenterol 2014;20(1):204-13.

Laiho P, Hienonen T, Karhu A, Lipton L, Aalto Y, Thomas HJ,Birkenkamp-Demtroder K, Hodgson S, Salovaara R, Mecklin JP, Järvinen H, Knuutila S, Halford S, Ørntoft TF, Tomlinson I, Launonen V, Houlston R, Aaltonen LA. Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess ofallelic imbalance in chromosome 20q in familial cases. Oncogene 2003;22(14):2206-14.

Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. Int J Cancer 2006;119(11):2657-64.

Lemire M, Qu C, Loo LW, Zaidi SH, Wang H, Berndt SI, Bézieau S, Brenner H, Campbell PT, Chan AT, Chang-Claude J, Du M, Edlund CK, Gallinger S, Haile RW, Harrison TA, Hoffmeister M, Hopper JL, Hou L, Hsu L, Jacobs EJ, Jenkins MA, Jeon J, Küry S, Li L, Lindor NM, Newcomb PA, Potter JD, Rennert G, Rudolph A, Schoen RE, Schumacher FR, Seminara D, Severi G, Slattery ML, White E, Woods MO, Cotterchio M, Le Marchand L, Casey G, Gruber SB, Peters U, Hudson TJ. A genome-wide association study for colorectal cancer identifies a risk locus in 14q23.1. Hum Genet 2015;134(11-12):1249-62.

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(2):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(1):112-7

Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, Fisk CJ, Li N, Smolyar A,Hill DE, Barabási AL, Vidal M, Zoghbi HY. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 2006;125(4):801-14.

Lim W, Olschwang S, Keller JJ, Westerman AM, Menko FH, Boardman LA, Scott RJ, Trimbath J, Giardiello FM, Gruber SB, Gille JJ, Offerhaus GJ, de Rooij FW, WilsonJH, Spigelman AD, Phillips RK, Houlston RS. Relative frequency and morphology of cancers in STK11 mutation carriers. Gastroenterology 2004;126(7):1788-94.

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(16):1979-85.

Lipton L, Sieber OM, Thomas HJ, Hodgson SV, Tomlinson IP, Woodford-Richens K. Germline mutations in the TGF-beta and Wnt signalling pathways are a rare cause of the "multiple" adenoma phenotype. J Med Genet 2003;40(4):e35.

Liu F, Yuan D, Wei Y, Wang W, Yan L, Wen T, Xu M, Yang J, Li B. Systematic review and meta-analysis of the relationship between EPHX1 polymorphisms and colorectal cancer risk. PLoS One 2012;7(8):e43821. Review.

Llor X, Pons E, Xicola RM, Castells A, Alenda C, Piñol V, Andreu M,Castellví-Bel S, Payá A, Jover R, Bessa X, Girós A, Roca A, Gassull MA; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. Clin Cancer Res 2005;11(20):7304-10.

Lowy AM, Kordich JJ, Gismondi V, Varesco L, Blough RI, Groden J. Numerous colonic adenomas in an individual with Bloom's syndrome. Gastroenterology 2001;121(2):435-9.

Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, Huang KL, Wyczalkowski MA, Jayasinghe R, Banerjee T, Ning J, Tripathi P, Zhang Q, Niu B, Ye K, Schmidt HK, Fulton RS, McMichael JF, Batra P, Kandoth C, Bharadwaj M, Koboldt DC, Miller CA, Kanchi KL, Eldred JM, Larson DE, Welch JS, You M, Ozenberger BA, Govindan R, Walter MJ, Ellis MJ, Mardis ER, Graubert TA, Dipersio JF, Ley TJ, Wilson RK, Goodfellow PJ, Raphael BJ, Chen F, Johnson KJ, Parvin JD, Ding L. Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun 2015;6:10086.

Lubbe SJ, Pittman AM, Matijssen C, Twiss P, Olver B, Lloyd A, Qureshi M, Brown N, Nye E, Stamp G, Blagg J, Houlston RS. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. Hum Mutat 2011;32(1):E1928-38

Luo G, Santoro IM, McDaniel LD, Nishijima I, Mills M, Youssoufian H, Vogel H, Schultz RA, Bradley A. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. Nat Genet 2000;26(4):424-9.

Lynch HT, Lynch JF, Lynch PM, Attard T. Hereditary colorectal cancersyndromes: molecular genetics, genetic counseling, diagnosis and management. Fam Cancer 2008;7(1):27-39. Review.

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):1-18. Review.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C. Guidelines for investigating causality of sequence variants in human disease. Nature 2014;508(7497):469-76.

Mahdavinia M, Bishehsari F, Ansari R, Norouzbeigi N, Khaleghinejad A, HormazdiM, Rakhshani N, Malekzadeh R. Family history of colorectal cancer in Iran. BMC Cancer 2005;5:112.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthyMI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, KruglyakL, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature 2009;461(7265):747-53. Review.

Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995; 268(5215):1336-8

Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med 2009;361(25):2449-60. Review

Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science 2015;349(6255):1483-9. Review.

McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010;141(2):210-7.

Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11(1):31-46.

Michels KB, Edward Giovannucci, Joshipura KJ, Rosner BA, Stampfer MJ, Fuchs CS, Colditz GA, Speizer FE, Willett WC. Prospective study of fruit and vegetable consumption and incidence of colon and rectal cancers. J Natl Cancer Inst 2000;92(21):1740-52.

Michels KB, Giovannucci E, Chan AT, Singhania R, Fuchs CS, Willett WC. Fruit and vegetable consumption and colorectal adenomas in the Nurses' Health Study. Cancer Res 2006;66(7):3942-53.

Middeldorp A, Jagmohan-Changur SC, van der Klift HM, van Puijenbroek M, Houwing-Duistermaat JJ, Webb E, Houlston R, Tops C, Vasen HF, Devilee P, Morreau H, van Wezel T, Wijnen J. Comprehensive genetic analysis of seven large families with mismatch repair proficient colorectal cancer. Genes Chromosomes Cancer 2010;49(6):539-48.

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(4):837-46.

Minoo P, Baker K, Goswami R, Chong G, Foulkes WD, Ruszkiewicz AR, Barker M,Buchanan D, Young J, Jass JR. Extensive DNA methylation in normal colorectalmucosa in hyperplastic polyposis. Gut 2006;55(10):1467-74.

Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, IgariT, Koike M, Chiba M, Mori T. Germline mutation of MSH6 as the cause of hereditarynonpolyposis colorectal cancer. Nat Genet 1997;17(3):271-2.

Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. Cancer Epidemiol Biomarkers Prev 2007;16(12):2533-47.

Mueller-Koch Y, Vogelsang H, Kopp R, Lohse P, Keller G, Aust D, Muders M, Gross M, Daum J, Schiemann 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(12):1733-40.

Mur P, Elena SC, Aussó S, Aiza G, Rafael VM, Pineda M, Navarro M, Brunet J, Urioste M, Lázaro C, Moreno V, Capellá G, Puente XS, Valle L. Scarce evidence of the causal role of germline mutations in UNC5C in hereditary colorectal cancer and polyposis. Sci Rep 2016;6:20697. doi: 10.1038/srep20697.

Murphy K, Darmawan H, Schultz A, Fidalgo da Silva E, Reha-Krantz LJ. A method to select for mutator DNA polymerase deltas in Saccharomyces cerevisiae. Genome 2006;49(4):403-10.

Murray JE, van der Burg M, IJspeert H, Carroll P, Wu Q, Ochi T, Leitch A, Miller ES, Kysela B, Jawad A, Bottani A, Brancati F, Cappa M, Cormier-Daire V, Deshpande C, Faqeih EA, Graham GE, Ranza E, Blundell TL, Jackson AP, Stewart GS, Bicknell LS. Mutations in the NHEJ component XRCC4 cause primordial dwarfism. Am J Hum Genet 2015;96(3):412-24.

Naim V, Rosselli F. The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. Nat Cell Biol 2009;11(6):761-8.

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(21):8993-7.

Nelen MR, van Staveren WC, Peeters EA, Hassel MB, Gorlin RJ, Hamm H, LindboeCF, Fryns JP, Sijmons RH, Woods DG, Mariman EC, Padberg GW, Kremer H. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. Hum Mol Genet 1997;6(8):1383-7.

Neufeld G, Sabag AD, Rabinovicz N, Kessler O. Semaphorins in angiogenesis and tumor progression. Cold Spring Harb Perspect Med 2012;2(1):a006718. Review.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010;42(1):30-5.

Ngeow J, Heald B, Rybicki LA, Orloff MS, Chen JL, Liu X, Yerian L, Willis J, Lehtonen HJ, Lehtonen R, Mester JL, Moline J, Burke CA, Church J, Aaltonen LA, Eng C. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. Gastroenterology 2013;144(7):1402-9

Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA,Haseltine WA, Fleischmann RD, Fraser CM, y col. Mutations of two PMS homologuesin hereditary nonpolyposis colon cancer. Nature 1994;371(6492):75-80.

Nieminen TT, Abdel-Rahman WM, Ristimäki A, Lappalainen M, Lahermo P, MecklinJP, Järvinen HJ, Peltomäki P. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. Gastroenterology 2011; 141(1):e23-6.

Nieminen TT, O'Donohue MF, Wu Y, Lohi H, Scherer SW, Paterson AD, Ellonen P, Abdel-Rahman WM, Valo S, Mecklin JP, Järvinen HJ, Gleizes PE, Peltomäki P. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. Gastroenterology 2014;147(3):595-598.e5.

O'Brien MJ, Winawer SJ, Zauber AG, Gottlieb LS, Sternberg SS, Diaz B,Dickersin GR, Ewing S, Geller S, Kasimian D, y col. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 1990;98(2):371-9.

O'Hagan RC, Chang S, Maser RS, Mohan R, Artandi SE, Chin L, DePinho RA. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. Cancer Cell 2002; 2(2):149-55.

O'Riordan JM, O'Donoghue D, Green A, Keegan D, Hawkes LA, Payne SJ, Sheahan K,Winter DC. Hereditary mixed polyposis syndrome due to a BMPR1A mutation.Colorectal Dis 2010;12(6):570-3.

Orlando G, Law PJ, Palin K, Tuupanen S, Gylfe A, Hänninen U, Cajuso T, Tanskanen T, Kondelin J, Kaasinen E, Sarin AP, Kaprio J, Eriksson JG, Rissanen H, Knekt P, Pukkala E, Jousilahti P, Salomaa V, Ripatti S, Palotie A, Järvinen H, Renkonen-Sinisalo L, Lepistö A, Böhm J, Meklin JP, Al-Tassan NA, Palles C, Martin L, Barclay E, Tenesa A, Farrington S, Timofeeva MN, Meyer BF, Wakil SM, Campbell H, Smith CG, Idziaszczyk S, Maughan TS, Kaplan R, Kerr R, Kerr D, Buchanan DD, Win AK, Hopper J, Jenkins M, Lindor NM, Newcomb PA, Gallinger S, Conti D, Schumacher F, Casey G, Taipale J, Cheadle JP, Dunlop MG, Tomlinson IP, Aaltonen LA, Houlston RS. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. Hum Mol Genet. 2016 Mar 22. pii: ddw087. [Epub ahead of print] PubMed PMID: 27005424.

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; CORGI Consortium; WGS500 Consortium, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsey SE, Thomas 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(2):136-44.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA,Haseltine WA, Fleischmann RD, Fraser CM, Adams MD, y col. Mutation of a mutLhomolog in hereditary colon cancer. Science 1994 Mar 18;263(5153):1625-9.

Papaemmanuil E, Carvajal-Carmona L, Sellick GS, Kemp Z, Webb E, Spain S, Sullivan K, Barclay E, Lubbe S, Jaeger E, Vijayakrishnan J, Broderick P, Gorman M, Martin L, Lucassen A, Bishop DT,

Evans DG, Maher ER, Steinke V, Rahner N, Schackert HK, Goecke TO, Holinski-Feder E, Propping P, Van Wezel T, Wijnen J, Cazier JB, Thomas H, Houlston RS, Tomlinson I; CORGI Consortium. Deciphering the genetics of hereditary non-syndromic colorectal cancer. Eur J Hum Genet 2008;16(12):1477-86.

Pellegata NS, Quintanilla-Martinez L, Siggelkow H, Samson E, Bink K, Höfler H, Fend F, Graw J, Atkinson MJ. Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. Proc Natl Acad Sci U S A 2006;103(42):15558-63.

Perea J, Rodríguez Y, Rueda D, Marín JC, Díaz-Tasende J, Álvaro E, Alegre C,Osorio I, Colina F, Lomas M, Hidalgo M, Benítez J, Urioste M. Early-onset colorectal cancer is an easy and effective tool to identify retrospectively Lynch syndrome. Ann Surg Oncol 2011;18(12):3285-91.

Peters U, Bien S, Zubair N. Genetic architecture of colorectal cancer. Gut 2015;64(10):1623-36.

Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, Berndt SI, Bézieau S, Brenner H, Butterbach K, Caan BJ, Campbell PT, Carlson CS, Casey G, Chan AT, Chang-Claude J, Chanock SJ, Chen LS, Coetzee GA, Coetzee SG, Conti DV, Curtis KR, Duggan D, Edwards T, Fuchs CS, Gallinger S, Giovannucci EL, Gogarten SM, Gruber SB, Haile RW, Harrison TA, Hayes RB, Henderson BE, Hoffmeister M, Hopper JL, Hudson TJ, Hunter DJ, Jackson RD, Jee SH, Jenkins MA, Jia WH, Kolonel LN, Kooperberg C, Küry S, Lacroix AZ, Laurie CC, Laurie CA, Le Marchand L, Lemire M, Levine D, Lindor NM, Liu Y, Ma J, Makar KW, Matsuo K, Newcomb PA, Potter JD, Prentice RL, Qu C, Rohan T, Rosse SA, Schoen RE, Seminara D, Shrubsole M, Shu XO, Slattery ML, Taverna D, Thibodeau SN, Ulrich CM, White E, Xiang Y, Zanke BW, Zeng YX, Zhang B, Zheng W, Hsu L; Colon Cancer Family Registry and the Genetics and Epidemiology of Colorectal Cancer Consortium. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 2013;144(4):799-807.e24.

Picelli S, Vandrovcova J, Jones S, Djureinovic T, Skoglund J, Zhou XL,Velculescu VE, Vogelstein B, Lindblom A. Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. BMC Cancer 2008;8:87.

Pineda M, González S, Lázaro C, Blanco I, Capellá G. Detection of genetic alterations in hereditary colorectal cancer screening. Mutat Res 2010; 693(1-2):19-31.

Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology 2010;138(6):2059-72. Review.

Piñol V, Castells A, Andreu M, Castellví-Bel S, Alenda C, Llor X, Xicola RM,Rodríguez-Moranta F, Payá A, Jover R, Bessa X; Gastrointestinal Oncology Group ofthe Spanish Gastroenterological Association. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 2005;293(16):1986-94.

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(4):357-66.

Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype.MScience 1997; 275(5302):967-9

Ratajska M, Antoszewska E, Piskorz A, Brozek I, Borg Å, Kusmierek H, Biernat W, Limon J. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. Breast Cancer Res Treat 2012;131(1):89-97

Rayner E, van Gool IC, Palles C, Kearsey SE, Bosse T, Tomlinson I, Church DN. A panoply of errors: polymerase proofreading domain mutations in cancer. Nat Rev Cancer 2016;16(2):71-81.

Real LM, Ruiz A, Gayán J, González-Pérez A, Sáez ME, Ramírez-Lorca R, Morón FJ, Velasco J, Marginet-Flinch R, Musulén E, Carrasco JM, Moreno-Rey C, Vázquez E, Chaves-Conde M, Moreno-Nogueira JA, Hidalgo-Pascual M, Ferrero-Herrero E, Castellví-Bel S, Castells A, Fernandez-Rozadilla C, Ruiz-Ponte C, Carracedo A, González B, Alonso S, Perucho M. A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis. PLoS One 2014;9(6):e101178.

Rehman M, Tamagnone L. Semaphorins in cancer: biological mechanisms and therapeutic approaches. Semin Cell Dev Biol 2013;24(3):179-89. Review.

Rivera B, Castellsagué E, Bah I, van Kempen LC, Foulkes WD. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 2015;373(20):1985-6

Rohlin A, Zagoras T, Nilsson S, Lundstam U, Wahlström J, Hultén L, Martinsson T, Karlsson GB, Nordling M. A mutation in POLE predisposing to a multi-tumour phenotype. Int J Oncol 2014;45(1):77-81.

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

Rosin N, Elcioglu NH, Beleggia F, Isgüven P, Altmüller J, Thiele H, Steindl K, Joset P, Rauch A, Nürnberg P, Wollnik B, Yigit G. Mutations in XRCC4 cause primary microcephaly, short stature and increased genomic instability. Hum Mol Genet 2015;24(13):3708-17. doi: 10.1093/hmg/ddv115. Epub 2015. PubMed PMID: 25839420.

Rudkjøbing LA, Eiberg H, Mikkelsen HB, Binderup ML, Bisgaard ML. The analysis of a large Danish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 11q24. Fam Cancer 2015;14(3):393-400.

Russo AA, Jeffrey PD, Patten AK, Massagué J, Pavletich NP. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. Nature 1996;382:325–331.

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(9377):39-41.

Sánchez-Tomé E, Rivera B, Perea J, Pita G, Rueda D, Mercadillo F, Canal A, Gonzalez-Neira A, Benitez J, Urioste M. Genome-wide linkage analysis and tumoral characterization reveal heterogeneity in familial colorectal cancer type X. J Gastroenterol 2015;50(6):657-66.

Sayed MG, Ahmed AF, Anderson ME, y col. Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis. Ann Surg Oncol 2002;9:901– 906.

Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, Feuk L. Challenges and standards in integrating surveys of structural variation. Nat Genet 2007;39(7 Suppl):S7-15. Review.

Schiemann U, Müller-Koch Y, Gross M, Daum J, Lohse P, Baretton G, Muders M, Mussack T, Kopp R, Holinski-Feder E. Extended microsatellite analysis inmicrosatellite stable, MSH2 and MLH1 mutation-negative HNPCC patients: genetic reclassification and correlation with clinical features. Digestion2004;69(3):166-76.

Schneppenheim R, Frühwald MC, Gesk S, y col. Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. Am J Hum Genet 2010;86:279–284

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

Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, Hsu L, Huang SC, Fischer CP, Harju JF, Idos GE, Lejbkowicz F, Manion FJ, McDonnell K, McNeil CE, Melas M, Rennert HS, Shi W, Thomas DC, Van Den Berg DJ, Hutter CM, Aragaki AK, Butterbach K, Caan BJ, Carlson CS, Chanock SJ, Curtis KR, Fuchs CS, Gala M, Giovannucc EL, Gogarten SM, Hayes RB, Henderson B, Hunter DJ, Jackson RD, Kolonel LN, Kooperberg C, Küry S, LaCroix A, Laurie CC, Laurie CA, Lemire M, Levine D, Ma J, Makar KW, Qu C, Taverna D, Ulrich CM, Wu K, Kono S, West DW, Berndt SI, Bezieau S, Brenner H, Campbell PT, Chan AT, Chang-Claude J, Coetzee GA, Conti DV, Duggan D, Figueiredo JC, Fortini BK, Gallinger SJ, Gauderman WJ, Giles G, Green R, Haile R, Harrison TA, Hoffmeister M, Hopper JL, Hudson TJ, Jacobs E, Iwasaki M, Jee SH, Jenkins M, Jia WH, Joshi A, Li L, Lindor NM, Matsuo K, Moreno V, Mukherjee B, Newcomb PA, Potter JD, Raskin L, Rennert G, Rosse S, Severi G, Schoen RE, Seminara D, Shu XO, Slattery ML, Tsugane S, White E, Xiang YB, Zanke BW, Zheng W, Le Marchand L, Casey G, Gruber SB, Peters U. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nat Commun 2015;6:7138.

Seguí N, Pineda M, Navarro M, Lázaro C, Brunet J, Infante M, Durán M, Soto JL, Blanco I, Capellá G, Valle L. GALNT12 is not a major contributor of familial colorectal cancer type X. Hum Mutat 2014;35(1):50-2.

Seguí N, Mina LB, Lázaro C, Sanz-Pamplona R, Pons T, Navarro M, Bellido F, López-Doriga A, Valdés-Mas R, Pineda M, Guinó E, Vidal A, Soto JL, Caldés T, Durán M, Urioste M, Rueda D, Brunet J, Balbín M, Blay P, Iglesias S, Garré P, Lastra E, Sánchez-Heras AB, Valencia A, Moreno V, Pujana MÁ, Villanueva A, Blanco I, Capellá G, Surrallés J, Puente XS, Valle L. Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. Gastroenterology 2015;149(3):563-6.

Shaheen R, Faqeih E, Ansari S, Abdel-Salam G, Al-Hassnan ZN, Al-Shidi T, Alomar R, Sogaty S, Alkuraya FS. Genomic analysis of primordial dwarfism reveals novel disease genes. Genome Res 2014 ;24(2):291-9.

Shevelev IV, Hübscher U. The 3' 5' exonucleases. Nat Rev Mol Cell Biol 2002;3(5):364-76. Review.

Shiovitz S, Copeland WK, Passarelli MN, Burnett-Hartman AN, Grady WM, Potter JD, Gallinger S, Buchanan DD, Rosty C, Win AK, Jenkins M, Thibodeau SN, Haile R, Baron JA, Marchand LL, Newcomb PA, Lindor NM; Colon Cancer Family Registry. Characterisation of familial colorectal cancer Type X, Lynch syndrome, and non-familial colorectal cancer. Br J Cancer 2014;111(3):598-602.

Short E, Thomas LE, Hurley J, Jose S, Sampson JR. Inherited predisposition to colorectal cancer: towards a more complete picture. J Med Genet 2015;52(12):791-6. Review.

Sieber OM, Lipton L, Crabtree M, Heinimann K, Fidalgo P, Phillips RK, BisgaardML, Orntoft TF, Aaltonen LA, Hodgson SV, Thomas HJ, Tomlinson IP. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N Engl J Med 2003;348(9):791-9.

Sill H, Schulz E, Steinke-Lange V, Boland CR. Correspondence: Reply to 'SEMA4A variation and risk of colorectal cancer'. Nat Commun 2016 Mar 10;7:10695. doi: 10.1038/ncomms10695.

Silla IO, Rueda D, Rodríguez Y, García JL, de la Cruz Vigo F, Perea J. Early-onset colorectal cancer: a separate subset of colorectal cancer. World J Gastroenterol 2014;20(46):17288-96.Review.

Skoglund J, Djureinovic T, Zhou XL, Vandrovcova J, Renkonen E, Iselius L, Bisgaard ML, Peltomäki 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(2):e7.

Smith CG, Naven M, Harris R, Colley J, West H, Li N, Liu Y, Adams R, Maughan TS, Nichols L, Kaplan R, Wagner MJ, McLeod HL, Cheadle JP. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. Hum Mutat 2013;34(7):1026-34.

Snover D, Ahnen D, Burt R, Odze RD. Serrated Polyps of the Colon and Rectum and Serrated Polyposis, WHO Classification of Tumours of the Digestive System, IARC, Lyon, France, 4th edition, 2010.

Snover, D.C., Ahnen, D.J., Burt, R.W. and Odze, R.D. in WHO Classification of Tumors of the Digestive System Vol.3 Ch. 8 (ed Bosman, F. T., Carneiro, F., Hruban, R.H., Theise N.D) 160-165 (IARC Press, 2010)

Soengas MS, Esteban JA, Lázaro JM, Bernad A, Blasco MA, Salas M, Blanco L. Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. EMBO J 1992;11(11):4227-37

Spier I, Holzapfel S, Altmüller J, Zhao B, Horpaopan S, Vogt S, Chen S, Morak M, Raeder S, Kayser K, Stienen D, Adam R, Nürnberg P, Plotz G, Holinski-Feder E, Lifton RP, Thiele H, Hoffmann P, Steinke V, Aretz S. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int J Cancer 2015;137(2):320-31.

Spier I, Kerick M, Drichel D, Horpaopan S, Altmüller J, Laner A, Holzapfel S, Peters S, Adam R, Zhao B, Becker T, Lifton RP, Holinski-Feder E, Perner S, Thiele H, Nöthen MM, Hoffmann P, Timmermann B, Schweiger MR, Aretz S. Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis 2016;15(2):281-8.

Stigliano V, Sanchez-Mete L, Martayan A, Anti M. Early-onset colorectal cancer: a sporadic or inherited disease? World J Gastroenterol 2014;20(35):12420-30.Review.

Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, Seruca R,Iacopetta B, Hamelin R. Evaluation of tumor microsatellite instability using fivequasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology 2002;123(6):1804-11.

Suzuki A, de la Pompa JL, Stambolic V, Elia AJ, Sasaki T, del Barco Barrantes I, Ho A, Wakeham A, Itie A, Khoo W, Fukumoto M, Mak TW. High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. Curr Biol 1998;8(21):1169-78.

Tanskanen T, Gylfe AE, Katainen R, Taipale M, Renkonen-Sinisalo L, Järvinen H, Mecklin JP, Böhm J, Kilpivaara O, Pitkänen E, Palin K, Vahteristo P, Tuupanen S, Aaltonen LA. Systematic search for rare variants in Finnish early-onset colorectal cancer patients. Cancer Genet 2015;208(1-2):35-40.

Taupin D, Lam W, Rangiah D, McCallum L, Whittle B, hang Y, Andrews D, Field M, Goodnow C, Cook M. A deleterious RNF43 germline mutation in a severely affected serrated polyposis kindred. Human Genome Variation 2015;2,15013; doi:10.1038/hgv.2015.13

Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Völzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IP, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CM, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Genome-wide association scan identifies a

colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 2008;40(5):631-7.

Thutkawkorapin J, Picelli S, Kontham V, Liu T, Nilsson D, Lindblom A. Exome sequencing in one family with gastric- and rectal cancer. BMC Genet 2016;17(1):41.

Timofeeva MN, Kinnersley B, Farrington SM, Whiffin N, Palles C, Svinti V, Lloyd A, Gorman M, Ooi LY, Hosking F, Barclay E, Zgaga L, Dobbins S, Martin L, Theodoratou E, Broderick P, Tenesa A, Smillie C, Grimes G, Hayward C, Campbell A, Porteous D, Deary IJ, Harris SE, Northwood EL, Barrett JH, Smith G, Wolf R, Forman D, Morreau H, Ruano D, Tops C, Wijnen J, Schrumpf M, Boot A, Vasen HF, Hes FJ, van Wezel T, Franke A, Lieb W, Schafmayer C, Hampe J, Buch S, Propping P, Hemminki K, Försti A, Westers H, Hofstra R, Pinheiro M, Pinto C, Teixeira M, Ruiz-Ponte C, Fernandez-Rozadilla C, Carracedo A, Castells A, Castellví-Bel S, Campbell H, Bishop DT, Tomlinson IP, Dunlop MG, Houlston RS. Recurrent Coding Sequence Variation Explains Only A Small Fraction of the Genetic Architecture of Colorectal Cancer. Sci Rep 2015;5:16286.

Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K; CORGI Consortium, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB, Houlston R. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39(8):984-8.

Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijayakrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, Gray R, Maher ER, Lucassen A, Kerr D, Evans DG; CORGI Consortium, Schafmayer C, Buch S, Völzke H, Hampe J, Schreiber S, John U, Koessler T, Pharoah P, van Wezel T, Morreau H, Wijnen JT, Hopper JL, Southey MC, Giles GG, Severi G, Castellví-Bel S, Ruiz-Ponte C, Carracedo A, Castells A; EPICOLON Consortium, Försti A, Hemminki K, Vodicka P, Naccarati A, Lipton L, Ho JW, Cheng KK, Sham PC, Luk J, Agúndez JA, Ladero JM, de la Hoya M, Caldés T, Niittymäki I, Tuupanen S, Karhu A, Aaltonen L, Cazier JB, Campbell H, Dunlop MG, Houlston RS. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 2008;40(5):623-30.

Tomlinson I. The Mendelian colorectal cancer syndromes. Ann Clin Biochem 2015;52(Pt 6):690-2.

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65(2):87-108.

Totzke G, Essmann F, Pohlmann S, Lindenblatt C, Jänicke RU, Schulze- Osthoff K. A novel member of the IkappaB family, human IkappaB-zeta, inhibits transactivation of p65 and its DNA binding. J Biol Chem 2006;281:12645– 12654

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(15):8681-6.

Ulahannan D, Kovac MB, Mulholland PJ, Cazier JB, Tomlinson I. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. Br J Cancer 2013109(4):827-35.

Umar A, Risinger JI, Hawk ET, Barrett JC. Testing guidelines for hereditary non-polyposis colorectal cancer. Nat Rev Cancer 2004 ;4(2):153-8. Review.

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(7):781-6.

Valle L, Hernández-Illán E, Bellido F, Aiza G, Castillejo A, Castillejo MI, Navarro M, Seguí N, Vargas G, Guarinos C, Juarez M, Sanjuán X, Iglesias S, Alenda C, Egoavil C, Segura Á, Juan MJ, Rodriguez-Soler M, Brunet J, González S, Jover R, Lázaro C, Capellá G, Pineda M, Soto JL, Blanco I. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Hum Mol Genet 2014;23(13):3506-12.

van der Luijt RB, Meera Khan P, Vasen HF, Breukel C, Tops CM, Scott RJ, Fodde R. Germline mutations in the 3' part of APC exon 15 do not result in truncatedproteins and are associated with attenuated adenomatous polyposis coli. Hum Genet 1996;98(6):727-34.

van Lier MG, Wagner A, Mathus-Vliegen EM, Kuipers EJ, Steyerberg EW, van Leerdam ME. High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. Am J Gastroenterol 2010;105(6):1258-64; authorreply 1265. Review.

Venkatachalam R, Ligtenberg MJ, Hoogerbrugge N, Schackert HK, Görgens 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(6):2221-4.

Venkatachalam R, Verwiel ET, Kamping EJ, Hoenselaar E, Görgens H, Schackert HK, van Krieken JH, Ligtenberg MJ, Hoogerbrugge N, van Kessel AG, Kuiper RP. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. Int J Cancer 2011;129(7):1635-42.

Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007;23(6):657-63.

Villacis RA, Miranda PM, Gomy I, Santos EM, Carraro DM, Achatz MI, Rossi BM, Rogatto SR. Contribution of rare germline copy number variations and common susceptibility loci in Lynch syndrome patients negative for mutations in the mismatch repair genes. Int J Cancer 2016;138(8):1928-35.

Vodicka P, Polivkova Z, Sytarova S, Demova H, Kucerova M, Vodickova L, Polakova V, Naccarati A, Smerhovsky Z, Ambrus M, Cerna M, Hemminki K. Chromosomal damage in peripheral blood lymphocytes of newly diagnosed cancer patients and healthy controls. Carcinogenesis 2010;31(7):1238-41.

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(9):525-32.

Vogelstein B, Kinzler KW. The genetic basis of human cancer. Part 3 584-585 (McGraw-Hill Medical, 2002)

Waite KA, Eng C. Protean PTEN: form and function. Am J Hum Genet 2002;70(4):829-44. Review.

Waite KA, Eng C. From developmental disorder to heritable cancer: it's all in the BMP/TGF-beta family. Nat Rev Genet 2003;4(10):763-73. Review.

Wang H, Burnett T, Kono S, Haiman CA, Iwasaki M, Wilkens LR, Loo LW, Van Den Berg D, Kolonel LN, Henderson BE, Keku TO, Sandler RS, Signorello LB, Blot WJ, Newcomb PA, Pande M, Amos CI, West DW, Bézieau S, Berndt SI, Zanke BW, Hsu L; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Lindor NM, Haile RW, Hopper JL, Jenkins MA, Gallinger S, Casey G; Colon Cancer Family Registry (CCFR), Stenzel SL, Schumacher FR, Peters U, Gruber SB; Colorectal Transdisciplinary Study (CORECT), Tsugane S, Stram DO, Le Marchand L. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. Nat Commun 2014;5:4613.

Weren RD, Venkatachalam R, Cazier JB, Farin HF, Kets CM, de Voer RM, Vreede L, Verwiel ET, van Asseldonk M, Kamping EJ, Kiemeney LA, Neveling K, Aben KK, Carvajal-Carmona L, Nagtegaal ID, Schackert HK, Clevers H, van de Wetering M, Tomlinson IP, Ligtenberg MJ, Hoogerbrugge N, Geurts van Kessel A, Kuiper RP. Germline deletions in the tumour suppressor gene FOCAD are associated with polyposis and colorectal cancer development. J Pathol 2015a;236(2):155-64.

Weren RD, Ligtenberg MJ, Kets CM, de Voer RM, Verwiel ET, Spruijt L, van Zelst-Stams WA, Jongmans MC, Gilissen C, Hehir-Kwa JY, Hoischen A, Shendure J, Boyle EA, Kamping EJ, Nagtegaal ID, Tops BB, Nagengast FM, Geurts van Kessel A, van Krieken JH, Kuiper RP, Hoogerbrugge N. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet 2015b;47(6):668-71.

Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, Lloyd A, Kinnersley B, Gorman M, Tenesa A, Broderick P, Wang Y, Barclay E, Hayward C, Martin L, Buchanan DD, Win AK, Hopper J, Jenkins M, Lindor NM, Newcomb PA, Gallinger S, Conti D, Schumacher F, Casey G, Liu T; Swedish Low-Risk Colorectal Cancer Study Group, Campbell H, Lindblom A, Houlston RS, Tomlinson IP, Dunlop MG. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Hum Mol Genet 2014;23(17):4729-37.

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(2):327-34.

Wiedemann HR, Burgio GR, Aldenhoff P, Kunze J, Kaufmann HJ, Schirg E. Theproteus syndrome. Partial gigantism of the hands and/or feet, nevi,hemihypertrophy, subcutaneous tumors, macrocephaly or other skull anomalies and possible accelerated growth and visceral affections. Eur J Pediatr 1983;140(1):5-12.

Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillenM, 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(22):12961-5.

Williams, R.S., Glover, J.N., 2003. Structural consequences of a cancer-causing BRCA1-BRCT missense mutation. J. Biol. Chem 278, 2630–2635.

Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer 2011;11:481–492

Win AK, Walters RJ, Buchanan DD, Jenkins MA, Sweet K, Frankel WL, de la Chapelle A, McKeone DM, Walsh MD, Clendenning M, Pearson SA, Pavluk E, Nagler B, Hopper JL, Gattas MR, Goldblatt J, George J, Suthers GK, Phillips KD, Woodall S, Arnold J, Tucker K, Field M, Greening S, Gallinger S, Aronson M, Perrier R, Woods MO, Green JS, Walker N, Rosty C, Parry S, Young JP. Cancer risks for relatives of patients with serrated polyposis. Am J Gastroenterol 2012;107(5):770-8.

Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, Waye JD, Schapiro M, Bond JH, Panish JF, y col. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993;329(27):1977-81.

Witkowski L, Carrot-Zhang J, Albrecht S, y col. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. Nat Genet 2014;46:438–443.

Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer 2009;100(4):611-6.

Woods MO, Hyde AJ, Curtis FK, Stuckless S, Green JS, Pollett AF, Robb JD,Green RC, Croitoru ME, Careen A, Chaulk JA, Jegathesan J, McLaughlin JR,Gallinger SS, Younghusband HB, Bapat BV, Parfrey PS. High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. Clin Cancer Res 2005;11(19 Pt 1):6853-61.

Xicola RM, Rodríguez-Moranta F, Payá A, Jover R, Bessa X; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 2005;293(16):1986-94.

Yamaguchi S, Ogata H, Katsumata D, Nakajima M, Fujii T, Tsutsumi S, Asao T,Sasaki K, Kuwano H, Kato H. MUTYH-associated colorectal cancer and adenomatous polyposis. Surg Today 2014;44(4):593-600.Epub 2013. Review.

Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kaushik P, Roa BB, Wenstrup RJ, Hartman AR, Syngal S. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. Gastroenterology 2015;149(3):604-13.e20.

Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, Blanche H, Sahbatou M, Tubacher E, Bonaiti-Pellié C, Buecher B, Riboli E, Kury S, Chanock SJ, Potter J, Thomas G, Gallinger S, Hudson TJ, Dunlop MG. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 2007;39(8):989-94

Zbuk KM, Eng C. Hamartomatous polyposis syndromes. Nat Clin Pract Gastroenterol Hepatol 2007;4(9):492-502. Review.

Zeng C, Matsuda K, Jia WH, Chang J, Kweon SS, Xiang YB, Shin A, Jee SH, Kim DH, Zhang B, Cai Q, Guo X, Long J, Wang N, Courtney R, Pan ZZ, Wu C, Takahashi A, Shin MH, Matsuo K, Matsuda F, Gao YT, Oh JH, Kim S, Jung KJ, Ahn YO, Ren Z, Li HL, Wu J, Shi J, Wen W, Yang G, Li B, Ji BT; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Brenner H, Schoen RE, Küry S; Colorectal Transdisciplinary (CORECT) Study, Gruber SB, Schumacher FR, Stenzel SL; Colon Cancer Family Registry (CCFR), Casey G, Hopper JL, Jenkins MA, Kim HR, Jeong JY, Park JW, Tajima K, Cho SH, Kubo M, Shu XO, Lin D, Zeng YX, Zheng W. Identification of Susceptibility Loci and Genes for Colorectal Cancer Risk. Gastroenterology. 2016 Mar 7. pii: S0016-5085(16)00299-7. doi: 10.1053/j.gastro.2016.02.076. [Epub ahead of print] PubMed PMID: 26965516.

Zhang B, Jia WH, Matsuo K, y col. Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. Int J Cancer 2014a;135:948–55.

Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, Shin A, Jee SH, Kim DH, Cai Q, Long J, Shi J, Wen W, Yang G, Zhang Y, Li C, Li B, Guo Y, Ren Z, Ji BT, Pan ZZ, Takahashi A, Shin MH, Matsuda F, Gao YT, Oh JH, Kim S, Ahn YO; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Chan AT, Chang-Claude J, Slattery ML; Colorectal Transdisciplinary (CORECT) Study, Gruber SB, Schumacher FR, Stenzel SL; Colon Cancer Family Registry (CCFR), Casey G, Kim HR, Jeong JY, Park JW, Li HL, Hosono S, Cho SH, Kubo M, Shu XO, Zeng YX, Zheng W. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet 2014b;46(6):533-42.

Zhang G, Lischetti T, Hayward DG, Nilsson J. Distinct domains in Bub1 localize RZZ and BubR1 to kinetochores to regulate the checkpoint. Nat Commun 2015;6:7162.

Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013;14

# ANEXO

**Genetic susceptibility variants associated with colorectal cancer prognosis**

Abulí A, Lozano JJ, Rodríguez-Soler M, Jover R, Bessa X, Muñoz J, Esteban-Jurado C, Fernández-Rozadilla C, Carracedo A, Ruiz-Ponte C, Cubiella J, Balaguer F, Bujanda L, Reñé JM, Clofent J, Morillas JD, Nicolás-Pérez D, Xicola RM, Llor X, Piqué JM, Andreu M, Castells A, Castellví-Bel S; Gastrointestinal Oncology Group of Spanish Gastroenterological Association. Genetic susceptibility variants associated with colorectal cancer prognosis. Carcinogenesis 2013;34:2286-91.

**New genes emerging for colorectal cancer predisposition**

Esteban-Jurado C, Garre P, Vila M, Lozano JJ, Pristoupilova A, Beltrán S, Abulí A, Muñoz J, Balaguer F, Ocaña T, Castells A, Piqué JM, Carracedo A, Ruiz-Ponte C, Bessa X, Andreu M, Bujanda L, Caldés T, Castellví-Bel S. New genes emerging for colorectal cancer predisposition. World J Gastroenterol 2014;20:1961-71. Review.

**The MLH1 c.1852_1853delinsGC (p.K618A) variant in colorectal cancer: genetic association study in 18,723 individuals**

Abulí A, Bujanda L, Muñoz J, Buch S, Schafmayer C, Valeria Maiorana M, Veneroni S, van Wezel T, Liu T, Westers H, Esteban-Jurado C, Ocaña T, Piqué JM, Andreu M, Jover R, Carracedo A, Xicola RM, Llor X, Castells A; EPICOLON Consortium, Dunlop M, Hofstra R, Lindblom A, Wijnen J, Peterlongo P, Hampe J, Ruiz-Ponte C, Castellví-Bel S. The MLH1 c.1852_1853delinsGC (p.K618A) variant in colorectal cancer: genetic association study in 18,723 individuals. PLoS One 2014;9:e95022.

**Genetic Variants associated with colorectal adenoma susceptibility**

Abulí A, Castells A, Bujanda L, Lozano JJ, Bessa X, Hernández C, Álvarez-Urturi C, Pellisé M, Esteban-Jurado C, Hijona E, Burón A, Macià F, Grau J, Guayta R, Castellví-Bel S, Andreu M; PROCOLON research group. Genetic Variants Associated with Colorectal Adenoma Susceptibility. PLoS One. 2016 Apr 14;11(4):e0153084

**Carcinogenesis** vol.34 no.10 pp.2286–2291, 2013
doi:10.1093/carcin/bgt179
*Advance Access publication May 27, 2013*

# Genetic susceptibility variants associated with colorectal cancer prognosis

**Anna Abulí[1,2], Juan José Lozano[3], María Rodríguez-Soler[4], Rodrigo Jover[4], Xavier Bessa[2], Jenifer Muñoz[1], Clara Esteban-Jurado[1], Ceres Fernández-Rozadilla[5], Angel Carracedo[5], Clara Ruiz-Ponte[5], Joaquín Cubiella[6], Francesc Balaguer[1], Luis Bujanda[7], Josep M.Reñé[8], Juan Clofent[9], Juan Diego Morillas[10], David Nicolás-Pérez[11], Rosa M.Xicola[12], Xavier Llor[12], Josep M.Piqué[1], Montserrat Andreu[2], Antoni Castells[1] and Sergi Castellví-Bel[1,*]; for the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association[13]**

[1]Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Catalonia 08036, Spain, [2]Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, Barcelona, Catalonia 08018, Spain, [3]Bioinformatics Platform, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Barcelona, Catalonia 08036, Spain, [4]Department of Gastroenterology, Hospital General d'Alacant, Alicante 03010, Spain, [5]Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, Santiago de Compostela, University of Santiago de Compostela, Santiago de Compostela 15704, Galicia, Spain, [6]Department of Gastroenterology, Complexo Hospitalario de Ourense, Ourense 32005, Spain, [7]Department of Gastroenterology, Hospital Donostia, Networked Biomedical Research Centre for Hepatic and Digestive Diseases (CIBEREHD), Basque Country University, San Sebastián 20014, Spain, [8]Department of Gastroenterology, Hospital Arnau de Vilanova, Lleida 25198, Spain, [9]Department of Gastroenterology, Hospital Meixoeiro, Vigo 36200, Spain, [10]Department of Gastroenterology, Hospital Clínico de Madrid, Madrid 28040, Spain, [11]Department of Gastroenterology, Hospital Universitario de Canarias, Santa Cruz de Tenerife, Canarias 38320, Spain, [12]Section of Digestive Diseases and Nutrition, University of Illinois at Chicago, Chicago, IL 60607, USA and [13]All authors are listed in a Supplementary Note, available at *Carcinogenesis* Online

*To whom correspondence should be addressed. Tel: +34 93 2275400 ext. 4183; Fax: +34 93 3129405;
Email: sbel@clinic.ub.es

**Colorectal cancer (CRC) is the second leading cause of cancer-related death among men and women in Western countries. Once a tumour develops, a differentiated prognosis could be determined by lifestyle habits or inherited and somatic genetic factors. Finding such prognostic factors will be helpful in order to identify cases with a shorter survival or at a higher risk of recurrence that may benefit from more intensive treatment and follow-up surveillance. Sixteen CRC genetic susceptibility variants were directly genotyped in a cohort of 1235 CRC patients recruited by the EPICOLON Spanish consortium. Univariate Cox and multivariate regression analyses were performed taking as primary outcomes overall survival (OS), disease-free survival and recurrence-free interval. Genetic variants rs9929218 at 16q22.1 and rs10795668 at 10p14 may have an effect on OS. The G allele of rs9929218 was linked with a better OS [GG genotype, genotypic model: hazard ratio (HR) = 0.65, 95% confidence interval (CI) 0.45–0.93, *P* = 0.0179; GG/GA genotypes, dominant model: HR = 0.66, 95% CI 0.47–0.94, *P* = 0.0202]. Likewise, the G allele of rs10795668 was associated with better clinical outcome (GG genotype, genotypic model: HR = 0.73, 95% CI 0.53–1.01, *P* = 0.0570; GA genotype, genotypic model: HR = 0.66, 95% CI 0.47–0.92, *P* = 0.0137; GG/GA genotypes, dominant model: HR = 0.68, 95% CI 0.50–0.94, *P* = 0.0194). In conclusion, CRC susceptibility variants rs9929218 and rs10795668 may exert some influence in modulating patient's survival and they deserve to be further tested in additional CRC cohorts in order to confirm their potential as prognosis or predictive biomarkers.**

**Abbreviations:** CI, confidence interval; CRC, colorectal cancer; DFS, disease-free survival; HR, hazard ratio; MSI, microsatellite instability; OS, overall survival; RFI, recurrence-free interval; SNPs, single nucleotide polymorphisms; TNM, tumour, node, metastasis.

## Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related death among men and women in Western countries (1). Despite recent improvements in CRC survival, it still remains a major public health problem with over 1 million cases worldwide and a disease-specific mortality of approximately 33% in the developed world. Moreover, a considerable portion of successfully treated CRC patients develop a recurrence or metastasis within 5 years of diagnosis (2). It remains as an important aim for the management of CRC to understand why some patients relapse or respond to chemotherapy treatment, whereas others do not.

The main CRC prognostic marker used in clinical practice at present is tumour stage at diagnosis [tumour, node, metastasis (TNM)], which represents depth of tumour invasion, number of affected lymph nodes and distant metastasis (3). Environmental factors such as smoking and diet, as well as some inherited genetic variants, are undoubtedly major risk factors for CRC, and likewise, several factors are known to influence CRC prognosis. Some studies reported that lifestyle factors, such as diet, physical activity and body mass index, influence tumour systemic inflammatory response, which affects cancer survival (4,5). Also, CRC genetic heterogeneity is known to influence patient's survival after diagnosis and the most studied markers of CRC prognosis have been somatic changes with a potential effect on cancer outcome. For instance, it is known that chromosomal instability is associated with a worse CRC prognosis, and microsatellite instability (MSI) and wild-type BRAF with a better prognosis (6,7). Additionally, inherited polymorphisms may have also the potential to affect CRC outcome. In fact, there is some evidence of familial concordance for survival in CRC, suggesting that inherited genetic variation can contribute to CRC prognosis (8). Several studies have reported associations with inherited genetic factors and CRC outcome, including single nucleotide polymorphisms (SNPs) in the mismatch repair genes (9), microRNA-related genes (10) and toll-like receptor genes (11).

Recently, genome-wide association studies have successfully identified 20 susceptibility SNPs in 18 risk loci for CRC at 1q41, 3q26.2, 5p15.33, 6p21, 8q23.3, 8q24.21, 10p14, 11q13.4, 11q23.11, 12q13.3, 14q22.2, 15q13.3, 16q22.1, 18q21.1, 19q13, 20p12.3, 20q13.33 and Xp22.2 (12–16). Besides affecting CRC predisposition, it could be plausible that these variants may also have a role in CRC prognosis. This hypothesis has been explored by previous studies to some extent, including several of these genetic variants in their analyses, but reaching contradictory results (17–19). Tenesa *et al.* (17) analysed 10 CRC susceptibility variants (rs16892766 at 8q23.3, rs6983267 at 8q24.21, rs10795668 at 10p14, rs3802842 at 11q23.11, rs4444235 at 14q22.2, rs4779584 at 15q13.3, rs9929218 at 16q22.1, rs4939827 at 18q21.1, rs10411210 at 19q13 and rs961253 at 20p12.3) in 2838 CRC cases and reported no association between them and all-cause or CRC-specific mortality after diagnosis. A subsequent report by Xing *et al.* (18) described the analysis of six CRC susceptibility loci (rs16892766, rs6983267, rs10795668, rs3802842, rs4779584 and rs4939827) and their association with recurrence and death in a reduced number of cases ($n$ = 380), suggesting that rs10795668 at 10p14 might be used as a survival biomarker to identify patients with a high risk of recurrence after chemotherapy. Very recently, Phipps *et al.* (19) examined the association between 16 CRC susceptibility variants (those analysed by Tenesa *et al.* (17) and rs6691170 and rs6687758 at

1q41, rs10936599 at 3q26.2, rs7136702 and rs11169552 at 12q13.3, rs4925386 at 20q13.33, whereas rs4813802 at 20p12.3 and rs1957636 at 14q22.2 were not included) and survival in 2611 cases, concluding that rs4779584 variant at 15q13.3 could affect progression of CRC. It should be noted that CRC susceptibility variants with positive results in Xing *et al*. (18) and Phipps *et al*. (19) were included in the study by Tenesa *et al*. (17), so results are not concordant between these studies. Hence, the prognosis significance of these CRC susceptibility variants remains controversial. Therefore, the aim of this study was to evaluate the prognostic or predictive value of 16 CRC genetic susceptibility SNPs in our population-based cohort of 1235 CRC patients in order to identify cases with worse clinical outcome that may benefit from more intensive treatment and follow-up surveillance.

## Materials and methods

*Patients*

We included 1235 CRC cases recruited by the EPICOLON consortium, a Spanish epidemiologic, prospective, multicentre and population-based study (20,21). Patients were selected from this cohort for whom follow-up data were available with a median follow-up of 36 months (range 0–72 months). Patients with inflammatory bowel disease, biallelic *MUTYH* mutations, Lynch syndrome (germline mutation carriers) or familial adenomatous polyposis were excluded. Colon cancer patients had received adjuvant or palliative 5-fluorouracil or FOLFOX-based chemotherapy, whereas rectal cancer patients had undergone 5-fluorouracil or FOLFOX-based chemotherapy or a combination of these regimens with radiation therapy.

Clinical characteristics considered in this study included gender, age, CRC location, TNM stage, chemotherapy treatment and tumour features such as MSI, histology type (adenocarcinoma and mucinous/signet-ring differentiation), degree of differentiation and presence of tumour-infiltrating lymphocytes.

This study was approved by the institutional ethics committee of each participant hospital in EPICOLON, and written informed consent was obtained from all patients.

*Genotyping*

Sixteen SNPs previously associated with CRC risk were genotyped in this study including rs6691170 (1q41), rs10936599 (3q26.2), rs16892766 (8q23.3), rs6983267 (8q24.21), rs10795668 (10p14), rs3802842 (11q23.1), rs11169552 (12q13.13), rs4444235 and rs1957636 (14q22.2), rs4779584 (15q13.3), rs9929218 (16q22.1), rs4939827 (18q22.1), rs10411210 (19q13), rs961253 and rs4813802 (20p12.3), and rs4925386 (20q13.33; Supplementary Table 1, available at *Carcinogenesis* Online). Genotyping was performed using germline DNA from all cases and TaqMan allelic discrimination (Applied Biosystems) or single-base primer extension chemistry matrix-assisted laser desorption/ionization time-of-flight mass spectrometry detection (Sequenom Inc, San Diego, CA) according to the manufacturer's instructions. Genotyping was performed directly for all SNPs and genotypes were never based on imputation. Successful genotyping rate was >97% for all included SNPs.

*Statistical analysis*

The genotype distributions of all SNPs did not deviate significantly from that expected for a population in Hardy–Weinberg equilibrium. There was no sign of underlying population stratification in EPICOLON as tested by an independent study (21).

The study endpoints were (i) overall survival (OS), defined as the time from diagnosis to death by any cause, patients alive being censored at the last follow-up visit; (ii) disease-free survival (DFS), defined as the time from diagnosis until radiologic evidence of disease recurrence or death from any cause, patients alive without tumour recurrence being censored at the last follow-up contact and (iii) recurrence-free interval (RFI), defined as the time from diagnosis until radiologic evidence of disease recurrence. For RFI, patients who died without known tumour recurrence were censored at the last follow-up visit. Patients with TNM stage IV tumours were not considered when analyzing RFI and DFS.

Univariate Cox regression analyses were performed on each covariate to examine its influence on prognosis, computing hazard ratio (HR) with the corresponding 95% confidence interval (95% CI). Every SNPs was included independently in the univariate analysis and evaluated under genotypic and dominant genetic models, considering risk alleles as defined by results of previous genome-wide association studies for CRC incidence (12–15). Multivariate stepwise Cox regression analysis was performed by including all variables achieving a *P*-value of <0.2 in the univariate analysis. Regarding confounding variables, TNM stage and chemotherapy treatment were considered as predictors of patients' outcome and included in the final multivariate model, adjusting also by age and gender. However, MSI status was not included due to the high number of cases with missing data. Both univariate and multivariate Cox proportional hazards were done using the coxph function from Survival R Package. For all statistical analyses, results were considered significant at the *P*-value of 0.05 threshold. When commented, a Bonferroni correction for multiple comparisons corresponded to a value of 0.003125 (0.05/16 SNPs). Results are presented without correction for multiple testing to prevent type II error.

## Results

A total of 1235 cases with follow-up data available were included. Table I summarizes their demographic and clinical characteristics. In 845 patients (68, 42%), the tumour was located in the colon, and in 381 patients (3, 85%), the tumour was located in the rectum; most of the patients were in TNM stages II–III (833, 67.45%), and around half of all CRC cases received chemotherapy (581, 47.04%).

After a median follow-up of 36 months (range 0–72 months) in the whole cohort, there were 433 (35.1%) events for OS analysis. DFS and RFI analyses were based on 1001 CRC patients with stage I–III disease. With a median follow-up period of 40 months (range 1–72 months), 351 (35.1%) patients had died or showed disease recurrence, and there were 225 (22.6%) events for RFI analyses.

SNP genotypes were analysed by Cox regression in a univariate model taking into account genotypic and dominant inheritance in order to test their putative influence on OS, DFS and RFI. Results are summarized in Supplementary Table 2, available at *Carcinogenesis* Online. To proceed with the multivariate analysis, SNPs with a $P < 0.2$ for OS, DFS and RFI in the univariate analysis were further tested and covariates with a $P > 0.2$ were not further analysed.

Results for multivariate analyses are shown in Table II. Regarding OS adjusted for gender, age, TNM stage and chemotherapy treatment, both rs9929218 and rs10795668 remained statistically significant with a $P < 0.05$. The G allele of rs9929218 seemed to be linked with a

**Table I.** Demographic, clinical and tumour-related characteristics of the 1235 CRC patients evaluated in this prognosis study

| Characteristics | CRC patients (*N* = 1235) |
|---|---|
| Age mean (range) | 70.78 (26–101) |
| ≤50 y.o. | 63 (5.1%) |
| >50 y.o. | 1172 (94.9%) |
| Gender | |
| Males | 733 (59.35%) |
| Females | 502 (40.65%) |
| Tumour location | |
| Colon | 845 (68.42%) |
| Rectum | 381 (30.85%) |
| TNM tumour stage | |
| I | 168 (13.6%) |
| II | 469 (37.98%) |
| III | 364 (29.47%) |
| IV | 210 (17%) |
| Chemotherapy | |
| Yes | 581 (47.04%) |
| No | 654 (52.96%) |
| Tumour MSI | |
| Unstable | 81 (6.56%) |
| Stable | 966 (78.22%) |
| Tumour histology type | |
| Adenocarcinoma | 1142 (92.47%) |
| Mucinous/signet-ring | 82 (6.64%) |
| Tumour degree of differentiation | |
| Well/moderate | 1063 (86.07%) |
| Poor | 90 (7.29%) |
| Tumour-infiltrating lymphocytes | |
| Present | 306 (24.78%) |
| Absent | 882 (71.42%) |

y.o., years old.

**Table II.** Multivariate analysis results for OS, DFS and RFI and its association with CRC genetic susceptibility variants

| SNP_ID | Chromosomal region | Genotype | OS HR (95% CI)[a] | *P*-value |
|---|---|---|---|---|
| rs16892766 | 8q23.3 | AA | 1 | |
| | | AC + CC | 1.14 (0.87–1.49) | 0.3309 |
| | | AA | 1 | |
| | | AC | 1.16 (0.88–1.51) | 0.2877 |
| | | CC | 1.06 (0.26–4.31) | 0.9341 |
| rs10795668 | 10p14 | AA | 1 | |
| | | GA+ GG | 0.68 (0.50–0.94) | 0.0194 |
| | | AA | 1 | |
| | | GA | 0.66 (0.47–0.92) | 0.0137 |
| | | GG | 0.73 (0.53–1.01) | 0.0591 |
| rs3802842 | 11q23.1 | AA | 1 | |
| | | AC + CC | 0.87 (0.72–1.07) | 0.1809 |
| | | AA | 1 | |
| | | AC | 0.85 (0.68–1.05) | 0.1210 |
| | | CC | 0.97(0.68–1.39) | 0.8838 |
| rs9929218 | 16q22.1 | AA | 1 | |
| | | GA + GG | 0.66 (0.47–0.94) | 0.0202 |
| | | AA | 1 | |
| | | GA | 0.72(0.50–1.04) | 0.0780 |
| | | GG | 0.65 (0.45–0.93) | 0.0179 |
| rs10411210[b] | 19q13.11 | TT | 1 | |
| | | CT | 0.77 (0.28–2.11) | 0.6050 |
| | | CC | 1.11 (0.41–3.01) | 0.8362 |
| rs961253[b] | 20p12.3 | CC | 1 | |
| | | CA + AA | 1.12 (0.91–1.37) | 0.2938 |
| SNP_ID | Chromosomal region | Genotype | DFS HR (95% CI)[a] | *P*-value |
| rs16892766 | 8q23.3 | AA | 1 | |
| | | AC + CC | 1.08 (0.81–1.43) | 0.6133 |
| | | AA | 1 | |
| | | AC | 1.09 (0.82–1.46) | 0.5424 |
| | | CC | 0.55 (0.13–2.23) | 0.4017 |
| rs3802842 | 11q23.1 | AA | 1 | |
| | | AC + CC | 0.94 (0.76–1.17) | 0.5880 |
| | | AA | 1 | |
| | | AC | 0.90 (0.72–1.14) | 0.3884 |
| | | CC | 1.15 (0.79–1.68) | 0.4592 |
| rs9929218 | 16q22.1 | AA | 1 | |
| | | GA + GG | 0.88 (0.60–1.28) | 0.4988 |
| | | AA | 1 | |
| | | GA | 0.88 (0.59–1.31) | 0.5395 |
| | | GG | 0.83 (0.57–1.23) | 0.3545 |
| rs10411210 | 19q13.11 | TT | 1 | |
| | | CT + CC | 0.49 (0.16–1.55) | 0.2245 |
| | | TT | 1 | |
| | | CT | 0.37 (0.11–1.18) | 0.0929 |
| | | CC | 0.45 (0.14–1.45) | 0.1817 |
| rs961253[b] | 20p12.3 | AA | 1 | |
| | | CA | 0.91 (0.66–1.24) | 0.5515 |
| | | AA | 0.77 (0.56–1.07) | 0.1171 |
| SNP_ID | Chromosomal region | Genotype | RFI HR (95% CI)[a] | *P*-value |
| rs6983267 | 8q24.21 | TT | 1 | |
| | | GT + GG | 0.78 (0.57–1.08) | 0.1398 |
| | | TT | 1 | |
| | | GT | 0.72 (0.50–1.02) | 0.0652 |
| | | GG | 0.86 (0.59–1.24) | 0.4082 |
| rs9929218 | 16q22.1 | AA | 1 | |
| | | GA + GG | 0.77 (0.49–1.21) | 0.2573 |
| | | AA | 1 | |
| | | GA | 0.76 (0.47–1.24) | 0.2721 |
| | | GG | 0.75 (0.46–1.20) | 0.2316 |
| rs4939827[b] | 18q22.1 | CC | 1 | |
| | | CT | 1.01 (0.71–1.45) | 0.9400 |
| | | TT | 0.88 (0.59–1.30) | 0.5091 |
| rs10411210 | 19q13.11 | TT | 1 | |
| | | CT + CC | 0.46 (0.11–1.85) | 0.2720 |
| | | TT | 1 | |
| | | CT | 0.37 (0.09–1.55) | 0.1718 |
| | | CC | 0.45 (0.11–1.85) | 0.2664 |

**Table II.** *Continued*

| SNP_ID | Chromosomal region | Genotype | RFI HR (95% CI)[a] | *P*-value |
|---|---|---|---|---|
| rs4813802 | 20p12.3 | TT | 1 | |
| | | GT + GG | 1.27 (0.97–1.67) | 0.0843 |
| | | TT | 1 | |
| | | GT | 1.20 (0.90–1.61) | 0.2098 |
| | | GG | 1.48 (0.97–2.24) | 0.0691 |

SNP_ID, single nucleotide polymorphism identification.
[a]Adjusted for age, gender, TNM stage and chemotherapy treatment.
[b]Dominant and/or genotypic model were selected for multivariate analysis.

**Table III.** Multivariate analyses to evaluate the role of rs10795668 and rs9929218 on the benefit from chemotherapy in CRC patients

| SNP_ID | Chemotherapy treatment | Genotype | OS HR (95% CI)[a] | *P*-value |
|---|---|---|---|---|
| rs10795668 | Chemotherapy | AA | 1 | |
| | | GA + GG | 0.62 (0.38–1.04) | 0.0679 |
| | | AA | 1 | |
| | | GA | 0.69 (0.41–1.17) | 0.1699 |
| | | GG | 0.57 (0.34–0.97) | 0.0384 |
| | No chemotherapy | AA | 1 | |
| | | GA + GG | 0.74 (0.50–1.10) | 0.1394 |
| | | AA | 1 | |
| | | GA | 0.63 (0.41–0.96) | 0.0324 |
| | | GG | 0.85 (0.57–1.28) | 0.4457 |
| rs9929218 | Chemotherapy | AA | 1 | |
| | | GA + GG | 0.60 (0.36–1.02) | 0.0601 |
| | | AA | 1 | |
| | | GA | 0.63 (0.36–1.09) | 0.0998 |
| | | GG | 0.58 (0.34–1.01) | 0.0531 |
| | No chemotherapy | AA | 1 | |
| | | GA + GG | 0.71 (0.45–1.12) | 0.1439 |
| | | AA | 1 | |
| | | GA | 0.78 (0.48–1.25) | 0.3008 |
| | | GG | 0.67 (0.42–1.07) | 0.0963 |

SNP_ID, single nucleotide polymorphism identification.
[a]Adjusted for age, gender and TNM stage.

better OS (GG genotype, genotypic model: HR = 0.65, 95% CI 0.45–0.93, $P = 0.0179$; GG/GA genotypes, dominant model: HR = 0.66, 95% CI 0.47–0.94, $P = 0.0202$). Similarly, the G allele of rs10795668 was associated with better clinical outcome (GG genotype, genotypic model: HR = 0.73, 95% CI 0.53–1.01, $P = 0.0570$; GA genotype, genotypic model: HR = 0.66, 95% CI 0.47–0.92, $P = 0.0137$; GG/GA genotypes, dominant model: HR = 0.68, 95% CI 0.50–0.94, $P = 0.0194$). However, it should be noted that these results should be not considered formally significant if Bonferroni correction for multiple testing was applied ($P < 0.003125$). Finally, there were no statistically significant associations in the DFS and IFR analyses.

Further analyses were performed in a subset of patients in order to specifically evaluate the role of rs10795668 and rs9929218 on the benefit from chemotherapy as summarized in Table III. In a multivariate analysis performed in CRC patients treated with chemotherapy, OS seemed to be different depending on rs10795668 genotype, showing a borderline significant association between better outcome and the G allele (GG genotype, genotypic model: HR = 0.62, 95% CI 0.38–1.04, $P = 0.0679$; GG/GA genotypes, dominant model: HR = 0.57, 95% CI 0.34–0.97, $P = 0.0384$). On the other hand, in patients who did not receive chemotherapy, rs10795668 was not associated with clinical outcome (GG genotype: HR = 0.74, 95% CI 0.50–1.10, $P = 0.4457$; GG/GA genotypes: HR = 0.74, 95% CI 0.50–1.10, $P = 0.1394$). Likewise, in the stratified analysis by chemotherapy for rs9929218, we found that the effect of this SNP on OS remained borderline significant in patients treated with chemotherapy (GG genotype: HR = 0.58, 95% CI 0.34–1.01, $P = 0.0531$; GG/GA genotypes: HR = 0.60, 95% CI 0.36–1.02, $P = 0.0601$), but not in those not receiving chemotherapy (GG genotype: HR = 0.67, 95% CI 0.42–1.07, $P = 0.0963$; GG/GA genotypes: HR = 0.71, 95% CI 0.45–1.12, $P = 0.1439$).

## Discussion

In addition to influencing the risk of developing CRC, inherited genetic variants may play a role in determining the natural course of the disease and therapeutic response. Identification of inherited polymorphisms as prognostic biomarkers in CRC may allow using them in order to identify cases at higher probability of relapse or worse clinical outcome that may benefit from more intensive treatment and follow-up surveillance.

This study evaluated the prognostic or predictive value of the 16 CRC genetic susceptibility variants in a cohort of 1235 CRC patients from Spain. It is worth mentioning that the genetic variants evaluated in this study corresponded to all known SNPs for CRC genetic susceptibility except for four of them that were identified very recently. Thus, this analysis corresponds to the more complete study up to now since it included the highest number of variants and used always direct genotyping. Phipps *et al.* (19) genotyped 14 variants (rs4813802 and rs1957636 were not included) and used imputation for some SNPs.

Our results demonstrate that both rs9929218 and rs10795668 are independently associated to OS in CRC patients. Thus, the G allele of both rs9929218 and rs10795668 seemed to be linked with improved OS. Interestingly, we also performed a subanalysis in a subgroup of patients in order to know the predictive role of rs9929218 and rs10795668 on treatment efficacy. In this stratified analysis, we found that the effect of both SNPs on OS seemed more evident in patients treated with chemotherapy than in those not receiving chemotherapy. These results, if validated, could suggest a beneficial response to chemotherapy for patients carriers of the G allele of both rs9929218 and rs10795668 compared with patients carriers of the A allele. However, it should be mentioned that statistical power was limited

in these stratified analyses and results should be taken into account as exploratory and further tested in additional CRC cohorts.

rs9929218 localizes to intron 1 of the *CDH1* gene (also known as *E-cadherin*) at 16q22.1, a gene with an established role in CRC. Loss of CDH1 expression is an indicator of high tumour aggressiveness and it seems to be associated with factors of poor CRC prognosis such as vascular invasion, presence of lymph node metastases, advanced stages and poor tumour differentiation (22). The G allele of rs9929218 was linked to a higher CRC risk by previous studies (12), and it seemed to be linked with a better OS in this study. Then, it could be hypothesized that this association may be driven through a minimal downregulation of this gene. On the other hand, rs10795668 maps to an 82 kb linkage disequilibrium block within 10p14, but no known or predicted protein-coding genes are present in the 400 kb region harbouring this SNP. This genetic variant was reported to be associated with CRC risk in the European population (12) and it seemed to confer a better clinical outcome in this study.

So far, four studies have investigated the association between CRC susceptibility variants with disease progression and survival. Although there is some overlap in variants and outcomes, there are also some differences that may explain the lack of agreement between them. However, we found some coincidences after careful examination of results in previous studies that may be reinforcing our conclusions. Phipps *et al.* (19) found that the G allele of rs10795668 was associated with better OS in a multivariate analysis adjusted for gender, age and TNM stage in agreement with our results although this association was not statistically significant after Bonferroni correction (HR = 1.14, 95% CI 1.02–1.28, $P = 0.02$). They also reported a borderline association between the G allele of rs9929218 and improved OS after CRC diagnosis (HR = 1.12, 95% CI 0.99–1.26, $P = 0.07$). Likewise, Tenesa *et al.* (17) observed a trend of decreased all-cause mortality after adjustment for age and gender in patients carrying the G allele of rs9929218 (HR = 0.91, 95% CI 0.83–1.01, $P = 0.08$).

Interestingly, for both rs9929218 and rs10795668, we found a suggestive evidence of association with OS, but this association was not present in DFS and RFI analyses, suggesting that SNPs might act differentially on initiation and progression of CRC. Therefore, both rs9929218 and rs10795668 could have some influence on the aggressiveness of colorectal tumours.

Finally, this study has some limitations. First, it should be commented that our cohort sample size is probably not large enough to be able to reach stronger conclusions for the analysed variants. Also, as our results should be formally considered not statistically significant when applying multiple testing correction, additional studies in other CRC cohorts are needed in order to confirm the potential role of susceptibility variants in patient outcome. Second, the follow-up period used in this study may be causing a higher degree of censored cases. Third, we included clinical variables in our multivariate analysis that are well established as predictors of patients' outcome such TNM stage and chemotherapy treatment. However, although a number of studies have demonstrated that CRC with MSI have improved prognosis, we did not take into consideration tumours' MSI status in this analysis, due to the high percentage of missing data in our cohort. In addition, a more homogenous CRC cohort consisting of only stage II–III or MSS tumours may have provided more consistent results. Finally, patients included in this study did not receive the same first-line chemotherapy regimens and there is a possibility of confounding effects of the heterogeneous therapeutic modalities. Future studies with more detailed treatment information could elucidate the potential interaction between inherited genetic variation and treatment response.

In summary, genetic variants rs9929218 at 16q22.1 and rs10795668 at 10p14 may exert some influence in modulating patient's survival and they deserve to be further tested in additional CRC cohorts in order to confirm their potential as prognosis or predictive biomarkers.

## Supplementary material

Supplementary Note and Supplementary Tables 1 and 2 and can be found at http://carcin.oxfordjournals.org/



## Funding

Fondo de Investigación Sanitaria (CP 03-0070 to S.C.B., PS09/02368 to C.F.R.); Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd; to J.M.); Instituto de Salud Carlos III (to CIBERehd and Centro de Investigación Biomédica en Red de Enfermedades Raras); Fondo de Investigación Sanitaria/FEDER (08/0024, 08/1276, PS09/02368, 10/00918, 11/00219, 11/00681); Instituto de Salud Carlos III (Acción Transversal de Cáncer); Xunta de Galicia (07PXIB9101209PR); Ministerio de Ciencia e Innovación (SAF2010-19273); Asociación Española contra el Cáncer (Fundación Científica y Junta de Barcelona); Fundació Olga Torres (to S.C.B. and C.R.P.); FP7 CHIBCHA Consortium (to S.C.B., A.C.); COST Action (BM1206 to S.C.B., C.R.P.).

## Acknowledgements

We are sincerely grateful to all patients participating in this study who were recruited in 25 (EPICOLON 1) and 14 (EPICOLON 2) Spanish hospitals as part of the EPICOLON project. We are also indebted to the USC and UPF nodes of Spanish National Genotyping Centre (CEGEN-ISCIII) and the Genomics Unit of the Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) for technical help. This work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

*Conflict of Interest Statement:* None declared.

## References

1. Ferlay,J. *et al.* (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
2. Chang,G.J. *et al.* (2009) Practical application of a calculator for conditional survival in colon cancer. *J. Clin. Oncol.*, **27**, 5938–5943.
3. Edge,S.B. *et al.* (eds) (2010) *AJCC Cancer Staging Manual.* 7th edn. Springer, New York, NY.
4. Dray,X. *et al.* (2003) Influence of dietary factors on colorectal cancer survival. *Gut*, **52**, 868–873.
5. Haydon,A.M. *et al.* (2006) Effect of physical activity and body size on survival after diagnosis with colorectal cancer. *Gut*, **55**, 62–67.
6. Hong,S.P. *et al.* (2012) The differential impact of microsatellite instability as a marker of prognosis and tumour response between colon cancer and rectal cancer. *Eur. J. Cancer*, **48**, 1235–1243.
7. Kalady,M.F. *et al.* (2012) BRAF mutations in colorectal cancer are associated with distinct clinical characteristics and worse prognosis. *Dis. Colon Rectum*, **55**, 128–133.
8. Lindström,L.S. *et al.* (2007) Familial concordance in cancer survival: a Swedish population-based study. *Lancet Oncol.*, **8**, 1001–1006.
9. Koessler,T. *et al.* (2009) Common germline variation in mismatch repair genes and survival after a diagnosis of colorectal cancer. *Int. J. Cancer*, **124**, 1887–1891.
10. Lin,M. *et al.* (2012) Genetic polymorphisms in MicroRNA-related genes as predictors of clinical outcomes in colorectal adenocarcinoma patients. *Clin. Cancer Res.*, **18**, 3982–3991.
11. Slattery,M.L. *et al.* (2012) Toll-like receptor genes and their association with colon and rectal cancer development and prognosis. *Int. J. Cancer*, **130**, 2974–2980.
12. Tenesa,A. *et al.* (2009) New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat. Rev. Genet.*, **10**, 353–358.
13. Houlston,R.S. *et al.* (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.*, **42**, 973–977.
14. Tomlinson,I.P. *et al.* (2011) Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.*, **7**, e1002105.
15. Dunlop,M.G. *et al.* (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat. Genet.*, **44**, 770–776.
16. Kinnersley,B. *et al.* (2012) The TERT variant rs2736100 is associated with colorectal cancer risk. *Br. J. Cancer*, **107**, 1001–1008.
17. Tenesa,A. *et al.* (2010) Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. *Clin. Cancer Res.*, **16**, 3754–3759.
18. Xing,J. *et al.* (2011) GWAS-identified colorectal cancer susceptibility locus associates with disease prognosis. *Eur. J. Cancer*, **47**, 1699–1707.

19. Phipps,A.I. *et al.* (2012) Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. *Gastroenterology*, **143**, 51–4.e4.
20. Piñol,V. *et al.* (2005) Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA*, **293**, 1986–1994.
21. Fernández-Rozadilla,C. *et al.* (2010) Colorectal cancer susceptibility quantitative trait loci in mice as a novel approach to detect low-penetrance variants in humans: a two-stage case-control study. *Cancer Epidemiol. Biomarkers Prev.*, **19**, 619–623.
22. Peña,C. *et al.* (2005) E-cadherin and vitamin D receptor regulation by SNAIL and ZEB1 in colon cancer: clinicopathological correlations. *Hum. Mol. Genet.*, **14**, 3361–3370.

*Received February 26, 2013; revised April 26, 2013; accepted May 21, 2013*

World Journal of Gastroenterology

Online Submissions: http://www.wjgnet.com/esps/
bpgoffice@wjgnet.com
doi:10.3748/wjg.v20.i8.1961

*World J Gastroenterol* 2014 February 28; 20(8): 1961-1971
ISSN 1007-9327 (print) ISSN 2219-2840 (online)


TOPIC HIGHLIGHT

WJG 20th Anniversary Special Issues (5): Colorectal cancer

# New genes emerging for colorectal cancer predisposition

Clara Esteban-Jurado, Pilar Garre, Maria Vila, Juan José Lozano, Anna Pristoupilova, Sergi Beltrán, Anna Abulí, Jenifer Muñoz, Francesc Balaguer, Teresa Ocaña, Antoni Castells, Josep M Piqué, Angel Carracedo, Clara Ruiz-Ponte, Xavier Bessa, Montserrat Andreu, Luis Bujanda, Trinidad Caldés, Sergi Castellví-Bel

Clara Esteban-Jurado, Anna Abulí, Jenifer Muñoz, Francesc Balaguer, Teresa Ocaña, Antoni Castells, Josep M Piqué, Sergi Castellví-Bel, Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, 08036 Barcelona, Catalonia, Spain
Anna Abulí, Xavier Bessa, Montserrat Andreu, Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, 08003 Barcelona, Spain
Maria Vila, Juan José Lozano, Bioinformatics platform, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), 08036 Barcelona, Spain
Anna Pristoupilova, Sergi Beltrán, Centre Nacional d'Anàlisi Genòmica (CNAG), Parc Científic de Barcelona, 08028 Barcelona, Spain
Angel Carracedo, Clara Ruiz-Ponte, Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, Santiago de Compostela, University of Santiago de Compostela, 15706 Galicia, Spain
Angel Carracedo, Center of Excellence in Genomic Medicine Research, King Abdulaziz University, 21589 Jeddah, Kingdom of Saudi Arabia
Luis Bujanda, Gastroenterology Department, Hospital Donostia, Networked Biomedical Research Centre for Hepatic and Digestive Diseases (CIBEREHD), Basque Country University, 20080 San Sebastián, Spain
Pilar Garre, Trinidad Caldés, Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), 28040 Madrid, Spain
Author contributions: Esteban-Jurado C, Garre P, Vila M, Lozano JJ, Pristoupilova A, Beltrán S, Abulí A, Muñoz J, Balaguer F, Ocaña T, Castells A, Piqué JM, Carracedo A, Ruiz-Ponte C, Bessa X, Andreu M, Bujanda L, Caldés T and Castellví-Bel S contributed to this paper.
Supported by SCB is supported by a contract from the Fondo de Investigación Sanitaria, No. CP 03-0070; CEJ and JM are supported by a contract from CIBERehd; CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III; Fondo de Investigación Sanitaria/FEDER, No.11/00219 and No. 11/00681, Instituto de Salud Carlos III (Acción Transversal de Cáncer), Xunta de Galicia, No. 07PXIB9101209PR, Ministerio de Ciencia e Innovación, No. SAF2010-19273, Asociación Española contra el Cáncer (Fundación Científica GCB13131592CAST y Junta de Barcelona), Fundació Olga Torres (SCB and CRP), FP7 CHIBCHA Consortium (SCB and ACar), and COST Action BM1206 (SCB and CRP)
Correspondence to: Sergi Castellví-Bel, PhD, Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Rosselló 153 planta 4, 08036 Barcelona, Catalonia, Spain. sbel@clinic.ub.es
Telephone: +34-93-2275418 Fax: +34-93-3129405
Received: October 1, 2013 Revised: November 7, 2013
Accepted: January 14, 2014
Published online: February 28, 2014

## Abstract

Colorectal cancer (CRC) is one of the most frequent neoplasms and an important cause of mortality in the developed world. This cancer is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian syndromes account for about 5% of the total burden of CRC, with Lynch syndrome and familial adenomatous polyposis the most common forms. Excluding hereditary forms, there is an important fraction of CRC cases that present familial aggregation for the disease with an unknown germline genetic cause. CRC can be also considered as a complex disease taking into account the common disease-commom variant hypothesis with a polygenic model of inheritance where the genetic components of common complex diseases correspond mostly to variants of low/moderate effect. So far, 30 common, low-penetrance susceptibility variants have been identified for CRC. Recently, new sequencing technologies including exome- and whole-genome sequencing have permitted to add

a new approach to facilitate the identification of new genes responsible for human disease predisposition. By using whole-genome sequencing, germline mutations in the *POLE* and *POLD1* genes have been found to be responsible for a new form of CRC genetic predisposition called polymerase proofreading-associated polyposis.



**Key words:** Colorectal neoplasm, genetic predisposition to disease; Next generation sequencing; Genotype-phenotype correlation; Genetic variant; Single nucleotide polymorphism

**Core tip:** Colorectal cancer (CRC) is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian syndromes account for about 5% of the total burden of CRC. Excluding hereditary forms, there is an important fraction of CRC cases that present familial aggregation for the disease with an unknown germline genetic cause. Recently, new sequencing technologies have permitted to add a new approach to identify new genes responsible for human disease predisposition. By doing so, germline mutations in the *POLE* and *POLD1* genes have been found to be responsible for a new form of CRC genetic predisposition.

Esteban-Jurado C, Garre P, Vila M, Lozano JJ, Pristoupilova A, Beltrán S, Abulí A, Muñoz J, Balaguer F, Ocaña T, Castells A, Piqué JM, Carracedo A, Ruiz-Ponte C, Bessa X, Andreu M, Bujanda L, Caldés T, Castellví-Bel S. New genes emerging for colorectal cancer predisposition. *World J Gastroenterol* 2014; 20(8): 1961-1971 Available from: URL: http://www.wjgnet.com/1007-9327/full/v20/i8/1961.htm DOI: http://dx.doi.org/10.3748/wjg.v20.i8.1961

## INTRODUCTION

Colorectal cancer (CRC) is one of the most frequent neoplasms and an important cause of mortality in the developed world. Approximately 5% of the population develops CRC and this figure is expected to rise as life expectancy increases[1]. For 2015, approximately 473200 new cases are predicted and 233900 individuals will die from this disease in Europe[2]. When taking into account both genders together, it corresponds to the most frequent neoplasm in Spain. Although there has been recent progress in CRC clinical management and treatment that has permitted to reduce the number of cases in the developed countries, it is foreseen that its incidence will increase worldwide with developing nations bearing the brunt of the rise. The incidence of CRC varies widely between countries, depending on their degree of development and also on the quality of their cancer registries[3]. Around 60% of cases are diagnosed in the developed world[3]. The highest incidence rates are found in Australia and New Zealand, North America and Europe, whereas the lowest rates are registered in Africa and South-Central Asia[2] (Figure 1).

CRC survival depends on the stage of disease at diagnosis and typically ranges from a 90% 5-year survival rate for cancers detected at the localized stage to 10% for people diagnosed of a distant metastatic cancer[4]. The lifetime risk of CRC in the general population is about 5% in Western countries, but the likelihood of CRC diagnosis increases progressively with age, being more than 90% in individuals over age 50, and 70% of these over 65[4].

CRC is believed to develop from polyps, which have been traditionally classified as either hyperplastic or adenomatous. Until recently, according to the adenoma-carcinoma sequence proposed by Vogelstein *et al*[5] the adenoma was considered the exclusive precursor lesion while hypeplastic polyps were deemed to have no malignant potential. However, it is now recognized that lesions, formerly classified as hyperplastic, represent a heterogeneous group of polyps with a characteristic serrated morphology, some of which have a significant risk of malignant transformation through the serrated neoplasia pathway[4].

## GENETIC AND ENVIRONMENTAL RISK FACTORS

As other complex diseases, CRC is caused by both genetic and environmental factors. The role of environmental factors on colorectal carcinogenesis is indicated by the increase in CRC incidence in parallel with economic development and adoption of Western diets and lifestyles, responsible for the high incidence of CRC in industrialized countries[7]. Although the majority of CRC occur mostly in industrialized countries, their incidence rates are rapidly rising in economically transitioning countries in the world[8]. These observations highlight the importance of environmental influences on CRC development and suggest that Western lifestyle risk factors play an important role in the etiology of the disease. However, although environmental causes such as smoking and diet are undoubtedly risk factors for CRC, twin studies have shown that 35% of the variation in CRC susceptibility involves inherited genetic differences[9,10]. In that sense, a minority of CRC cases (about 5%) show strong familial aggregation and belong to the well-known hereditary CRC forms mainly caused by germline mutations in *APC*, *MUTYH* and the DNA mismatch repair genes[11]. Approximately 30% of CRC cases show some family history of the disease but do not fit in the previous category and are regarded as familial CRC, whereas a majority of cases do not show any familial aggregation and correspond to sporadic CRC. For instance, familial CRC accounted for about 30% of all CRC cases in an

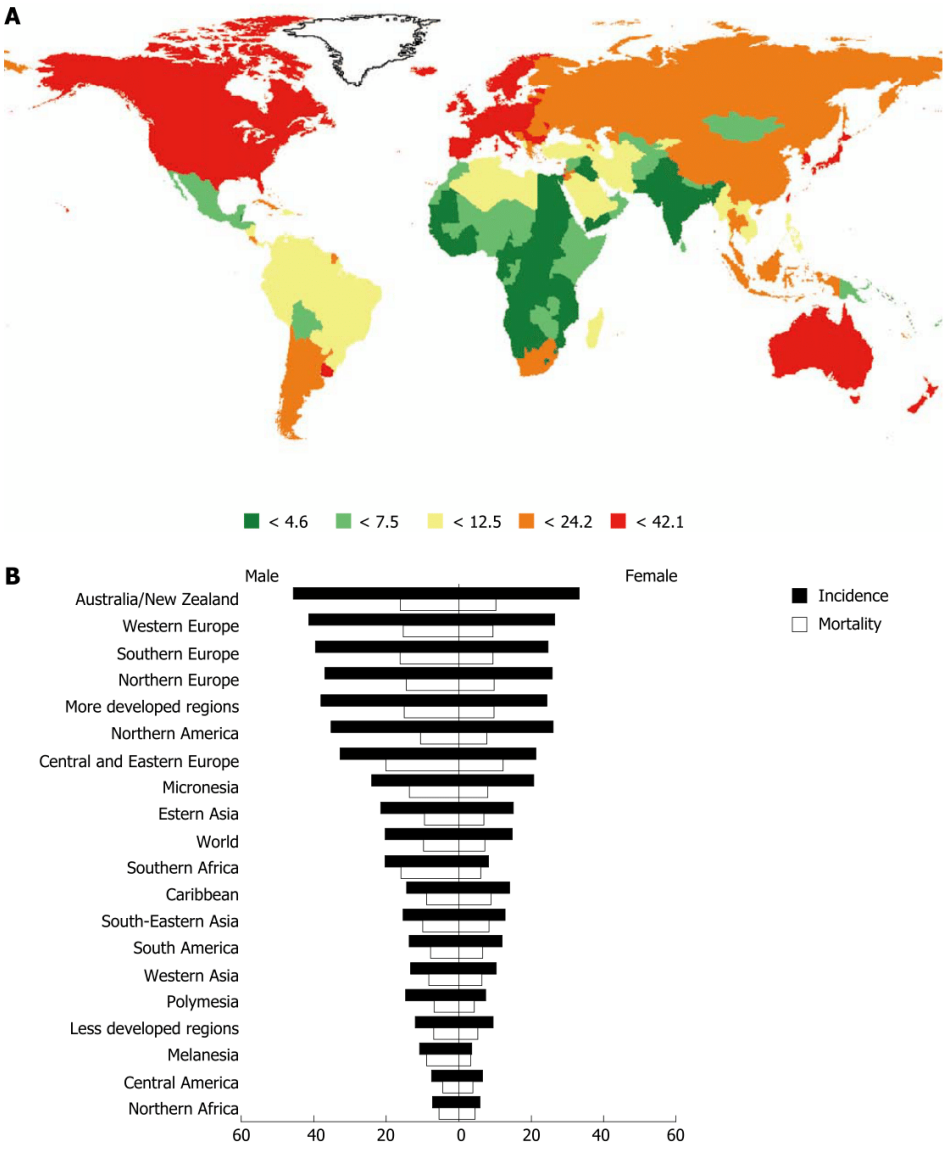



**Figure 1 Colorectal cancer in the world.** A: Estimated age-standarized incidence rate per 100000 individuals (both genders and all ages); B: Estimated age-standarized incidence and mortality rate per 100000 individuals by genders (data adapted from Ferlay *et al*[2]).

epidemiological study in the Spanish population[12].

## HEREDITARY CRC

Mendelian cancer syndromes account for about 5% of the total burden of CRC[11]. The genetic components involved in these less frequent hereditary forms were successfully identified using linkage analysis in the past two decades and they correspond to rare highly penetrant alleles that predispose to CRC. Two major subgroups can be clinically divided on the presence or absence of colorectal polyposis. An overview of all CRC syndromes is provided in Table 1. The most frequent forms are hereditary nonpolyposis colorectal cancer and familial polyposis syndrome, which are further described below.

Hereditary Nonpolyposis Colorectal Cancer (HNPCC; MIM No.120435), also known as Lynch syndrome, is the most common form of hereditary CRC accounting for at least 3% of all CRC. HPNCC is an autosomal dominant syndrome defined clinically by the Amsterdam criteria (Table 2), which are used in clinical practice to identify individuals at risk for this disease who require further evaluation and are based on strong familial aggregation and early onset. It is characterized by early-onset CRC (mean age at diagnosis, approximately 45 years), excess synchronous and metachronous colorectal neoplasms and right-sided predominance compared to sporadic neoplasms. In addition, there is an increased

**Table 1 Hereditary colorectal cancer genes**

| | Gene | Chromosome | Mendelian pattern | Function |
|---|---|---|---|---|
| Familial adenomatous polyposis | *APC* | 5q | AD | Regulation of canonical Wnt signaling pathway |
| | *MUTYH* | 1p | AR | Base-excision repair |
| Hereditary non-polyposis CRC (Lynch syndrome) | *MLH1* | 3p | AD | Mismatch repair |
| | *MSH2* | 2p | AD | Mismatch repair |
| | *MSH6* | 2p | AD | Mismatch repair |
| | *PMS2* | 7p | AD | Mismatch repair |
| Peutz-Jeghers | *LKB1* | 19p | AD | Regulation of Wnt signaling pathway |
| Juvenile polyposis | *SMAD4* | 18q | AD | TGFBR signaling pathway |
| | *BMPR1A* | 10q | AD | TGFBR signaling pathway |
| Cowden's disease | *PTEN* | 10q | AD | Negative regulation of PI3K signaling |
| | *KLLN* | 10q | AD | Apoptotic process |

CRC: Colorectal cancer; AD: Autosomal dominant; AR: Autosomal recessive; TGFBR: Transforming growth factor beta receptor; PI3K: Phosphatidylinositol 3-kinase.

**Table 2 Amsterdam criteria in Lynch syndrome**

| Amsterdam criteria I | Amsterdam criteria II |
|---|---|
| At least three relatives with CRC; all of the following must be met: | At least three relatives with colorectal, endometrial, small bowel, ureter, or renal pelvis cancer; all of the following must be met: |
| One affected individual is a first degree relative of the other two | One affected individual is a first degree relative of the other two |
| At least two successive generations affected | At least two successive generations affected |
| At least one CRC diagnosed before the age of 50 years | At least one tumor diagnosed before the age of 50 years |
| Familial adenomatous polyposis has been excluded | Familial adenomatous polyposis has been excluded |

CRC: Colorectal cancer.

incidence of extracolonic neoplasms (endometrial, small bowel, gastric, upper urinary tract, ovarian, brain and pancreatic tumors) being endometrial cancer the most common malignancy associated with Lynch syndrome. Indeed, Lynch syndrome is responsible for approximately 2% of all endometrial cancers[13]. The lifetime risk for developing CRC in individuals affected with Lynch syndrome have been estimated in approximately 66% for men and about 43% for women. The cumulative risk of endometrial cancer is approximately 40% and the lifetime risk of endometrial cancer or CRC in women is approximately 73%[14]. Lynch syndrome tumors develop as a consequence of defective DNA mismatch repair (MMR) associated with germline mutations in the MMR genes, including *MSH2* on chromosome 2p16, *MLH1* on chromosome 3p21, *MSH6* on chromosme 2p16, and *PMS2* on chromosome 7q11. In addition, germline epigenetic inactivation of *MLH1*, by hypermethylation of its promoter, can also lead to Lynch syndrome[15]. Recently, germline deletions of the 3' region of *EPCAM* gene were found in a subset of families with Lynch syndrome. This deletion leads to promoter hypermetilation of *MSH2*, located upstream of the deleted gene[16]. The MMR system is necessary to maintaining genomic fidelity by correcting single-base mismatches and insertion-deletion loops during DNA replication. As a consequence, Lynch syndrome tumors accumulate errors in short repetitive sequences, a phenomenon called microsatellite instability (MSI), which is considered a landmark for this disease. It is noteworthy to mention that in sporadic MSI CRC cancers, loss of expression of MLH1 due to hypermethilation of its promoter is a frequent event, and it is linked with the somatic mutation V600E in the *BRAF* gene[17].

Familial Adenomatous Polyposis (FAP; MIM No.175100) is the most common polyposis syndrome, classically characterized by the development of hundreds to thousands of adenomatous polyps in the rectum and colon. FAP is an autosomal dominant disease and accounts for approximately 1% of all CRC cases. In the majority of patients polyps begin to develop during the second decade of life and nearly 100% of untreated patients will have malignancy by ages 40-50 years. Individuals with FAP can also develop a variety of extracolonic manifestations, including cutaneous lesions such as fibromas, lipomas, sebaceous and epidermoid cysts, facial osteomas, congenital hypertrophy of the retinal pigment epithelium, desmoid tumours and extracolonic cancers (tyroid, liver, biliary tract and central nervous system)[18]. Duodenal cancer is the second most common malignancy in FAP, with a lifetime risk of approximately 4%-12%. *Adenomatous polyps* are also found in the stomach and duodenum, especially the *periampullary area* and can develop into adenocarcinomas. After colectomy, periampullary carcinoma is the most common malignancy, occurring in approximately 5%-6% of the patients[19]. Some lesions such as skull and mandible osteomas, dental abnormalities and fibromas are indicative of the Gardner syndrome, a clinical variant of FAP where the extracolonic features are prominent. FAP is caused by germline mutations in the *APC* gene on

chromosome 5q22, which encodes a tumor suppressor protein that plays an important role in the *Wnt* signaling pathway. Most patients have a family history of colorectal polyps and cancer, but de novo *APC* mutations are responsible for approximately 25% of cases[11].

## APPROACHES TO IDENTIFY GENETIC VARIANTS FOR CRC RISK

Among CRC cases of unknown inherited cause, there are large families with a clear positive family history of CRC, which are likely caused by highly penetrant risk loci. In the last few years, it has been described that approximately 40%-50% of CRC families that fulfill the Amsterdam Criteria for Lynch syndrome do not show evidence of MMR deficiency. Studying relatives in such families showed that CRC risk is lower than in those families with Lynch syndrome, that CRC diagnosis is in average 10 years later and that there is no increased incidence of extracolonic malignancies[20,21]. The designation of Familial CRC type X was proposed to describe this type of CRC clustering[20]. Meanwhile, genes responsible for this new entity are unknown, and most patients are included in the heterogeneous group of non-syndromic familial CRC.

Recently, there have been several efforts to identify additional genetic factors that predispose to CRC with uneven success. Linkage analysis in affected families were able to pinpoint chromosomal regions of interest such as 9q22 and 3q22 but no clear CRC predisposition genes were identified after screening for interesting candidates within these areas[22,23].

Since the known high-risk syndromes only account for a small minority of CRC cases, there has been an intensified search for low-penetrance genetic variants that probably underlie part of the hereditary predisposition and together with environmental interactions are responsible for CRC as a complex disease. Therefore, the common disease-common variant hypothesis has been also considered, being a polygenic model of inheritance where the genetic components of common complex diseases correspond mostly to variants of low/moderate effect (typically < 1.5-fold increased risk) that appeared at an elevated frequency in the population (> 5%), each exerting a small influence on disease risk. In this regard, case-control genome-wide association studies (GWAS) have been more successful by discovering up to now 31 common, low-penetrance genetic variants involved in CRC susceptibility[24-32] (Table 3).

## OVERVIEW OF NEW SEQUENCING TECHNOLOGIES

Until recently, the Sanger method was the dominant approach and gold standard for DNA sequencing[33]. Next generation sequencing (NGS), also called massive parallel sequencing, is based in sequencing millions of DNA fragments at the same time[34]. It consists in a mix of techniques of DNA shearing, PCR amplification and sequencing through modified nucleotides attached to a reversible terminator and a fluorophore, which permits fluorescent detection with an imaging system. Once the fragments are sequenced, they are assembled de novo or aligned with a reference genome by bioinformatics tools and positions that differ are designated as variants. Variants are annotated assigning their position in a gene, retrieving frequency information from genetic variation databases and categorizing them by their functional class (nonsense, missense, synonymous, frameshift, splicing, intronic, untranslated regions, regulatory).

The advantage of NGS comparing with conventional Sanger sequencing is that millions of DNA fragments are sequenced at the same time which permits to have an entire human genome sequenced in few days, and the cost is greatly reduced. However, data analysis that includes filtering of the false positives and prioritization of the candidate variants for the studied phenotypic condition is the main bottleneck of NGS, being time consuming and requiring different strategies that will be discussed later. Another disadvantage of NGS is that PCR amplification and sequencing reaction steps systematically introduce mistakes, producing base-calling errors and shorter sequenced fragments that difficult the mapability to the reference sequence. Due to recent technology and variant calling algorithm improvements, NGS is probably nowadays more accurate than conventional Sanger sequencing[35]. However, although there is a very small error rate associated with NGS, a huge amount of false positives are still detected since millions of variants are sequenced per genome[36]. Thus, after data analysis and selection of the candidate variants it is necessary to validate them using a technology with a different systematic error associated, such as conventional Sanger sequencing, which increases the costs and time of the analysis.

In order to detect genomic sequence variation by NGS, it is possible to sequence the entire genome (whole-genome sequencing, WGS) or capture and sequence only specific regions of interest (targeted enrichment). The most commonly used application for NGS target enrichment in the human genome is whole-exome sequencing (WES) that captures and amplifies the entire protein coding sequence (1% genome), flanking intronic regions and some noncoding RNAs[37]. It is a cost effective approach for detecting rare high penetrance variants based on the fact that for Mendelian disorders over the 85% of causative mutations are in coding regions. One advantage of WES is that is about much cheaper than WGS, which allows sequencing a larger number of samples with better accuracy or coverage. The term coverage corresponds to the read depth or depth and it is the average number of times that a nucleotide has been sequenced in a different sequencing read. Also, the data analysis pipelines are simpler in WES than for

**Table 3 Genetic variants associated with colorectal cancer susceptibility identified by genome-wide association studies (as for September 2013)**

| SNP | Region | Gene | Sample size (cases/controls) | Effect size OR (95%CI) |
|---|---|---|---|---|
| rs693267 | 8q24.21 | *MYC* | 8264/6206 | 1.21 (1.15-1.27) |
| rs4939827 | 18q21.1 | *SMAD7* | 8413/6949 | 1.18 (1.12-1.23) |
| rs16892766 | 8q23.3 | *EIF3H* | 18831/18540 | 1.25 (1.19-1.32) |
| rs3802842 | 11q23.1 | ? | 14500/13294 | 1.12 (1.07-1.17) |
| rs4779584 | 15q13.3 | *GREM1* | 7922/6741 | 1.26 (1.19-1.34) |
| rs10795668 | 10p14 | ? | 18831/18540 | 1.12 (1.10-1.16) |
| rs4444235 | 14q22.2 | *BMP4* | 20288/20971 | 1.11 (1.08-1.15) |
| rs9929218 | 16q22.1 | *CDH1* | 20288/20971 | 1.10 (1.06-1.12) |
| rs10411210 | 19q13 | *RHNP2* | 20288/20971 | 1.15 (1.10-1.20) |
| rs961253 | 20p12.3 | *BMP2* | 20288/20971 | 1.12 (1.08-1.16) |
| rs6691170 | 1q41 | *DUSP10* | 18185/20197 | 1.06 (1.03-1.09) |
| rs10936599 | 3q26.2 | *TERC* | 18185/20197 | 0.93 (0.91-0.96) |
| rs11169552 | 12q13.3 | ? | 18185/20197 | 0.92 (0.90-0.95) |
| rs4925386 | 20q13.33 | *LAMA5* | 18,185/20,197 | 0.93 (0.91-0.95) |
| rs1957636 | 14q22.2 | *BMP4* | 24910/26275 | 1.08 (1.05-1.11) |
| rs4813802 | 20p12.3 | *BMP2* | 24910/26275 | 1.09 (1.06-1.12) |
| rs2736100 | 5p15.33 | *TERT* | 16039/16430 | 1.07 (1.04-1.10) |
| rs1321311 | 6p21 | *CDKN1A* | 21096/19555 | 1.10 (1.07-1.13) |
| rs3824999 | 11q13.4 | *POLD3* | 21096/19555 | 1.10 (1.07-1.13) |
| rs5934683 | Xp22.2 | *SHROOM2* | 21096/19555 | 1.07 (1.04-1.10) |
| rs12080929 | 1p33 | *SLC5A9* | 2317/2447 | 0.86 (0.78-0.95) |
| rs11987193 | 8p12 | *DUSP4* | 2317/2447 | 0.78 (0.70-0.87) |
| rs10774214 | 12p13.32 | *CCND2* | 11870/14190 | 1.04 (1.00-1.09) |
| rs647161 | 5q31.1 | *PITX1* | 11870/14190 | 1.07 (1.02-1.11) |
| rs2423279 | 20p12.3 | *HAQ1* | 11870/14190 | 1.07 (1.03-1.12) |
| rs11903757 | 2q32.3 | *NABP1* | 15752/21771 | 1.16 (1.10-1.22) |
| rs10911251 | 1q25.3 | *LAMC1* | 15752/21771 | 1.09 (1.06-1.13) |
| rs3217810 | 12p13.32 | *CCND2* | 13654/16022 | 1.20 (1.12-1.28) |
| rs3217901 | 12p13.32 | *CCND2* | 15752/21771 | 1.10 (1.06-1.14) |
| rs59336 | 12q24.21 | *TBX3* | 15752/21771 | 1.09 (1.06-1.13) |

WGS. However, WES need for larger amounts of DNA sample and only covering coding variants are among the shortcomings for this technique. It is noteworthy mentioning that NGS target enrichment can also be used to sequence a panel of known genes for clinical diagnosis[33] or regions of linkage disequilibrium for a disease.

The election of individuals to sequence is a critical process to take into account for further analysis and will depend of the disease phenotype and pattern of genetic inheritance. Also, it should be noted that is possible to obtain good results with NGS when using carefully selected patients in contrast to GWAS, where number of cases and controls that are compared needs to be much higher in order to obtain statistically significant findings. For diseases with genetic heterogeneity as human cancers, different strategies can be used including the selection of families with strong disease aggregation or sequencing sporadic cases with early onset for the disease. Both situations are suggestive of the involvement of a germline predisposition. When focusing in families with several affected members, sequencing can be performed in several cases in each family and only those shared variants will be taken into account. On the other hand, if sporadic early-onset cases are chosen, genes with variants in different individuals can be selected. Sequencing non affected individuals of the same family can be useful to discard the variants shared with patients, as long as the disease has complete penetrance or it is quite likely that the non affected individuals will not express the disease in their lifetime.

### *Data filtering and prioritization in NGS*

Based on several recently sequenced individual genomes a pattern has been recognized that, in general, approximately 3-4 million variants are expected to be found in a human genome by WGS[38] and 20000 single nucleotide variants are to be found in a human exome by WES[39], so it is necessary to do a filtering strategy in order to eliminate as many false positives as possible. The first filter to apply is for those variants that do not pass a coverage threshold (typically 5-10x).

The second filtering process is based on the kind of inheritance, penetrance and frequency of the disease. Regarding the inheritance, for monogenic diseases where unrelated affected individuals have been sequenced, it is necessary to select only the genes that have variants in all of them. If a disease with genetic heterogeneity is studied, variants shared between the affected members of the same family and not shared by the unaffected ones will be chosen. Also, if dominant inheritance is present heterozygous mutations will be expected, whereas homozygous or compound heterozygous mutations will be selected in the case of recessive inheritance. However, variants in the non pseudoautosomal regions of X chromosome for dominant inheritance have to take into account also. In men, they will be annotated as

homozygous and it is necessary to select these variants too and not filter them out. Regarding variant effect on protein, it is assumed that high penetrance mutations are causative of Mendelian disorders with a large effect on protein function. Therefore, a positive selection for variants with a strong effect on the protein is advised including those affecting canonical splice sites, as well as frameshift, nonsense and missense mutations.

Proportionally, more deleterious than polymorphic variants are expected to be rare so a causative mutation is not expected to be present at a high frequency in the general population[40]. Thus, variants present at high frequency at reference genetic variation databases can be removed as potential candidates to be causative mutations.

However, many variants can still remain for each individual as putative causative mutations for the disease after filtering. A logical approach to reduce the number of candidate variants is to prioritize the mutations in genes previously implicated with the studied disease. Also, since the protein products of genes responsible for the same disorder tend to physically interact with each other so as to carry out certain biological functions, another approach for the prioritization strategy will be to include genes interacting with those previously implicated with the studied disease[41]. Finally, knowledge of the pathways implicated in a disease can be helpful also to prioritize those genes related with those pathways. After filtering and prioritization, a list of candidate variants will be available.

Sequencing validation by Sanger sequencing or any other PCR technology designed to detect a specific nucleotide change is necessary after NGS to confirm the prioritized variants and exclude sequencing artifacts. Also, segregation analysis in families permits to check if a candidate variant segregates correctly with the disease. Therefore, affected members need to be carriers and non-affected individuals old enough to be expressing the disease should be non-carriers in order to find correct segregation of the candidate variant with the studied disease. Additionally in the case of hereditary cancer, when heterozygous candidate variants with correct segregation are identified, it is necessary to confirm if there is loss of the second allele in the tumor DNA in order to establish the candidate gene as a tumor suppressor gene. Case-control screening studies can also be performed in order to identify additional carriers of the candidate variants in ample disease cohorts and further demonstrate its absence in controls. Finally, functional assessment of the candidate variant and affected gene will be also necessary to further confirm the negative effect of the variant in the protein and prove its involvement in disease development by *in vitro* studies and animal models.

## NEW GENES IDENTIFIED FOR CRC GENETIC PREDISPOSITION

New sequencing technologies made available recently including exome- and whole-genome sequencing have permitted to add a new approach to facilitate the identification of new genes responsible for human disease predisposition. Indeed, some seminal efforts have been already completed very recently for CRC. However, before these high-throughput technologies have yielded results in CRC families, some previous low-throughput sequencing studies reported directed screening of some plausible gene candidates for various reasons. Most studies have not been replicated in additional cohorts and, therefore, there is a strong need to further validate them before considering these genes as hereditary CRC genes perse.

A truncating mutation was found in the *CDH1* gene in a family with predisposition to CRC and gastric cancer, suggesting that germline mutations in this gene could contribute to early onset CRC and gastric cancer[42]. Later on, the *AXIN2* gene, a component of the Wnt signaling, was found to be mutated in a Finnish family with severe permanent tooth agenesis and CRC[43]. In a subsequent study in patients with unexplained hamartomatous or hyperplastic/mixed polyposis, two early-onset disease patients were found to have germline mutations in *ENG*, encoding endoglin, previously associated only with hereditary hemorrhagic telangiectasia[44]. This study suggested *ENG* as a new predisposition gene for juvenile polyposis, however this gene was found to be mutated in an additional study only in patients with $\geqslant$ 5 cumulative lifetime gastrointestinal polyps but not in juvenile polyposis[45]. *EPHB2* was also evaluated as a candidate tumor suppressor gene for CRC and found mutated in 3 out 116 population-based familial CRC cases, suggesting this gene may contribute to a small fraction of hereditary CRC[46]. In 2009, the *GALNT12* gene was also found mutated in the germline of 6 CRC patients[47]. This gene encodes one of the proteins involved in mucin type O-linked glycosylation and it is located in chromosomal region 9q22, previously involved in familial CRC. A more recent study detected additional deleterious variants in this gene reinforcing its role as a new candidate gene for hereditary CRC[48]. Also, an inherited duplication affecting the protein tyrosine phosphatase *PTPRJ* and causing epigenetic silencing of this gene was detected in a CRC family without polyposis and MMR alteration, being indicative of its contribution to a fraction of hereditary CRC with unknown basis[49]. Afterwards, *BMP4*, a gene close to 2 of the CRC genetic susceptibility variants identified by GWAS, was also screened in 504 genetically enriched CRC and 3 pathogenic mutations were identified[50]. Then, it could be plausible that some genes identified by CRC GWAS could be also involved in hereditary CRC. In 2011, the *BMPR1A* gene, previously involved in juveline polyposis and mixed polyposis germline predisposition, was also found mutated in familial CRC type X cases, expanding its phenotype also to this CRC hereditary form[51]. Finally, Cowden syndrome individuals without germline *PTEN* mutations were found to carry germline mutations in *PIK3CA* and *AKT1*, expanding the genetic spec-

trum of this hereditary CRC condition[52].

Regarding NGS studies to identify new CRC predisposition genes, Palles *et al*[53] reported very recently the identification of germline mutations in the *POLE* (polymerase (DNA directed), epsilon, catalytic subunit) and *POLD1* (polymerase (DNA directed), delta 1, catalytic subunit) genes in individuals with multiple colorectal adenomas, carcinoma or both, using whole-genome sequencing[54]. *POLE* and *POLD1* encode the catalytic and proofreading activities of the leading-strand DNA polymerase ε and the lagging-strand polymerase δ. The proofreading capacity of the exonuclease domain is essential for the maintenance of replication fidelity and may act not only on newly misincorporated bases but also on mismatches produced by non-proofreading polymerases. They identified a heterozygous p.Leu424Val missense variant in *POLE* DNA polymerase in a family affected with adenomas and CRC and a p.Ser478Asn missense variant in *POLD1* in a second family with CRC. The same *POLD1* p.Ser478Asn variant was also identified in the affected members of an independent family. These findings were further validated in a screen of 3,085 individuals with CRC, enriched for a family history of colorectal tumors, in which they detected 12 individuals with the p.Leu424Val variant in *POLE* and one additional individual with the pSer478Asn in *POLD1*. Functional assessment supported the importance of these mutations in *POLE* and *POLD1*. Mutagenesis studies of Polδ and Pol3 in yeast showed that the mutation of the equivalent residue produces a mutator phenotype and loss of the proofreading activity of the protein[53,55,56]. Also, mice expressing proofreading-impaired Pole and Pold1 in a homozygous state developed spontaneous intestinal adenocarcinomas or a spectrum of cancers[57]. Thus, germline variants in *POLE* and *POLD1* predispose to individuals to either a multiple colorectal adenoma phenotype similar to that observed in MUTYH-associated polyposis or a HNPCC phenotype, in which carriers develop early-onset CRC. Although additional studies will be needed to evaluate these rare germline variants in *POLD1* and *POLE* and their associated phenotypes, the authors suggest that screening for these variants should be considered in patients with an unexplained personal or family history of multiple adenomas, early onset CRC or both. On the other hand, carriers are potential candidates for regular and frequent colonoscopic surveillance starting at an early age.

Two additional reports using exome sequencing have also been published very recently but their results are not as solid as those for the polymerase genes previously mentioned. A cohort of 50 sporadic CRC patients was sequenced including 18 early-onset cases with a relatively low coverage in the first study[58]. Variants were biased selected when found in a list of 1,138 genes likely to play a role in CRC. Further selection to include only those genes undergoing bialleic inactivation yielded *FANCM*, *LAMB4*, *PTCHD3*, *LAMC3* and *TREX2* as potential tumor suppressor candidates. In the second study, exome sequencing was completed for 40 familial cases from 16 families by selecting distant relatives to decrease the number of shared, non-predisposition variants[59]. Data was analyzed firstly by an agnostic search for CRC predisposition genes not taking into account a biased list of candidates, and secondly by selecting genes previously involved in CRC predisposition or within CRC linkage regions. Two missense variants in the *CENPE* and *KIF23* genes that complied with family segregation and belong to regions on chromosomes 1 and 15 formerly linked to CRC were considered the more plausible candidates for CRC predisposition but additional studies are needed to further elucidate their role.

## CONCLUSION

CRC is one of the most frequent neoplasms and an important cause of mortality in the developed world. CRC is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian cancer syndromes account for about 5% of the total burden of CRC, being Lynch syndrome and familial adenomatous polyposis the most common forms. Familial CRC type X is an example of CRC with unknown inherited cause. A clear positive family history of CRC is present (Amsterdam criteria for Lynch syndrome are fulfilled) although MMR is proficient. When considering CRC as a complex disease, low-penetrance genetic variants probably underlie part of the hereditary predisposition together with environmental interactions. So far, 30 susceptibility variants have been identified for CRC. New sequencing technologies made available recently including exome- and whole-genome sequencing have permitted to add a new approach to facilitate the identification of new genes responsible for human disease predisposition. Germline mutations in the *POLE* and *POLD1* genes are responsible for a new form of CRC genetic predisposition called polymerase proofreading-associated polyposis.

## ACKNOWLEDGMENTS

We are sincerely grateful to the Centre Nacional d'Anàlisi Genòmica and the Biobank of Hospital Clínic, Barcelona-IDIBAPS for technical help; The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

## REFERENCES

1 **Cohen AM**, Shank B, Friedman MA. Colorectal cancer. In: De Vita V, Hellman S, Rosenberg S (eds): Cancer: Principles and Practice of Oncology. Philadelphia: Lippincott, 1989: 895
2 **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**: 2893-2917 [PMID: 21351269 DOI: 10.1002/ijc.25516]
3 **Jemal A**, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69-90 [PMID: 21296855 DOI: 10.3322/caac.20107]

4 **Parker SL**, Tong T, Bolden S, Wingo PA. Cancer statistics, 1996. *CA Cancer J Clin* 1996; **46**: 5-27 [PMID: 8548526]

5 **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**: 525-532 [PMID: 2841597]

6 **Leggett B**, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* 2010; **138**: 2088-2100 [PMID: 20420948 DOI: 10.1053/j.gastro.2009.12.066]

7 **Gingras D**, Béliveau R. Colorectal cancer prevention through dietary and lifestyle modifications. *Cancer Microenviron* 2011; **4**: 133-139 [PMID: 21909875 DOI: 10.1007/s12307-010-0060-5]

8 **O'Callaghan T**. Introduction: The prevention agenda. *Nature* 2011; **471**: S2-S4 [PMID: 21430716 DOI: 10.1038/471S2a]

9 **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 [PMID: 10891514]

10 **Hemminki K**, Chen B. Familial risk for colorectal cancers are mainly due to heritable causes. *Cancer Epidemiol Biomarkers Prev* 2004; **13**: 1253-1256 [PMID: 15247139]

11 **Jasperson KW**, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010; **138**: 2044-2058 [PMID: 20420945 DOI: 10.1053/j.gastro.2010.01.054]

12 **Piñol V**, Castells A, Andreu M, Castellví-Bel S, Alenda C, Llor X, Xicola RM, Rodríguez-Moranta F, Payá A, Jover R, Bessa X. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005; **293**: 1986-1994 [PMID: 15855432]

13 **Hampel H**, Frankel W, Panescu J, Lockman J, Sotamaa K, Fix D, Comeras I, La Jeunesse J, Nakagawa H, Westman JA, Prior TW, Clendenning M, Penzone P, Lombardi J, Dunn P, Cohn DE, Copeland L, Eaton L, Fowler J, Lewandowski G, Vaccarello L, Bell J, Reid G, de la Chapelle A. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res* 2006; **66**: 7810-7817 [PMID: 16885385]

14 **Stoffel E**, Mukherjee B, Raymond VM, Tayob N, Kastrinos F, Sparr J, Wang F, Bandipalliam P, Syngal S, Gruber SB. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. *Gastroenterology* 2009; **137**: 1621-1627 [PMID: 19622357 DOI: 10.1053/j.gastro.2009.07.039]

15 **Suter CM**, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004; **36**: 497-501 [PMID: 15064764]

16 **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**: 112-117 [PMID: 19098912 DOI: 10.1038/ng.283]

17 **McGivern A**, Wynter CV, Whitehall VL, Kambara T, Spring KJ, Walsh MD, Barker MA, Arnold S, Simms LA, Leggett BA, Young J, Jass JR. Promoter hypermethylation frequency and BRAF mutations distinguish hereditary non-polyposis colon cancer from sporadic MSI-H colon cancer. *Fam Cancer* 2004; **3**: 101-107 [PMID: 15340260]

18 **Half E**, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis* 2009; **4**: 22 [PMID: 19822006 DOI: 10.1186/1750-1172-4-22]

19 **Bülow S**, Björk J, Christensen IJ, Fausa O, Järvinen H, Moesgaard F, Vasen HF. Duodenal adenomatosis in familial adenomatous polyposis. *Gut* 2004; **53**: 381-386 [PMID: 14960520]

20 **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**: 1979-1985 [PMID: 15855431]

21 **Mueller-Koch Y**, Vogelsang H, Kopp R, Lohse P, Keller G, Aust D, Muders M, Gross M, Daum J, Schiemann 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**: 1733-1740 [PMID: 15955785]

22 **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 USA* 2003; **100**: 12961-12965 [PMID: 14566058]

23 **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; **15**: 2903-2910 [PMID: 16923799]

24 **Tenesa A**, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 2009; **10**: 353-358 [PMID: 19434079 DOI: 10.1038/nrg2574]

25 **Fernandez-Rozadilla C**, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, Baiget M, López-Fernández LA, Brea-Fernández A, Abulí A, Bujanda L, Clofent J, Gonzalez D, Xicola R, Andreu M, Bessa X, Jover R, Llor X, Moreno V, Castells A, Carracedo Á, Castellvi-Bel S, Ruiz-Ponte C. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. *BMC Genomics* 2013; **14**: 55 [PMID: 23350875 DOI: 10.1186/1471-2164-14-55]

26 **Jia WH**, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, Kim DH, Ren Z, Cai Q, Long J, Shi J, Wen W, Yang G, Delahanty RJ, Ji BT, Pan ZZ, Matsuda F, Gao YT, Oh JH, Ahn YO, Park EJ, Li HL, Park JW, Jo J, Jeong JY, Hosono S, Casey G, Peters U, Shu XO, Zeng YX, Zheng W. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet* 2013; **45**: 191-196 [PMID: 23263487 DOI: 10.1038/ng.2505]

27 **Peters U**, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, Berndt SI, Bézieau S, Brenner H, Butterbach K, Caan BJ, Campbell PT, Carlson CS, Casey G, Chan AT, Chang-Claude J, Chanock SJ, Chen LS, Coetzee GA, Coetzee SG, Conti DV, Curtis KR, Duggan D, Edwards T, Fuchs CS, Gallinger S, Giovannucci EL, Gogarten SM, Gruber SB, Haile RW, Harrison TA, Hayes RB, Henderson BE, Hoffmeister M, Hopper JL, Hudson TJ, Hunter DJ, Jackson RD, Jee SH, Jenkins MA, Jia WH, Kolonel LN, Kooperberg C, Küry S, Lacroix AZ, Laurie CC, Laurie CA, Le Marchand L, Lemire M, Levine D, Lindor NM, Liu Y, Ma J, Makar KW, Matsuo K, Newcomb PA, Potter JD, Prentice RL, Qu C, Rohan T, Rosse SA, Schoen RE, Seminara D, Shrubsole M, Shu XO, Slattery ML, Taverna D, Thibodeau SN, Ulrich CM, White E, Xiang Y, Zanke BW, Zeng YX, Zhang B, Zheng W, Hsu L. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* 2013; **144**: 799-807.e24 [PMID: 23266556 DOI: 10.1053/j.gastro.2012.12.020]

28 **Houlston RS**, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS, Kerr D, Cazier JB, Niittymäki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IP, Farrington SM, Tene-

sa A, Prendergast JG, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PD, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Völzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008; **40**: 1426-1435 [PMID: 19011631 DOI: 10.1038/ng.262]

29 **Houlston RS**, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, Spain SL, Broderick P, Domingo E, Farrington S, Prendergast JG, Pittman AM, Theodoratou E, Smith CG, Olver B, Walther A, Barnetson RA, Churchman M, Jaeger EE, Penegar S, Barclay E, Martin L, Gorman M, Mager R, Johnstone E, Midgley R, Niittymäki I, Tuupanen S, Colley J, Idziaszczyk S, Thomas HJ, Lucassen AM, Evans DG, Maher ER, Maughan T, Dimas A, Dermitzakis E, Cazier JB, Aaltonen LA, Pharoah P, Kerr DJ, Carvajal-Carmona LG, Campbell H, Dunlop MG, Tomlinson IP. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 2010; **42**: 973-977 [PMID: 20972440 DOI: 10.1038/ng.670]

30 **Tomlinson IP**, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K, Palles C, Broderick P, Jaeger EE, Farrington S, Lewis A, Prendergast JG, Pittman AM, Theodoratou E, Olver B, Walker M, Penegar S, Barclay E, Whiffin N, Martin L, Ballereau S, Lloyd A, Gorman M, Lubbe S, Howie B, Marchini J, Ruiz-Ponte C, Fernandez-Rozadilla C, Castells A, Carracedo A, Castellvi-Bel S, Duggan D, Conti D, Cazier JB, Campbell H, Sieber O, Lipton L, Gibbs P, Martin NG, Montgomery GW, Young J, Baird PN, Gallinger S, Newcomb P, Hopper J, Jenkins MA, Aaltonen LA, Kerr DJ, Cheadle J, Pharoah P, Casey G, Houlston RS, Dunlop MG. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet* 2011; **7**: e1002105 [PMID: 21655089 DOI: 10.1371/journal.pgen.1002105]

31 **Dunlop MG**, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi LY, Domingo E, Smillie C, Henrion M, Frampton M, Martin L, Grimes G, Gorman M, Semple C, Ma YP, Barclay E, Prendergast J, Cazier JB, Olver B, Penegar S, Lubbe S, Chander I, Carvajal-Carmona LG, Ballereau S, Lloyd A, Vijayakrishnan J, Zgaga L, Rudan I, Theodoratou E, Starr JM, Deary I, Kirac I, Kovacević D, Aaltonen LA, Renkonen-Sinisalo L, Mecklin JP, Matsuda K, Nakamura Y, Okada Y, Gallinger S, Duggan DJ, Conti D, Newcomb P, Hopper J, Jenkins MA, Schumacher F, Casey G, Easton D, Shah M, Pharoah P, Lindblom A, Liu T, Smith CG, West H, Cheadle JP, Midgley R, Kerr DJ, Campbell H, Tomlinson IP, Houlston RS. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet* 2012; **44**: 770-776 [PMID: 22634755 DOI: 10.1038/ng.2293]

32 **Kinnersley B**, Migliorini G, Broderick P, Whiffin N, Dobbins SE, Casey G, Hopper J, Sieber O, Lipton L, Kerr DJ, Dunlop MG, Tomlinson IP, Houlston RS. The TERT variant rs2736100 is associated with colorectal cancer risk. *Br J Cancer* 2012; **107**: 1001-1008 [PMID: 22878375 DOI: 10.1038/bjc.2012.329]

33 **Sanger F**, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; **74**: 5463-5467 [PMID: 271968]

34 **Wheeler DA**, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008; **452**: 872-876 [PMID: 18421352 DOI: 10.1038/nature06884]

35 **Ulahannan D**, Kovac MB, Mulholland PJ, Cazier JB, Tomlinson I. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *Br J Cancer* 2013; **109**: 827-835 [PMID: 23887607 DOI: 10.1038/bjc.2013.416]

36 **Koboldt DC**, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Brief Bioinform* 2010; **11**: 484-498 [PMID: 20519329 DOI: 10.1093/bib/bbq016]

37 **Ng SB**, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30-35 [PMID: 19915526 DOI: 10.1038/ng.499]

38 **Desai AN**, Jere A. Next-generation sequencing: ready for the clinics? *Clin Genet* 2012; **81**: 503-510 [PMID: 22375550 DOI: 10.1111/j.1399-0004.2012.01865.x]

39 **DePristo MA**, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491-498 [PMID: 21478889 DOI: 10.1038/ng.806]

40 **Kimura M**. The Neutral Theory of Molecular Evolution. New York: Cambridge Press, 1983

41 **Lim J**, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, Barabási AL, Vidal M, Zoghbi HY. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 2006; **125**: 801-814 [PMID: 16713569]

42 **Richards FM**, McKee SA, Rajpar MH, Cole TR, Evans DG, Jankowski JA, McKeown C, Sanders DS, Maher ER. Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. *Hum Mol Genet* 1999; **8**: 607-610 [PMID: 10072428]

43 **Lammi L**, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I, Pirinen S, Nieminen P. Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* 2004; **74**: 1043-1050 [PMID: 15042511]

44 **Sweet K**, Willis J, Zhou XP, Gallione C, Sawada T, Alhopuro P, Khoo SK, Patocs A, Martin C, Bridgeman S, Heinz J, Pilarski R, Lehtonen R, Prior TW, Frebourg T, Teh BT, Marchuk DA, Aaltonen LA, Eng C. Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis. *JAMA* 2005; **294**: 2465-2473 [PMID: 16287957]

45 **Ngeow J**, Heald B, Rybicki LA, Orloff MS, Chen JL, Liu X, Yerian L, Willis J, Lehtonen HJ, Lehtonen R, Mester JL, Moline J, Burke CA, Church J, Aaltonen LA, Eng C. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterology* 2013; **144**: 1402-1409, 1409.e1-5 [PMID: 23399955 DOI: 10.1053/j.gastro.2013.02.001]

46 **Zogopoulos G**, Jorgensen C, Bacani J, Montpetit A, Lepage P, Ferretti V, Chad L, Selvarajah S, Zanke B, Hudson TJ, Pawson T, Gallinger S. Germline EPHB2 receptor variants in familial colorectal cancer. *PLoS One* 2008; **3**: e2885 [PMID: 18682749 DOI: 10.1371/journal.pone.0002885]

47 **Guda K**, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, Parmigiani 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**: 12921-12925 [PMID: 19617566 DOI: 10.1073/pnas.0901454106]

48 **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**: 1056-1058 [PMID: 22461326 DOI:

10.1002/humu.22088]

49 **Venkatachalam R**, Ligtenberg MJ, Hoogerbrugge N, Schackert HK, Görgens 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**: 2221-2224 [PMID: 21036128 DOI: 10.1053/j.gastro.2010.08.063]

50 **Lubbe SJ**, Pittman AM, Matijssen C, Twiss P, Olver B, Lloyd A, Qureshi M, Brown N, Nye E, Stamp G, Blagg J, Houlston RS. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. *Hum Mutat* 2011; **32**: E1928-E1938 [PMID: 20949628 DOI: 10.1002/humu.21376]

51 **Nieminen TT**, Abdel-Rahman WM, Ristimäki A, Lappalainen M, Lahermo P, Mecklin JP, Järvinen HJ, Peltomäki P. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 2011; **141**: e23-e26 [PMID: 21640116 DOI: 10.1053/j.gastro.2011.03.063]

52 **Orloff MS**, He X, Peterson C, Chen F, Chen JL, Mester JL, Eng C. Germline PIK3CA and AKT1 mutations in Cowden and Cowden-like syndromes. *Am J Hum Genet* 2013; **92**: 76-80 [PMID: 23246288 DOI: 10.1016/j.ajhg.2012.10.021].]

53 **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, Kearsey 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**: 136-144 [PMID: 23263490 DOI: 10.1038/ng.2503]

54 **Seshagiri S**. The burden of faulty proofreading in colon cancer. *Nat Genet* 2013; **45**: 121-122 [PMID: 23358219 DOI: 10.1038/ng.2540]

55 **Jin YH**, Ayyagari R, Resnick MA, Gordenin DA, Burgers PM. Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick. *J Biol Chem* 2003; **278**: 1626-1633 [PMID: 12424237]

56 **Murphy K**, Darmawan H, Schultz A, Fidalgo da Silva E, Reha-Krantz LJ. A method to select for mutator DNA polymerase deltas in Saccharomyces cerevisiae. *Genome* 2006; **49**: 403-410 [PMID: 16699561]

57 **Albertson TM**, Ogawa M, Bugni JM, Hays LE, Chen Y, Wang Y, Treuting PM, Heddle JA, Goldsby RE, Preston BD. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci USA* 2009; **106**: 17101-17104 [PMID: 19805137 DOI: 10.1073/pnas.0907147106]

58 **Smith CG**, Naven M, Harris R, Colley J, West H, Li N, Liu Y, Adams R, Maughan TS, Nichols L, Kaplan R, Wagner MJ, McLeod HL, Cheadle JP. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* 2013; **34**: 1026-1034 [PMID: 23585368 DOI: 10.1002/humu.22333]

59 **DeRycke MS**, Gunawardena SR, Middha S, Asmann YW, Schaid DJ, McDonnell SK, Riska SM, Eckloff BW, Cunningham JM, Fridley BL, Serie DJ, Bamlet WR, Cicek MS, Jenkins MA, Duggan DJ, Buchanan D, Clendenning M, Haile RW, Woods MO, Gallinger SN, Casey G, Potter JD, Newcomb PA, Le Marchand L, Lindor NM, Thibodeau SN, Goode EL. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev* 2013; **22**: 1239-1251 [PMID: 23637064 DOI: 10.1158/1055-9965.EPI-12-1226]

**P- Reviewers**: Bustamante-Balen M, Hamfjord J, Ventham NT
**S- Editor**: Qi Y **L- Editor**: A **E- Editor**: Liu XM

百世登
Baishideng®

Published by **Baishideng Publishing Group Co., Limited**
Flat C, 23/F., Lucky Plaza,
315-321 Lockhart Road, Wan Chai, Hong Kong, China
Fax: +852-65557188
Telephone: +852-31779906
E-mail: bpgoffice@wjgnet.com
http://www.wjgnet.com

ISSN 1007-9327
0 8>
9 771007 932045

OPEN ACCESS Freely available online

PLOS ONE

# The *MLH1* c.1852_1853delinsGC (p.K618A) Variant in Colorectal Cancer: Genetic Association Study in 18,723 Individuals

**Anna Abulí[1,2], Luis Bujanda[3], Jenifer Muñoz[1], Stephan Buch[4], Clemens Schafmayer[5], Maria Valeria Maiorana[6], Silvia Veneroni[7], Tom van Wezel[8], Tao Liu[9], Helga Westers[10], Clara Esteban-Jurado[1], Teresa Ocaña[1], Josep M. Piqué[1], Montserrat Andreu[2], Rodrigo Jover[11], Angel Carracedo[12,13], Rosa M. Xicola[14], Xavier Llor[14], Antoni Castells[1], The EPICOLON Consortium¶, Malcolm Dunlop[15], Robert Hofstra[10], Annika Lindblom[9], Juul Wijnen[16], Paolo Peterlongo[6], Jochen Hampe[4], Clara Ruiz-Ponte[12], Sergi Castellví-Bel[1]***

**1** Department of Gastroenterology, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Catalonia, Spain, **2** Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, Barcelona, Catalonia, Spain, **3** Gastroenterology Department, Hospital Donostia, Networked Biomedical Research Centre for Hepatic and Digestive Diseases (CIBEREHD), Basque Country University, San Sebastián, Spain, **4** Department of Medine I, University Hospital Dresden, Dresden, Germany, **5** Department of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Kiel, Germany, **6** IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy, **7** Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, **8** Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, **9** Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden, **10** Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands, **11** Department of Gastroenterology, Hospital General d'Alacant, Alicante, Spain, **12** Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, Santiago de Compostela, University of Santiago de Compostela, Galicia, Spain, **13** Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia, **14** Section of Digestive Diseases and Nutrition, University of Illinois at Chicago, Chicago, Illinois, United States of America, **15** Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh and MRC Human Genetics Unit, Edinburgh, United Kingdom, **16** Departments of Human Genetics and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

## Abstract

Colorectal cancer is one of the most frequent neoplasms and an important cause of mortality in the developed world. Mendelian syndromes account for about 5% of the total burden of CRC, being Lynch syndrome and familial adenomatous polyposis the most common forms. Lynch syndrome tumors develop mainly as a consequence of defective DNA mismatch repair associated with germline mutations in *MLH1*, *MSH2*, *MSH6* and *PMS2*. A significant proportion of variants identified by screening these genes correspond to missense or noncoding changes without a clear pathogenic consequence, and they are designated as ''variants of uncertain significance'', being the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene a clear example. The implication of this variant as a low-penetrance risk variant for CRC was assessed in the present study by performing a case-control study within a large cohort from the COGENT consortium-COST Action BM1206 including 18,723 individuals (8,055 colorectal cancer cases and 10,668 controls) and a case-only genotype-phenotype correlation with several clinical and pathological characteristics restricted to the Epicolon cohort. Our results showed no involvement of this variant as a low-penetrance variant for colorectal cancer genetic susceptibility and no association with any clinical and pathological characteristics including family history for this neoplasm or Lynch syndrome.

**Citation:** Abulí A, Bujanda L, Muñoz J, Buch S, Maiorana CSMV, et al. (2014) The *MLH1* c.1852_1853delinsGC (p.K618A) Variant in Colorectal Cancer: Genetic Association Study in 18,723 Individuals. PLoS ONE 9(4): e95022. doi:10.1371/journal.pone.0095022

**Editor:** Nathan A. Ellis, University of Illinois at Chicago, United States of America

**Received** December 16, 2013; **Accepted** March 21, 2014; **Published** April 17, 2014



**Funding:** This work was supported by COST Action BM1206. Groups 1, 4, 5, 6, 7, 8, 10 and 13 are active members of this consortium. Edinburgh: This work was supported by a Cancer Research UK Programme Grant (C348/A12076) and a Centre Grant from the CORE Charity. Epicolon: SCB is supported by a contract from the Fondo de Investigación Sanitaria (CP 03-0070). JM and CEJ are supported by contracts from the CIBERehd. CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III. This work was supported by grants from the Fondo de Investigación Sanitaria/FEDER (11/00219, 11/00681), Instituto de Salud Carlos III (Acción Transversal de Cáncer), Xunta de Galicia (07PXIB9101209PR), Ministerio de Ciencia e Innovación (SAF2010-19273), Asociación Española contra el Cáncer (Fundación Científica GCB13131592CAST y Junta de Barcelona), Fundació Olga Torres (SCB and CRP) and FP7 CHIBCHA Consortium (SCB and ACar). Kiel: This work was supported by the German Ministry for Education and Research through the German National Genome Research Network (NGFNplus) Colon Cancer Network (CCN). Leiden: This work was supported by the Dutch Cancer Society with grant KWF-UL-2005-3247. Stockholm: The Swedish sample and data resource were funded by the Swedish Cancer Society, the Swedish Scientific Research Council and the Stockholm Cancer Foundation. The funding agencies 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: sbel@clinic.ub.es

¶ Membership of the EPICOLON Consortium is provided in Appendix S1.

## Introduction

Colorectal cancer (CRC) is one of the most frequent neoplasms and an important cause of mortality in the developed world. This cancer is caused by both genetic and environmental factors although 35% of the variation in CRC susceptibility involves inherited genetic differences. Mendelian syndromes account for about 5% of the total burden of CRC, being Lynch syndrome and familial adenomatous polyposis the most common forms. Lynch syndrome tumors develop mainly as a consequence of defective DNA mismatch repair (MMR) associated with germline mutations in the *MLH1*, *MSH2*, *MSH6* and *PMS2* genes [1]. Once clinical criteria for this syndrome are complied, genetic screening of these genes is performed when a MMR defect is detected in the patient's tumor. When a pathogenic variant is detected, management of this disease can be significantly improved by identifying carriers that will benefit from specific screening, preventive, and therapeutic measures. Also, identifying non-carriers in additional family members permit to release these individuals from intensive surveillance. Noteworthy, a significant proportion of variants identified in the MMR genetic screening correspond to missense or noncoding changes without a clear pathogenic consequence, and they are designated as "variants of uncertain significance" (VUS). Therefore, differentiating pathogenic and neutral genetic variants is still a major challenge in clinical genetics [2].

The c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene corresponds to a clear example of VUS in Lynch syndrome. When consulting the Leiden Open Variation Database (LOVD v.2.0), there are 120 entries for this variant [3]. Available past studies reached contradictory conclusions about its pathogenicity reporting harmful *in silico* predictions [4], absence of splicing or mRNA alteration [5], presence in patients with a defective MMR tumor [6], co-occurrence with clearly pathogenic MMR mutations [7], apparent segregation with disease [8], and a majority of non-altered *in vitro* functional studies [9,10]. All previous data permitted to categorize it in LOVD as a class 1 variant (non-pathogenic/low clinical significance) [11]. Therefore, it should be considered as a neutral variant in terms of its implication with Lynch syndrome.

Recently, genome-wide association studies (GWAS) successfully identified so far 30 common, low-penetrance susceptibility variants in 25 risk loci for CRC [12–19]. Some genetic variants in hereditary CRC genes labeled as VUS could constitute low-penetrance risk alleles for CRC. Indeed, this hypothesis has been previously tested for some variants in those genes [20]. In agreement with this rationale, the main aim of the present study was to assess the implication of the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene as a low-penetrance risk variant for CRC by performing a case-control study within a large cohort from the COGENT consortium-COST Action BM1206, an international effort to facilitate the study of inherited genetic predisposition to CRC [21,22].

## Materials and Methods

### Study population

The current genetic association study totalized 8,055 CRC cases and 10,668 controls from 7 different cohorts (Edinburgh, Epicolon, Groningen, Kiel, Leiden, Milano, Stockholm) and recruitment details are summarized below. The study was approved by the institutional ethical committee of each participating hospital and written informed consent was obtained from all patients.

**Edinburgh cohort (1,553 CRC cases and 932 controls).** A population-based series of patients from throughout Scotland, who were diagnosed with colorectal cancer when they were less than 55 years of age, were recruited to the study between February 1999 and June 2004. During the same period, unaffected controls were ascertained from a population-based register (community health index) and were invited to participate.

**Epicolon (2,001 CRC cases and 1,647 controls).** Cases and controls were recruited through the EPICOLON Consortium that is based on a prospective, multicenter and population-based epidemiology survey of the incidence and features of CRC in the Spanish population [23]. Briefly, cases were selected as patients with *de novo* histologically confirmed diagnosis of colorectal adenocarcinoma. Exclusion criteria were hereditary CRC forms, such as Lynch syndrome and familial adenomatous polyposis (FAP) and a personal history of inflammatory bowel disease. Controls were from the Spanish National DNA bank and were confirmed not to have cancer or history of neoplasm and no family history of CRC. All cases and controls were of Caucasian ethnicity.

**Groningen (559 CRC cases and 501 controls).** Unselected CRC cases and hospital patient controls from the Netherlands included in the SCOPE project.

**Kiel (1,768 CRC cases and 2.030 controls).** Cases and controls from population-based biobank projects including POPGEN (Population Genetic Cohort) from Schleswig-Holstein, north Germany, and SHIP (Survey of Health in Pommerania) from east and north-east Germany.

**Leiden (505 CRC cases and 836 controls).** Cases and controls were recruited as previously described [24]. Briefly, most of the cases were recruited through the clinical genetics department. All cases were diagnosed with CRC and had early onset and/or positive family history for CRC. Known dominant polyposis syndromes, HNPCC/Lynch syndrome or bi-allelic MutYH mutation carriers were excluded. A single proband from each family was included in this study. Controls were healthy blood donors from the southwest region of the Netherlands. All cases and controls were of Caucasian ethnicity.

**Milano (619 CRC cases and 2,526 controls).** Briefly, the cases were consecutive individuals affected with CRC who underwent surgery at the Fondazione IRCCS Istituto Nazionale Tumori in Milan (INT). The controls were blood donors recruited through the Immunohematology and Transfusion Medicine Service of INT the Associazione Volontari Italiani Sangue Comunale in Milan. All cases and controls were of Caucasian ethnicity.

**Stockholm (1,729 CRC cases and 1,487 controls).** Unselected cases ascertained through 12 hospitals serving the Stockholm-Gotland and Uppsala-Örebro health-care regions in Sweden and blood donor controls.

### Genotyping

DNA was obtained from peripheral blood by standard extraction procedures. Allelic discrimination to genotype the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene was performed by using a custom assay with the TaqMan allelic discrimination system (Life Technologies, Foster City, USA). As quality control, DNA from a known carrier of this variant was used as positive control, as well as duplicates and negative controls for amplification. Data could be available upon request. An example of allelic discrimination for this variant is shown in **Figure 1**.

### Statistical analysis

To test the association between the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene and CRC risk, odds ratios

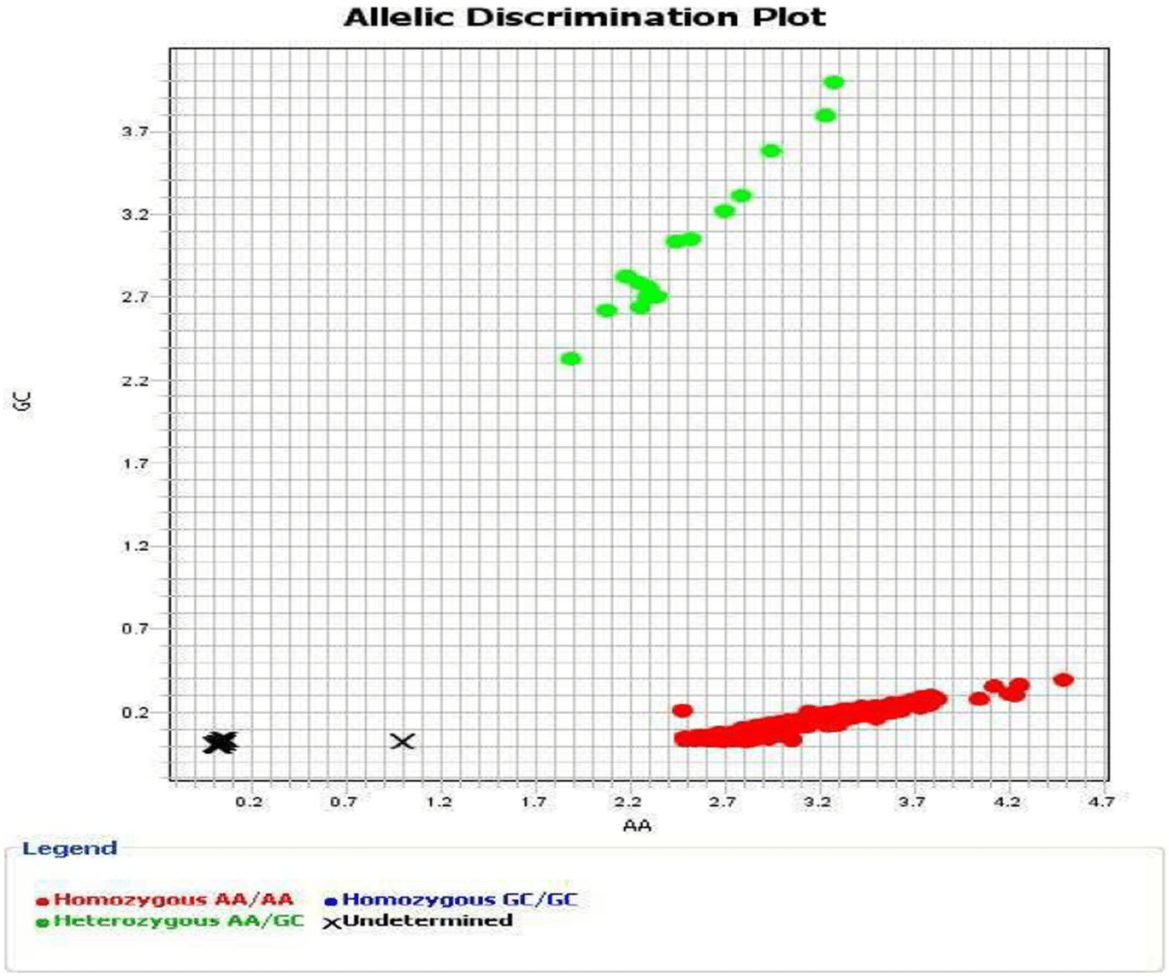



**Figure 1. Allelic discrimination for c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene by using the TaqMan system.** Red dots correspond to non-carriers (AA/AA genotype) and green dots to heterozygous carriers (AA/GC).
doi:10.1371/journal.pone.0095022.g001

(OR) and 95%CI were calculated for each genotype by using PLINK v1.07 [25], separately in each cohort and globally. No deviation of the genotype frequency in controls from those expected under Hardy-Weinberg equilibrium (HWE) was detected by $\chi^2$ test (1 df) (P-value = 0.6294) [26].

In order to explore if personal and/or familial characteristics were associated with the presence of the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene, univariate analysis was performed restricted to the CRC cases from the Epicolon cohort due to data availability in this cohort. The selected clinical variables to be evaluated were gender, age (dichotomized by 50 y.o.), location of CRC, previous neoplasm, previous/synchronous adenoma, CRC familiy history (any relative with CRC), Lynch syndrome family history (any relative affected), microsatellite instability (MSI) and TNM tumor stage. Categorical variables were compared by the $\chi^2$ test (1 df), applying the Yates' correction when needed. All *P*-values were two-sided, and a value less than 0.05 was considered statistically significant. Calculations were performed using the SPSS software version 18.0 (SPSS Inc, Chicago, Ill).

## Results and Discussion

Genotyping for the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene was successful in 8,055 CRC cases and 10,668 controls from 7 independent cohorts. Percentage of carriers varied between 0.4–2.6% in CRC cases and 0.5–3.1% for controls in the different cohorts, being 1.4% and 1.5% in the entire cohort for CRC cases and controls, respectively. Genotypic association results are shown in **Table 1** for each cohort and globally. No association of this variant with CRC risk was detected neither in a specific cohort nor globally.

In order to further explore the putative implication of this *MLH1* variant with CRC risk, we performed a case-only genotype-phenotype correlation restricted to the Epicolon cohort (2,001 CRC cases) with several clinical and pathological characteristics. Results are shown in **Table 2**. Again, none of the analyzed variables showed a distinct association with the presence of the c.1852_1853delinsGC (p.K618A) variant. Results for CRC family history and Lynch syndrome family history were statistically significant but the presence of any of these variables was linked

**Table 1.** Genotypic association results for the *MLH1* c.1852_1853delinsGC (p.K618A) variant in 18,723 individuals from 7 cohorts.

| Cohort | Controls | % | Cases | % | OR | lower | upper | *P*-value |
|---|---|---|---|---|---|---|---|---|
| *Edinburgh* | | | | | | | | |
| AA/AA | 1,539 | 99.1 | 916 | 98.3 | 1.000 | | | 0.087 |
| AA/GC | 14 | 0.9 | 16 | 1.7 | 1.904 | 0.934 | 3.884 | |
| Total | 1,553 | | 932 | | | | | |
| *Epicolon* | | | | | | | | |
| AA/AA | 1,596 | 96.9 | 1,949 | 97.4 | 1.000 | | | 0.368 |
| AA/GC | 51 | 3.1 | 52 | 2.6 | 0.839 | 0.574 | 1.228 | |
| Total | 1,647 | | 2,001 | | | | | |
| *Groningen* | | | | | | | | |
| AA/AA | 555 | 99.3 | 497 | 99.2 | 1.000 | | | 1.000 |
| AA/GC | 4 | 0.7 | 4 | 0.8 | 1.116 | 0.281 | 4.438 | |
| Total | 559 | | 501 | | | | | |
| *Kiel* | | | | | | | | |
| AA/AA | 2,003 | 98.7 | 1,752 | 99.1 | 1.000 | | | 0.282 |
| AA/GC | 27 | 1.3 | 16 | 0.9 | 0.680 | 0.368 | 1.259 | |
| Total | 2,030 | | 1,768 | | | | | |
| *Leiden* | | | | | | | | |
| AA/AA | 832 | 99.5 | 503 | 99.6 | 1.000 | | | 1.000 |
| AA/GC | 4 | 0.5 | 2 | 0.4 | 0.828 | 0.152 | 4.503 | |
| Total | 836 | | 505 | | | | | |
| *Milano* | | | | | | | | |
| AA/AA | 2,526 | 98.8 | 614 | 99.2 | 1.000 | | | 0.525 |
| AA/GC | 30 | 1.2 | 5 | 0.8 | 0.688 | 0.268 | 1.767 | |
| Total | 2,556 | | 619 | | | | | |
| *Stockholm* | | | | | | | | |
| AA/AA | 1,466 | 98.6 | 1,700 | 98.3 | 1.000 | | | 0.571 |
| AA/GC | 21 | 1.4 | 29 | 1.7 | 1.188 | 0.680 | 2.074 | |
| Total | 1,487 | | 1,729 | | | | | |
| ***GLOBAL*** | | | | | | | | |
| AA/AA | 10,517 | 98.6 | 7,931 | 98.5 | 1.000 | | | 0.501 |
| AA/GC | 151 | 1.4 | 124 | 1.5 | 1.088 | 0.859 | 1.377 | |
| **Total** | **10,668** | | **8,055** | | | | | |

OR, odds ratio.

doi:10.1371/journal.pone.0095022.t001

with the wild-type genotype (AA/AA). The rest of variables showed a similar distribution between carriers and non-carriers.

Obviously, genetic variants causing a missense mutation have a less clear pathogenic interpretation than those causing a premature termination of the protein. The c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene is a prominent example of a VUS that has been controversial for many years in the context of Lynch syndrome genetic diagnosis. However, recent functional studies have permitted to characterize more thoroughly its real effect of the MLH1 protein and it can be concluded that its effect is neutral or with very subtle effect [5,8–11].

Regarding its putative implication in CRC risk as a rare low-penetrance variant, previous studies were sparse and included a small number of CRC cases and controls [27,28]. Consequently, it was justified to perform a case-control association study in a large cohort in order to reach more solid conclusions. Our results showed no involvement of this variant in CRC risk as a low-penetrance variant in the *MLH1* gene.

Regarding its putative implication in familial CRC, this variant was also seen to be over-represented in families with suspected Lynch syndrome in a previous study [29]. Our results will be not in agreement with this previous observation since the K618A variant was not linked in the Epicolon cohort to the presence of CRC family history and Lynch syndrome family history. Therefore, our study is adding to the existing literature by showing that this variant is not linked to familial CRC.

Finally, we can conclude from our results and previous evidence that the c.1852_1853delinsGC (p.K618A) variant in the *MLH1* gene should be regarded from now on as a polymorphism without functional effect on the MLH1 protein, no role in genetic predisposition to Lynch syndrome, as well as no apparent effect as a low-penetrance variant for CRC genetic susceptibility.

**Table 2.** Genotype-phenotype correlation of the *MLH1* c.1852_1853delinsGC (p.K618A) variant with clinical and pathological characteristics in colorectal cancer cases from the Epicolon cohort.

| | CRC≤50 | % | CRC>50 | % | OR | lower | upper | *P*-value |
|---|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | | |
| AA/AA | 97 | 5 | 1,841 | 95 | 1.000 | | | 1.000 |
| AA/GC | 2 | 3.8 | 50 | 96.2 | 1.317 | 0.316 | 5.493 | |
| Total | 99 | | 1,891 | | | | | |
| | Female | % | Male | % | OR | lower | upper | *P*-value |
| **Gender** | | | | | | | | |
| AA/AA | 766 | 39.5 | 1,172 | 60.5 | 1.000 | | | 0.388 |
| AA/GC | 17 | 32.7 | 35 | 67.3 | 1.346 | 0.749 | 2.419 | |
| Total | 783 | | 1,207 | | | | | |
| | Colon | % | Rectum | % | OR | lower | upper | *P*-value |
| **CRC location** | | | | | | | | |
| AA/AA | 1,267 | 65.9 | 656 | 34.1 | 1.000 | | | 0.882 |
| AA/GC | 33 | 64.7 | 18 | 35.3 | 1.053 | 0.589 | 1.885 | |
| Total | 1,300 | | 674 | | | | | |
| | No | % | Yes | % | OR | lower | upper | *P*-value |
| **Previous neoplasm** | | | | | | | | |
| AA/AA | 1,290 | 73.8 | 458 | 26.2 | 1.000 | | | 0.624 |
| AA/GC | 39 | 78 | 11 | 22 | 0.794 | 0.403 | 1.564 | |
| Total | 1,329 | | 469 | | | | | |
| | No | % | Yes | % | OR | lower | upper | *P*-value |
| **Prev/sync adenoma** | | | | | | | | |
| AA/AA | 1,268 | 71.2 | 513 | 28.8 | 1.000 | | | 0.112 |
| AA/GC | 41 | 82 | 9 | 18 | 0.543 | 0.262 | 1.124 | |
| Total | 1,309 | | 522 | | | | | |
| | No | % | Yes | % | OR | lower | upper | *P*-value |
| **CRC FH** | | | | | | | | |
| AA/AA | 1,652 | 85.2 | 286 | 14.8 | 1.000 | | | 0.026 |
| AA/GC | 50 | 96.2 | 2 | 3.8 | 0.231 | 0.056 | 0.955 | |
| Total | 1,702 | | 288 | | | | | |
| | No | % | Yes | % | OR | lower | upper | *P*-value |
| **Lynch FH** | | | | | | | | |
| AA/AA | 1,401 | 81.5 | 317 | 18.5 | 1.000 | | | 0.048 |
| AA/GC | 42 | 93.3 | 3 | 6.7 | 0.316 | 0.097 | 1.025 | |
| Total | 1,443 | | 320 | | | | | |
| | No | % | Yes | % | OR | lower | upper | *P*-value |
| **MSI** | | | | | | | | |
| AA/AA | 1,308 | 94 | 84 | 6 | 1.000 | | | 0.731 |
| AA/GC | 37 | 92.5 | 3 | 7.5 | 1.263 | 0.381 | 4.180 | |
| Total | 1,345 | | 87 | | | | | |
| | I–II | % | III–IV | % | OR | lower | upper | *P*-value |
| **TNM** | | | | | | | | |
| AA/AA | 909 | 53.7 | 783 | 46.3 | 1.000 | | | 1.000 |
| AA/GC | 26 | 53.1 | 23 | 46.9 | 1.027 | 0.581 | 1.814 | |
| Total | 935 | | 806 | | | | | |

CRC, colorectal cancer; OR, odds ratio; Prev/Sync, Previous/Synchronous; FH, family history; MSI, microsatellite instability; TNM, tumor-node-metastasis.
doi:10.1371/journal.pone.0095022.t002

## Supporting Information

**Appendix S1 Members of the EPICOLON Consortium (Gastrointestinal Oncology Group of the Spanish Gastroenterological Association).**
(DOCX)

## Acknowledgments

**Epicolon**: We are sincerely grateful to all patients participating in this study who were recruited in 25 (EPICOLON 1) and 14 (EPICOLON 2) Spanish hospitals as part of the EPICOLON project. We are also indebted to the USC and UPF nodes of Spanish National Genotyping Center (CEGEN-ISCIII), the Genomics Unit of the Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) and the Biobanks of Hospital Clínic – IDIBAPS and Hospital Donostia for technical help. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

**Leiden**: Shanti Jagmohan-Changur and Melanie Schrumpf are acknowledged for technical support.

**Milano**: We thank all individuals who agreed to participate in the study. We also thank the personnel of Tissue Bank of Fondazione IRCCS Istituto dei Tumori for sample collection and all pathologists for their contribution and collaboration.

**Stockholm**: We acknowledge the contribution to recruitment and data collection of the Swedish Low-Risk Colorectal Cancer Study Group.

## Author Contributions

Conceived and designed the experiments: TW MD RH AL JW PP JH CRP SCB. Performed the experiments: AA JM SB MVM TW TL HW CEJ MD RH AL JW PP JH CRP SCB. Analyzed the data: AA JM SB MVM TW TL HW CEJ MD RH AL JW PP JH CRP SCB. Contributed reagents/materials/analysis tools: AA LB JM SB CS MVM SV TW TL HW CEJ TO JMP MA RJ AC RMX XL AC EC MD RH AL JW PP JH CRP SCB. Wrote the paper: AA TW TL HW MD RH AL JW PP JH CRP SCB.

## References

1. Jasperson KW, Tuohy TM, Neklason DW, Burt RW (2010) Hereditary and familial colon cancer. Gastroenterology 138: 2044–58.
2. Rasmussen LJ, Heinen CD, Royer-Pokora B, Drost M, Tavtigian S, et al. (2012) Pathological assessment of mismatch repair gene variants in Lynch syndrome: past, present, and future. Hum Mutat 33: 1617–25.
3. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, et al. (2011) LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 2011 32: 557–63.
4. Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, et al. (2008) Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). Hum Mutat 29: 852–60.
5. Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, et al. (2008) A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat 29: 1412–24.
6. Caldes T, Godino J, de la Hoya M, Garcia Carbonero I, Perez Segura P, et al. (2002) Prevalence of germline mutations of MLH1 and MSH2 in hereditary nonpolyposis colorectal cancer families from Spain. Int J Cancer 98: 774–9.
7. Liu T, Tannergård P, Hackman P, Rubio C, Kressner U, et al. (1999) Missense mutations in hMLH1 associated with colorectal cancer. Hum Genet 105: 437–41.
8. Pastrello C, Pin E, Marroni F, Bedin C, Fornasarig M, et al. (2011) Integrated analysis of unclassified variants in mismatch repair genes. Genet Med 13: 115–24.
9. Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, et al. (2007) Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. Cancer Res 67: 4595–604.
10. Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G (2013) Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis. Clin Cancer Res 19: 2432–41.
11. InSiGHT (2014). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 46: 107–15.
12. Tenesa A, Dunlop MG (2009) New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet 10: 353–8.
13. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, et al. (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet 42: 973–977.
14. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, et al. (2011) Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS Genet 7: e1002105.
15. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, et al. (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 44:770–776.
16. Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins SE, et al. (2012) The TERT variant rs2736100 is associated with colorectal cancer risk. Br J Cancer 107: 1001–8.
17. Fernández-Rozadilla C, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, et al. (2013) A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. BMC Genomics 26: 14:55.
18. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, et al. (2013) Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet 45: 191–196.
19. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, et al. (2013) Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 144: 799–807.
20. Picelli S, Lorenzo Bermejo J, Chang-Claude J, Hoffmeister M, Fernández-Rozadilla C, et al. (2013) Meta-Analysis of Mismatch Repair Polymorphisms within the Cogent Consortium for Colorectal Cancer Susceptibility. PLoS One 8: e72091.
21. Tomlinson IP, Dunlop M, Campbell H, Zanke B, Gallinger S, et al. (2010) COGENT (COlorectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. Br J Cancer 102: 447–54.
22. Houlston RS; members of COGENT (2012) COGENT (COlorectal cancer GENeTics) revisited. Mutagenesis 27: 143–51.
23. Piñol V, Castells A, Andreu M, Castellví-Bel S, Alenda C, et al. (2005) Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 293: 1986–94.
24. Middeldorp A, Jagmohan-Changur S, van Eijk R, Tops C, Devilee P, et al. (2009) Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev 18: 3062–7.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
26. Wigginton, JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005, 76(5):887–93.
27. Christensen LL, Madsen BE, Wikman FP, Wiuf C, Koed K, et al. (2008) The association between genetic variants in hMLH1 and hMSH2 and the development of sporadic colorectal cancer in the Danish population. BMC Med Genet 11:9: 52.
28. Castillejo A, Guarinos C, Martinez-Canto A, Barbera VM, Egoavil C, et al. (2011) Evidence for classification of c.1852_1853AA>GC in MLH1 as a neutral variant for Lynch syndrome. BMC Med Genet 12: 12.
29. Medeiros F, Lindor NM, Couch FJ, Highsmith WE Jr (2012) The germline MLH1 K618A variant and susceptibility to Lynch syndrome-associated tumors. J Mol Diagn 14: 264–73.

PLOS | ONE

RESEARCH ARTICLE

# Genetic Variants Associated with Colorectal Adenoma Susceptibility

**Anna Abulí[1], Antoni Castells[2], Luis Bujanda[3], Juan José Lozano[4], Xavier Bessa[1], Cristina Hernández[5], Cristina Álvarez-Urturi[1], Maria Pellisé[2], Clara Esteban-Jurado[2], Elizabeth Hijona[3], Andrea Burón[5], Francesc Macià[5], Jaume Grau[6], Rafael Guayta[7], Sergi Castellví-Bel[2☯], Montserrat Andreu[1☯*], PROCOLON research group[¶]**

**1** Department of Gastroenterology, Hospital del Mar, IMIM (Hospital del Mar Medical Research Institute), Pompeu Fabra University, Passeig Marítim 25–29, 08003, Barcelona, Catalonia, Spain, **2** Department of Gastroenterology, Hospital Clínic, CIBERehd, IDIBAPS, University of Barcelona, Villarroel 170, 08036, Barcelona, Catalonia, Spain, **3** Department of Gastroenterology, Hospital Donostia/Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Universidad del País Vasco (UPV/EHU), Doctor Begiristain Kalea, 20014, Donostia/Gipuzkoa, Spain, **4** Plataforma de Bioinformática, CIBERehd, Villarroel 170, 08036, Barcelona, Catalonia, Spain, **5** Department of Epidemiology and Evaluation, Hospital del Mar, IMIM (Hospital del Mar Medical Research Institute), Passeig Marítim 25–29, 08003, Barcelona, Catalonia, Spain, **6** Unitat d'Avaluació, Suport i Preventiva, Hospital Clínic, Roselló 138, 08036, Barcelona, Catalonia, Spain, **7** Planning and Research Unit, Consell de Collegis Farmacèutics de Catalunya, Girona 64, 08009, Barcelona, Catalonia, Spain

☯ These authors contributed equally to this work.
¶ All members of the PROCOLON group are listed in the Appendix.
* mandreu@parcdesalutmar.cat





**Citation:** Abulí A, Castells A, Bujanda L, Lozano JJ, Bessa X, Hernández C, et al. (2016) Genetic Variants Associated with Colorectal Adenoma Susceptibility. PLoS ONE 11(4): e0153084. doi:10.1371/journal.pone.0153084

**Editor:** Hiromu Suzuki, Sapporo Medical University, JAPAN

**Received:** December 8, 2015

**Accepted:** March 23, 2016

**Published:** April 14, 2016



**Data Availability Statement:** Due to ethical restrictions related to protecting patient privacy, data cannot be made publicly available. Data are available on request from Montserrat Andreu (mandreu@parcdesalutmar.cat).

**Funding:** This work was supported by grants from: Instituto de Salud Carlos III-FEDER (RD09/0076/00036), the Xarxa de Bancs de tumors sponsored by Pla Director d'Oncologia de Catalunya (XBTC), Fondo de Investigación Sanitaria/FEDER (PI10/00918, PI11/00219, PI14/00173, PI14/00441), Ministerio de Economía y Competitividad (SAF2010-19273), Asociación Española contra el Cáncer

## Abstract

### Background

Common low-penetrance genetic variants have been consistently associated with colorectal cancer risk.

### Aim

To determine if these genetic variants are associated also with adenoma susceptibility and may improve selection of patients with increased risk for advanced adenomas and/or multiplicity (≥ 3 adenomas).

### Methods

We selected 1,326 patients with increased risk for advanced adenomas and/or multiplicity and 1,252 controls with normal colonoscopy from population-based colorectal cancer screening programs. We conducted a case-control association study analyzing 30 colorectal cancer susceptibility variants in order to investigate the contribution of these variants to the development of subsequent advanced neoplasia and/or multiplicity.

### Results

We found that 14 of the analyzed genetic variants showed a statistically significant association with advanced adenomas and/or multiplicity: the probability of developing these lesions

(Fundación Científica GCB13131592CAST), and COST Action BM1206 (SCB). SCB is supported by a contract from the Fondo de Investigación Sanitaria (CP 03-0070) and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255, 2014SGR135). CIBERehd is funded by the Instituto de Salud Carlos III.

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

increased with the number of risk alleles reaching a 2.3-fold risk increment in individuals with $\geq$ 17 risk alleles.

## Conclusions

Nearly half of the genetic variants associated with colorectal cancer risk are also related to advanced adenoma and/or multiplicity predisposition. Assessing the number of risk alleles in individuals within colorectal cancer screening programs may help to identify better a subgroup with increased risk for advanced neoplasia and/or multiplicity in the general population.

## Introduction

Colorectal cancer (CRC) is one of the most common malignancies in the Western world and represents an important health problem worldwide.[1] Most CRCs arise from adenomatous polyps but only some adenomas acquire additional genetic alterations at the somatic level that make them grow, develop advanced histological features, and progress to cancer.[2] Patients presenting adenomas at the baseline colonoscopy with villous histology or high grade dysplasia or $\geq$10 mm in size, or $\geq$3 adenomas are considered at an increased risk for a subsequent advanced neoplasia, either cancer or advanced adenoma.[3] Transition from detectable adenoma to CRCs is estimated to take at least 10 years in most cases, providing an excellent window for early detection of the disease. This is the rationale for population-based CRC screening programs, which are aimed to identify malignant lesions at an early stage [4] or, even better, to detect and remove adenomatous polyps before CRC develops, thus reducing CRC incidence and mortality.[5] Population-based CRC screening programs are designed for average risk population. Among the accepted screening strategies in average-risk population, annual or biennial fecal occult blood testing is the most widely used.[6–8] Indeed, different trials have proved the effectiveness of fecal occult blood test, demonstrating a CRC mortality reduction of 15–33%.[9] Currently, the target population is defined only by age. However, it is well recognized that the outcome of patients with apparently similar risk at baseline is quite heterogeneous, thus emphasizing the need of more accurate predictors of CRC development.

Advances in genomic technologies have made it possible to genotype and evaluate many single-nucleotide polymorphisms (SNPs) throughout the human genome to identify novel disease susceptibility genes. Common, low-penetrance genetic variation for CRC have been identified by genome-wide association studies (GWAS) during the past years, allowing to point out so far 30 genetic variants in 25 risk loci at 1p33, 1q41, 1q25.3, 2q32.3, 3q26.2, 5p15.33, 5q31.1, 6p21, 8q24.21, 8p12, 8q23.3, 10p14, 11q13.4, 11q23.1, 12p13.32, 12q13.3, 12q24.21, 14q22.2, 15q13.3, 16q22.1, 18q21.1, 19q13, 20p12.3, 20q13.33, Xp22.2.[10–23] However, most of these studies mainly focused on CRC risk and, therefore, they only partially assessed the contribution of these variants to colorectal adenoma (CRA) susceptibility. This fact is especially true in patients with advanced adenomas or multiplicity, the main precursors of CRC.[24–27] Genetic predisposition variants shared by CRA and CRC could lead to additional knowledge on cancer initiation and progression and could elucidate why only a subset of CRA patients ends up developing CRC. Indeed, the identification of genetic factors involved in the early events of the adenoma-carcinoma sequence may offer the greatest potential benefit for CRC prevention.

Accordingly, the primary objective in our study was to know whether some of these common, low-penetrance CRC genetic variants solidly identified for CRC risk were also associated with CRA development. As secondary objective, we wanted to assess the cumulative impact of

these genetic variants on the probability of advanced adenoma and/or multiplicity and to explore a risk prediction model based on age, gender and genetic susceptibility variants, aiding to modulate risk stratification in population-based screening programs.

## Materials and Methods

### Ethics statement

The study was approved by the institutional ethic committee of each participating hospital (Hospital del Mar, IMIM (Hospital del Mar Medical Research Institute). Hospital Clínic and Hospital Donostia/Instituto Biodonostia), and a written informed consent was obtained from all patients.

### Study population

The current case-control study included 1,351 patients with advanced adenomas and/or 3 adenomas or more diagnosed at baseline colonoscopy and 1,266 control individuals with normal colonoscopy from the Spanish population. Individuals were recruited prospectively through the first round of the population-based CRC screening program at 3 hospitals from Spain, between September 2011 and November 2012. Asymptomatic men and women aged 50 through 69 years with an average risk of developing CRC were eligible to undergo colonoscopy after a positive FIT. Criteria for exclusion in the population-based CRC screening program included a personal history of CRC, adenoma, or inflammatory bowel disease, a family history of hereditary or familial colorectal cancer (i.e. $\geq$2 first-degree relatives with CRC or 1 first-degree relative diagnosed before the age of 60), a severe coexisting illness, or a previous colectomy. Environmental data were not considered in our study.

All colonoscopies were performed by expert endoscopists (those who had performed >400 colonoscopies per year). The quality of the bowel preparation in each colonoscopy was adequate and it was evaluated by the Boston Bowel Preparation Scale (each colon segment (right, transverse, left) had to reach a minimum score of 2 per segment (maximum 3) with a total score $\geq$6). Adenomas were classified by size (<10 mm or $\geq$10 mm), histology (tubular, tubulovillous or villous), degree of dysplasia (low or high-grade dysplasia) and number. After total colonoscopy, patients with advanced adenoma (adenomas with villous histology or high grade dysplasia or $\geq$10mm in size) and/or $\geq$3 adenomas were selected as cases. Controls were polyp-free individuals after complete colonoscopy. Patients with low-risk adenomas ($\leq$ 2 tubular adenomas, <10mm and low-grade dysplasia) or serrated polyps [28] were excluded from the study.

### SNP genotyping and quality control

DNA was obtained from frozen peripheral blood for all samples by standard extraction procedures in each participating hospital. SNPs were genotyped by using the TaqMan® OpenArray™ Genotyping System (Applied Biosystems Inc.). Genotyping of 1,351 cases and 1,266 controls for 30 SNPs including rs6983267, rs4939827, rs3802842, rs4779584, rs16892766, rs10795668, rs4444235, rs9929218, rs10411210, rs961253, rs6691170, rs10936599, rs11169552, rs4925386, rs1957636, rs4813802, rs2736100, rs1321311, rs3824999, rs5934683, rs12080929, rs11987193, rs10774214, rs647161, rs2423279, rs11903757, rs10911251, rs3217810, rs3217901 and rs5933 was performed at the Genomics Core Facility from the Pompeu Fabra University in Barcelona, Spain. SNP selection included genetic variants identified as linked to CRC risk by GWAS mainly conducted in European populations, and showing a genome-wide statistical significance (P-value$<5\times 10^{-8}$). Results in a prior Spanish GWAS[16] supported the CRC association of most of these genetic variants either by statistical significance or by showing odds

ratios in the same direction as those previously described. Also, all included SNPs can be considered independent genetic association signals including those located in the same genes ($R^2<0.1$). Genotyping call rates for the 30 SNPs varied from 87.9% to 99.7%. In order to test for genotyping quality, 10 duplicates were included, as well as 5 additional DNA samples with previously known results for the tested SNPs by using different platforms and available through previous studies.[13,16] Genotype concordance was 100% for all 15 samples. Quality control of the data was assessed using Genotyping Data Filter (http://bioinformatics.cesga.es/gdf/nav_input.php, GDF) and PLINK 1.07.[29] Samples with genotyping success rate below 90% were removed from subsequent analyses. Deviation of the genotype frequencies in controls from those expected under Hardy-Weinberg equilibrium (HWE) was assessed by $X^2$ test (1df).[30] Each SNP was in HWE (*P*-value >0.01) in controls (data not shown), thereby excluding the possibility of genotyping artifacts and any hidden population stratification. After quality control, the final cohort comprised 2,578 samples (1,326 cases and 1,252 controls) that remained to be analyzed. The overall genotyping success rate in the remaining individuals was >96%. Investigators responsible for genotyping were blinded to the clinical data.

## Statistical analysis

Genotypic and allelic association tests and logistic regression were performed using PLINK v1.07.[29] Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated for each genetic variant. Although there was already substantial prior evidence of an association between all SNPs examined and overall CRC risk, *P*-values were corrected for multiple comparisons by using the Benjamini Hochberg correction and false discovery rate (FDR)-corrected *P*-values (*Q*-values) <0.1 were considered to be significant.[31] Study power was estimated using CATS software [32] and power calculation was done under the assumption of an additive model with $\alpha = 0.05$.[33] The total number of significantly associated risk alleles was calculated for all samples and a two-sided t test was applied to compare the mean number of risk alleles between cases and controls. ORs with 95% CI and trend test for increasing risk alleles were estimated by counting two for homozygotes and one for heterozygotes in each genetic variant. The number of risk alleles was coded as 0, 1 or 2 for each SNP assuming a log-additive genetic effect. The method to compute the risk probabilities was based in a weighted way by multiplying the number of risk alleles at each locus (0, 1, or 2) for the corresponding $\beta$ coefficient from additive multivariate logistic regression model and then summing the products. Age was included in the equation as a numeric variable and gender as a factor. Significant variables obtained in the multivariate analysis were used to calculate the risk of having advanced adenoma and/or multiplicity for each patient according to the following equation: in which $\beta_0$ was the constant of the model, $\beta_1$ to $\beta_p$ were the regression coefficients of the independent variables, and $x_{1i}$ to $x_{pi}$ were the values of the variable for a particular patient *i*:

$$\text{Risk}_i = \frac{e^{\beta_0+\beta_1 x_{1i}+\cdots+\beta_p x_{pi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\cdots+\beta_p x_{pi}}}$$

PredictABEL R package was used to develop the equation risk and to predict the risk probabilities of the subjects.[34]

## Results

A total of 1,326 individuals with advanced adenomas and/or multiplicity and 1,252 control individuals were successfully genotyped for 30 SNPs previously known to confer genetic susceptibility to CRC. Table 1 summarizes their demographic and clinical characteristics. The mean age at recruitment of cases and controls was 60.35 (SD, 5.38) and 59.65 (SD, 5.64) years, respectively.

**Table 1. Summary of the demographic and clinical characteristics of individuals included in the study.**

| Characteristics | Cases (N = 1,326) | Controls (N = 1,252) | *P*-value |
|---|---|---|---|
| Mean age, y (SD) | 60.35 (5.38) | 59.65 (5.64) | 0.001 |
| Male, n (%) | 905 (68.3) | 509 (40.7) | 0.0001 |
| Female, n (%) | 421 (31.7) | 743 (59.3) | |
| Mean adenomas, n | 4.2 | - | |
| ≥ 3 adenomas, n (%)* | 788 (59.4) | - | |
| Adenoma ≥ 1 mm in size, n (%)* | 735 (55.4) | - | |
| Adenoma with villous histology, n (%)* | 831 (62.7) | - | |
| Adenoma with high grade dysplasia, n (%)* | 198 (15) | - | |

*One patient may have more than one characteristic. N. number; y, years; SD, standard deviation.

doi:10.1371/journal.pone.0153084.t001

## Association test for individual SNPs

Logistic regression adjusted for age and gender was used to detect risk alleles significantly enriched in patients with adenomas compared to controls and results are shown in Table 2. Although age and gender are associated with adenoma cases, they were not associated with SNP genotype and did not affect the statistical significance of any of the reported associations, as shown when genotype and allelic association were calculated not adjusting for these covariates *(S1 Table)*.

We found statistically significant associations with advanced adenomas and/or multiplicity for 14 out of the 30 SNPs analyzed (rs6983267, rs4939827, rs3802842, rs16892766, rs10795668, rs4444235, rs10411210, rs6691170, rs4925386, rs3824999, rs647161, rs2423279, rs3217810, rs59336) and these genetic associations were in the same direction as previously reported for CRC susceptibility (Table 2). Therefore, we selected these 14 SNPs that were associated with adenomas for subsequent analyses. The remaining SNPs, although not significant, showed ORs in the same directions as those previously described in the literature except for rs11169552, rs2736100 and rs11903757.

## Polygenic risk model

We also evaluated the presence of multiple risk alleles in the adenoma cohort when compared to controls. Distribution of risk by allele number for the 14 SNPs associated with adenoma is displayed for cases and controls in Fig 1. The distribution of risk alleles followed a normal distribution in both cases and controls with a shift towards a higher number of risk alleles in affected individuals consistent with a cumulative impact of risk alleles on adenoma predisposition. The mean number of risk alleles in controls subjects was 12.84 compared to 13.88 in cases (difference: -1.03 alleles, 95%CI 1.24–0.83) and there was a highly significant difference in the mean number of risk alleles between cases and controls (2-sided t-test $p<0.001$). In order to assess the risk of developing advanced adenoma and/or multiplicity associated with multiple alleles, we calculated ORs and 95% CI for cases when carrying an increasing number of risk alleles. Thirteen risk alleles were considered as reference since it was the median number in controls. Individuals were grouped for subjects carrying ≤9 risk alleles and ≥17 alleles because of very small number of subjects at these extremes. We observed that the risk of adenoma increased along with number of risk alleles for the 14 loci ($P_{trend} = 4.9x10^{-4}$, based on 1,073 cases and 1,021 controls). Individuals with ≥17 risk alleles had nearly a 2.5-fold increase in adenoma risk compared with those with 13 risk alleles.

## Assessment of risk prediction

In order to explore the possible clinical utility of the 14 genetic variants associated with advanced adenomas and/or multiplicity for individual risk prediction, we constructed a model

**Table 2. Case-control association results obtained by logistic regression analyses adjusted for age and gender.** Association results for cases (1,326) vs polyp-free controls (1,266). Results are based on the reported allele from previous CRC GWAS (reference number is shown). Statistically significant associations are denoted in bold (*P*-value<0.05 and multiple-comparison corrected *Q*-value<0.1).

| SNP | Locus | Gene | Reported allele | GWAS Ref. | MAF cases | MAF controls | GWAS OR (95% CI) | OR (95% CI) | *P*-value | *Q*-value |
|---|---|---|---|---|---|---|---|---|---|---|
| rs12080929 | 1p33 | *SLC5A9* | C | 16 | 0.291 | 0.280 | 0.86 (0.78–0.95) | 0.99 (0.87–1.12) | 0.842 | 0.902 |
| rs10911251 | 1q25.3 | *LAMC1* | A | 18 | 0.418 | 0.414 | 1.10 (1.06–1.14) | 1.04 (0.92–1.18) | 0.520 | 0.657 |
| **rs6691170** | 1q41 | *DUSP10* | T | 12 | 0.376 | 0.329 | 1.06 (1.03–1.09) | 1.22 (1.09–1.37) | **$9.1x10^{-4}$** | **0.009** |
| rs11903757 | 2q32.3 | *NABP1* | C | 18 | 0.147 | 0.153 | 1.15 (1.09–1.22) | 0.99 (0.85–1.16) | 0.896 | 0.926 |
| rs10936599 | 3q26.2 | *TERC* | T | 20 | 0.217 | 0.220 | 0.93 (0.91–0.96) | 0.94 (0.89–1.17) | 0.735 | 0.816 |
| rs2736100 | 5p15.33 | *TERT* | T | 15 | 0.468 | 0.487 | 1.07 (1.04–1.10) | 0.91 (0.81–1.02) | 0.111 | 0.221 |
| **rs647161** | 5q31.1 | *PITX1* | A | 17 | 0.345 | 0.359 | 1.07 (1.02–1.11) | 1.13 (1.01–1.27) | **0.030** | **0.069** |
| rs1321311 | 6p21 | *CDKN1A* | A | 14 | 0.254 | 0.239 | 1.10 (1.07–1.10) | 1.10 (0.96–1.26) | 0.170 | 0.300 |
| rs11987193 | 8p12 | *DUSP4* | T | 16 | 0.267 | 0.283 | 0.78 (0.70–0.87) | 0.90 (0.80–1.03) | 0.118 | 0.221 |
| **rs16892766** | 8q23.3 | *EIF3H* | C | 19 | 0.067 | 0.059 | 1.25 (1.19–1.32) | 1.29 (1.02–1.61) | **0.041** | **0.087** |
| **rs6983267** | 8q24.21 | *MYC* | G | 21 | 0.436 | 0.462 | 1.21 (1.15–1.27)[6] | 1.19 (1.05–1.32) | **$4.7x10^{-3}$** | **0.020** |
| **rs10795668** | 10p14 | - | A | 19 | 0.291 | 0.324 | 0.91 (0.86–0.96) | 0.83 (0.73–0.94) | **$3.3x10^{-3}$** | **0.019** |
| **rs3824999** | 11q13.4 | *POLD3* | G | 14 | 0.479 | 0.511 | 1.08 (1.05–1.10) | 1.15 (1.02–1.28) | **0.018** | **0.054** |
| **rs3802842** | 11q23.1 | *POU2AF1* | C | 10 | 0.297 | 0.268 | 1.21 (1.15–1.27) | 1.21 (1.07–1.37) | **$2.8x10^{-3}$** | **0.019** |
| rs10774214 | 12p13.32 | *CCND2* | T | 17 | 0.353 | 0.351 | 1.04 (1.00–1.09) | 0.95 (0.85–1.07) | 0.438 | 0.597 |
| **rs3217810** | 12p13.32 | *CCND2* | T | 18 | 0.103 | 0.082 | 1.19 (1.11–1.28) | 1.28 (1.05–1.56) | **0.012** | **0.045** |
| rs3217901 | 12p13.32 | *CCND2* | G | 18 | 0.368 | 0.351 | 1.10 (1.06–1.15) | 1.08 (0.96–1.22) | 0.199 | 0.331 |
| rs11169552 | 12q13.3 | *DIP2B* | T | 12 | 0.221 | 0.219 | 0.92 (0.90–0.95) | 1.02 (0.91–1.18) | 0.548 | 0.657 |
| **rs59336** | 12q24.21 | *TBX3* | T | 18 | 0.462 | 0.499 | 1.10 (1.06–1.14) | 1.14 (1.01–1.28) | **0.027** | **0.067** |
| **rs4444235** | 14q22.2 | *BMP4* | C | 11 | 0.449 | 0.485 | 1.12 (1.07–1.18) | 1.21 (1.08–1.36) | **$9.9x10^{-4}$** | **0.009** |
| rs1957636 | 14q22.2 | *BMP4* | A | 13 | 0.421 | 0.417 | 1.08 (1.06–1.11) | 1.00 (0.89–1.13) | 0.958 | 0.958 |
| rs4779584 | 15q13.3 | *GREM1* | T | 22 | 0.181 | 0.181 | 1.19 (1.12–1.26) | 1.05 (0.90–1.21) | 0.538 | 0.657 |
| rs9929218 | 16q22.1 | *CDH1* | A | 11 | 0.270 | 0.283 | 0.88 (0.83–0.92) | 0.95 (0.83–1.08) | 0.406 | 0.580 |
| **rs4939827** | 18q21.1 | *SMAD7* | T | 23 | 0.423 | 0.451 | 1.18 (1.12–1.23) | 1.15 (1.03–1.29) | **0.015** | **0.050** |
| **rs10411210** | 19q13 | *RHPN2* | T | 11 | 0.117 | 0.148 | 0.79 (0.72–0.86) | 0.74 (0.62–0.88) | **$5.8x10^{-4}$** | **0.009** |
| rs961253 | 20p12.3 | *BMP2* | A | 11 | 0.338 | 0.323 | 1.13 (1.08–1.19) | 1.08 (0.96–1.21) | 0.218 | 0.344 |
| rs4813802 | 20p12.3 | *BMP2* | G | 13 | 0.320 | 0.312 | 1.09 (1.06–1.12) | 1.03 (0.91–1.17) | 0.598 | 0.690 |
| **rs2423279** | 20p12.3 | *HAQ1* | C | 17 | 0.322 | 0.298 | 1.07 (1.03–1.12) | 1.15 (1.02–1.30) | **0.022** | **0.060** |
| **rs4925386** | 20q13.33 | *LAMA5* | T | 12 | 0.255 | 0.297 | 0.93 (0.91–0.95) | 0.83 (0.73–0.94) | **$4.6x10^{-3}$** | **0.020** |
| rs5934683 | Xp22.2 | *SHROOM2* | T | 14 | 0.402 | 0.392 | 1.07 (1.04–1.10) | 1.07 (0.93–1.22) | 0.335 | 0.502 |

GWAS Ref. = GWAS reference; MAF: minor allele frequency; OR, odds ratio; 95% CI, 95% confidence interval.

doi:10.1371/journal.pone.0153084.t002

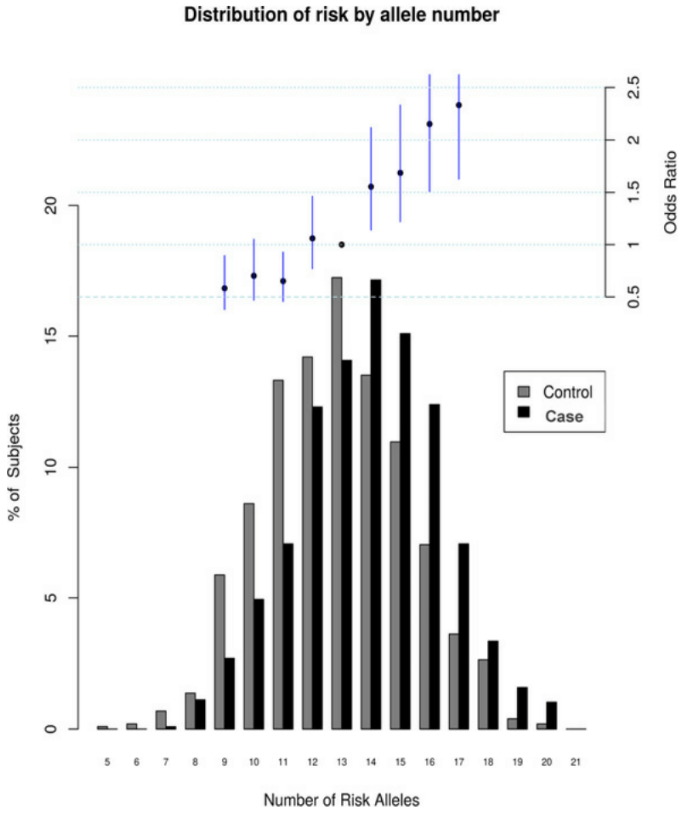
combining these genetic variants with gender and age. In this model, chance of advanced adenomas and/or multiplicity development was calculated for each subject with no missing data (1,073 cases and 1,021 controls) according to the following equation:

$$\text{Risk Score} = \frac{e^{\alpha}}{1 - e^{\alpha}},$$

where

$$\alpha = -3.9779 + 0.0154 \times \text{age} + 1.1493 \times \text{gender} + 0.2438 \times \text{rs10411210} + 0.1875 \times \text{rs10795668} + 0.3424 \times \text{rs16892766} + 0.3538 \times \text{rs3217810} + 0.1746 \times \text{rs3802842} + 0.1434 \times \text{rs3824999} + 0.2420 \times \text{rs4444235} + 0.1424 \times \text{rs4939827} + 0.1564 \times \text{rs59336} + 0.1821 \times \text{rs647161} + 0.2094 \times \text{rs6691170} + 0.1650 \times \text{rs6983267} + 0.1620 \times \text{rs4925386} + 0.1080 \times \text{rs2423279}.$$

The distribution of risk probabilities in patients with advanced adenomas and/or multiplicity and controls is shown in Fig 2. A tendency towards a higher risk score was noticeable in




| Nº of risk alleles | Cases | Controls | OR (95% CI)[a] | P-value |
|---|---|---|---|---|
| ≤9 | 42 | 84 | 0.58 (0.37-0.89) | 0.018 |
| 10 | 53 | 88 | 0.70 (0.47-1.05) | 0.106 |
| 11 | 76 | 136 | 0.65 (0.46-0.93) | 0.022 |
| 12 | 132 | 145 | 1.06 (0.77-1.46) | 0.778 |
| **13** | **151** | **176** | **1 (ref)** | |
| 14 | 184 | 138 | 1.55 (1.14-2.12) | $6.6x10^{-3}$ |
| 15 | 162 | 112 | 1.67 (1.21-2.33) | $2.1x10^{-3}$ |
| 16 | 133 | 72 | 2.15 (1.50-3.09) | $2.7x10^{-5}$ |
| **≥17** | **140** | **70** | **2.33 (1.63-3.34)** | **$3.7x10^{-6}$** |
| Total | 1073 | 1021 | $P_{trend} = 4.9x10^{-4}$ | |

**Fig 1. Cumulative impact of the 14 selected variants on adenoma risk.** Distribution of risk alleles for cases (black bars) and controls (grey bars). Upper panel and table: Plot of the ORs for cases with increasing number of risk alleles. ORs are relative to the median number of risk alleles in controls (13 risk alleles as reference group). Vertical bars correspond to 95% CI. Statistical significance is shown in the table for the different groups of multiple risk alleles.

doi:10.1371/journal.pone.0153084.g001

affected individuals. When comparing the upper and lower quantiles of the risk score, it was much more likely to find advanced adenomas and/or multiplicity cases with a higher risk (OR = 7.35, 95% CI 5.59–9.66, *P*-value = $2x10^{-16}$). Also, the median of risk score for advanced adenomas and/or multiplicity cases was 0.60 (95% CI 0.59–0.61) and 0.43 (95% CI 0.42–0.45) for controls. In general, risk score was significantly higher in the advanced adenomas and/or multiplicity group compared to controls (OR = 1.09, 95% CI 1.07–1.13, *P*-value = $7.93x10^{-13}$).

## Discussion

We found strong evidence that nearly half of CRC genetic variants were also involved in adenoma presentation. Additionally, we found that the risk of developing advanced adenomas and/or multiplicity increased along with the number of risk alleles, with an almost 2.5-fold increased risk in carriers of ≥17 risk alleles.

This study provides additional insight into the role of these genetic variants within the adenoma-carcinoma sequence and the association of these CRC risk alleles with advanced adenoma and/or multiplicity development. The cohort used was recruited as part of average-risk CRC screening programs and, therefore, cases and controls had merely age and gender as risk factors. Controls had normal colonoscopy, thus dismissing the presence of any colorectal neoplasia. A previous study [27] stated that the number of risk alleles was not a good variable for

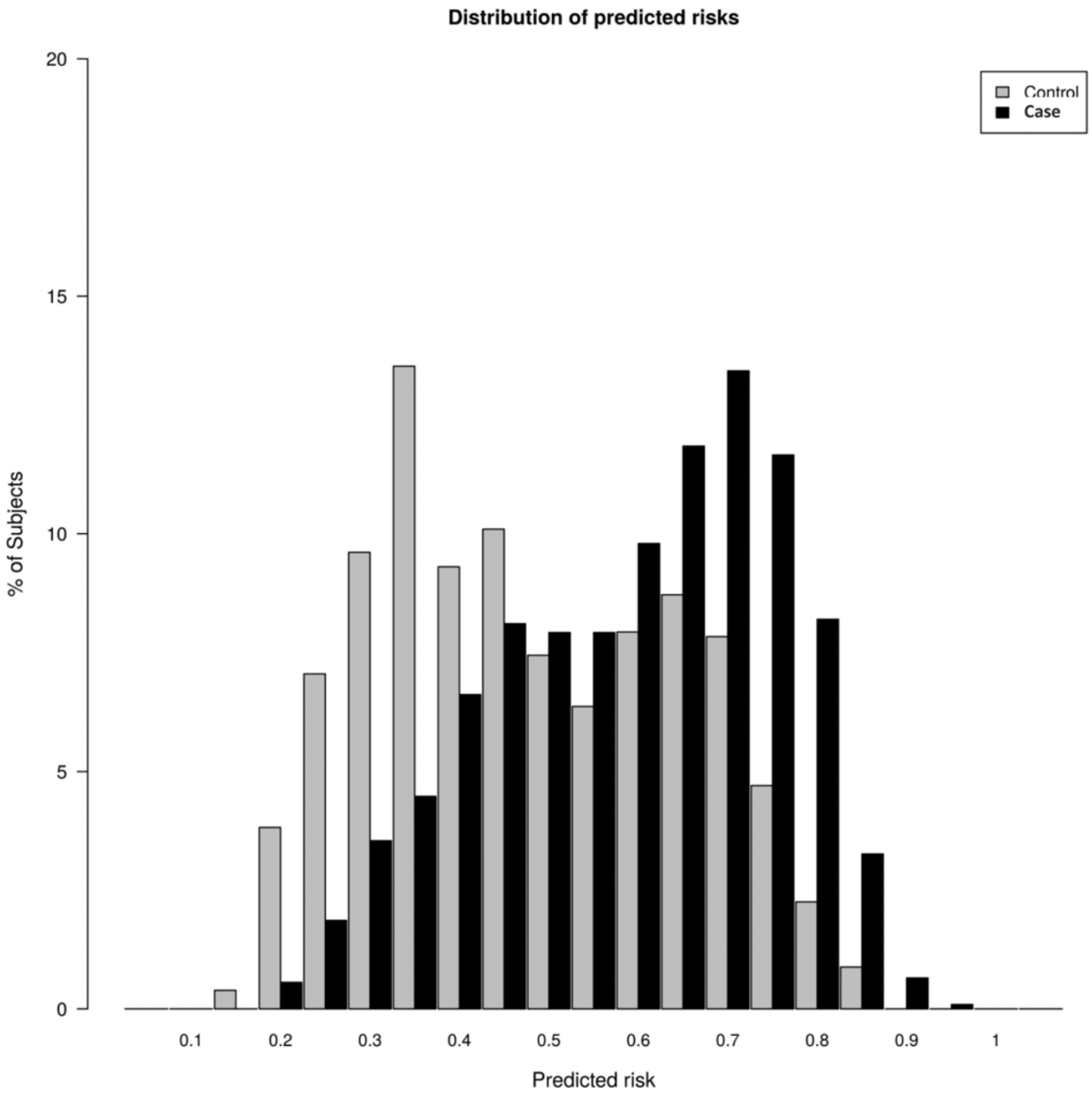



**Fig 2. Distributions of predicted risks in cases and controls.** The median of risk score was 0.60 (95% CI 0.59–0.61) for cases and 0.43 (95% CI 0.42–0.45) for controls. Risk score was significantly higher in advanced adenomas and/or multiplicity.

doi:10.1371/journal.pone.0153084.g002

differentiating between cases and control when considering their results. Importantly, we need to highlight that our control population is rather different. Population controls (some of them with adenomas) were tested by Cheng et al, whereas individuals with normal colonoscopy (no adenomas) were used in our study. We believe that this important difference is permitting us to obtain better results and be able to detect associations with the adenoma phenotype for almost half of the variants tested.

It could be postulated that CRC genetic variants may increase the risk of premalignant CRC precursors such as adenomas. However, CRC predisposition alleles can act either early in the adenoma-carcinoma sequence or later in the carcinogenic step, or even through adenoma-independent pathways. CRC genetic variants that only affect the progression into carcinogenic stage should not show an association for adenoma risk. However, our study showed that indeed practically half of the previously identified CRC risk variants were associated with an increased risk of adenoma and, therefore such variants seem to act through adenoma-carcinoma sequence. Our study also detected new variants associated with the development of adenomas not previously reported (rs16892766, rs10411210, rs6691170, rs4925386, rs3824999, rs647161, rs2423279). Additionally, we provided further evidence of the contribution of some of these variants in adenoma development (rs6983267, rs4939827, rs3802842, rs10795668, rs4444235, rs3217810, rs59336), already reported by previous studies.[24–26] One of the most significant associations was for rs6983267 (8q24.21).[21] This finding is in agreement with previous studies that already suggested the role of this variant in adenoma risk. Interestingly, Berndt et al. observed a stronger association for multiple adenomas than for single adenoma.[21,35] In line with these results, we previously reported in an independent study an interactive effect between rs6983267 (8q24.21) and rs9929218 (16q22.2) associated with a personal history of CRAs.[36] In addition, more recent studies reported association between rs6983267 and adenoma multiplicity,[24,25,27] supporting again the hypothesis that rs6983267 may have an effect on adenoma initiation or early CRC progression. Indeed, a recent study reported that the rs6983267 risk genotype (GG) affects the binding site for the Wnt regulated transcription factor TCF4 and, thereby, the transcription of *MYC* is upregulated.[37]

Besides, we also found statistically significant associations for rs3217810 (12p13.32) and rs59336 (12q24.21). Both SNPs are among the more recently reported variants as a result of a meta-analysis of several CRC GWAS.[24] This meta-analysis found stronger associations for adenoma compared to CRC for these 2 variants suggesting that some genes are related with early stages of CRC development while others may be more involved in the progression from adenoma to cancer. Additionally, our study found significant association for rs647161 (5q31.1) and rs2423279 (20p12.3), identified through a GWAS conducted in an East Asian population. [17] Although they also observed weaker associations in a case-control series of European ancestry, our study adds some more evidence of the implication of these variants in the European population, specifically with advanced adenoma lesions.

Another result to highlight is the highly significant difference between cases and controls regarding the mean number of risk alleles at the 14 selected adenoma susceptibility loci ($p<0.001$). Our results demonstrate the cumulative impact of multiple risk alleles on cases, especially in those individuals with at least 17 risk alleles. It could be suggested that a proportion of the general population with substantial increased risk of advanced adenomas or adenoma multiplicity, as determined by these genetic variants, could benefit from more intensive screening measures.

It is worth mentioning that our study had also a number of limitations. Although it was well powered to identify common variants (MAF $>0.3$) with OR $> 1.2$, the power to identify those loci with lower MAF or smaller genotypic risk was limited. Thus, the absence of association for the remaining variants with an expected relative risk about 1.1 may be explained by lack of power to detect association in our study. Indeed, most of these CRC variants with lower expected effect showed results in the same direction as previously reported for CRC susceptibility and therefore, we cannot exclude the possibility that some of them are also involved in advanced adenoma and/or multiplicity risk. Otherwise, the lack of association with adenoma risk in our study for these variants could also suggest that they have an effect on the later stages of colorectal tumorigenesis. It is also worth commenting that since our study only focused in

known genetic variants linked to CRC risk by previous solid GWAS studies and our hypothesis was to check if they were also implicated in an intermediate CRC phenotype, a replication of our findings in an independent cohort was not pursued. Finally, there is evidence that environmental factors such as smoking or body mass index are factors that modulate CRC risk but in this study they were not considered. However, cases and controls were age matched (±5 years) and all of them were of European ancestry from Spain and, by doing so, the influence of environmental differences between individuals was minimized at some extent. Also, it seems several of the genetic variants associated so far with CRC and adenoma risk are located close to genes involved in the TGF-beta pathway and BMP signaling.[38] These biological pathways are important in the adenoma-carcinoma sequence and, therefore, it could be hypothesized that their alteration by the functional effect of these genetic variants may be one of the mechanisms involved in adenoma predisposition.

In summary, our study provides evidence that nearly half of the CRC genetic risk variants are also associated with adenoma lesions. The presence of multiple risk alleles may allow identifying a subgroup of the population with a sufficient increased risk of advanced adenomas or adenoma multiplicity to be assigned to more intensive CRC prevention measures.

## Supporting Information

**S1 Table. Case-control association results obtained by logistic regression analyses with no adjusted for age and gender.** Association results for cases (1,326) vs polyp-free controls (1,266).
(DOCX)

## Acknowledgments

We are sincerely grateful to Pau Erola, the Genomics Core Facility from Pompeu Fabra University and the Biobanks of Hospital Clínic-IDIBAPS, Hospital del Mar and Hospital Donostia for technical help. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

PROCOLON is the research group of the Barcelona's Colorectal Cancer Screening Program, and currently includes the following members:

Cristina Álvarez, Montserrat Andreu, Josep M. Augé, Francesc Balaguer, Mercè Barau, Xavier Bessa, Felipe Bory, Andrea Burón, Antoni Castells, Xavier Castells, Mercè Comas, Rosa Costa, Ricard Courtier, Míriam Cuatrecasas, Josep M. Dedeu, Maria Estrada, Imma Garrell, Jaume Grau, Rafael Guayta, Cristina Hernández, Mar Iglesias, María López-Cerón, Francesc Macià, Leticia Moreira, M. Francisca Murciano, Gemma Navarro, Teresa Ocaña, Maria Pellisé, Mercè Pintanell, Mercè Piracés, Sandra Polbach, Àngels Pozo, Anna Reig, Cristina Rodríguez, Maria Sala, Agustín Seoane, Anna Serradesanferm, Judith Sivilla, Antoni Trilla.

## Author Contributions

Conceived and designed the experiments: AA AC JJL SCB MA. Performed the experiments: AA CH CEJ EH LB XB SCB. Analyzed the data: AA AC JJL CH SCB MA. Contributed reagents/materials/analysis tools: AA AC MA SCB LB XB CAU MP AB FM JG RG. Wrote the paper: AA MA JJL AC SCB. Selection of samples stored in the Biobank of Hospital del Mar and Hospital Donosti: AA EH.

## References

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013; 49:1374–403. doi: 10.1016/j.ejca.2012.12.027 PMID: 23485231

2. Neugut AI, Jacobson JS, Rella VA. Prevalence and incidence of colorectal adenomas and cancer in asymptomatic persons. Gastrointest Endosc Clin N Am. 1997; 7:387–99. PMID: 9177141
3. Winawer SJ, Zauber AG, Fletcher RH, Stillman JS, O'brien MJ, Levin B, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. Gastroenterology. 2006; 130:1872–85. PMID: 16697750
4. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, et al. 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. Gastroenterology. 2008; 134:1570–95. doi: 10.1053/j.gastro.2008.02.002 PMID: 18384785
5. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med. 2012; 366:687–96. doi: 10.1056/NEJMoa1100370 PMID: 22356322
6. Whitlock EP, Lin JS, Liles E, Beil TL, Fu R. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2008; 149:638–58. PMID: 18838718
7. van Rossum LG, van Rijn AF, Laheij RJ, van Oijen MG, Fockens P, van Krieken HH, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology. 2008; 135:82–90. doi: 10.1053/j.gastro.2008.03.040 PMID: 18482589
8. Guittet L, Bouvier V, Mariotte N, Vallee JP, Arsène D, Boutreux S, et al. Comparison of a guaiac based and an immunochemical faecal occult blood test in screening for colorectal cancer in a general average risk population. Gut.2007; 56:210–4. PMID: 16891354
9. Hardcastle JD, Thomas WM, Chamberlain J, Pye G, Sheffield J, James PD, et al. Randomised, controlled trial of faecal occult blood screening for colorectal cancer. Results for first 107,349 subjects. Lancet. 1989; 1:1160–1164. PMID: 2566735
10. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008; 40(5):631–7. doi: 10.1038/ng.133 PMID: 18372901
11. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008; 40:1426–35. doi: 10.1038/ng.262 PMID: 19011631
12. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet. 2010; 42:973–7. doi: 10.1038/ng.670 PMID: 20972440
13. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS Genet. 2011; 7:e1002105. doi: 10.1371/journal.pgen.1002105 PMID: 21655089
14. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet. 2012; 44:770–6. doi: 10.1038/ng.2293 PMID: 22634755
15. Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins SE, Casey G, et al. Colon Cancer Family Registry The TERT variant rs2736100 is associated with colorectal cancer risk. Br J Cancer. 2012; 107:1001–8. doi: 10.1038/bjc.2012.329 PMID: 22878375
16. Fernandez-Rozadilla C, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, et al. Colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. BMC Genomics. 2013; 14:55. doi: 10.1186/1471-2164-14-55 PMID: 23350875
17. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet. 2013; 45:191–6. doi: 10.1038/ng.2505 PMID: 23263487
18. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013; 144:799–807. doi: 10.1053/j.gastro.2012.12.020 PMID: 23266556
19. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, et al A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet. 2008; 40(5):623–30. doi: 10.1038/ng.111 PMID: 18372905

20. Jones AM, Beggs AD, Carvajal-Carmona L, Farrington S, Tenesa A, Walker M, et al, TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. Gut. 2012; 61(2):248–54. doi: 10.1136/gut.2011.239772 PMID: 21708826

21. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet. 2007; 39:984–8. PMID: 17618284

22. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet. 2008; 40(1):26–8. PMID: 18084292

23. Broderick P1, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007; 39(11):1315–7. PMID: 17934461

24. Zhang B, Shrubsole MJ, Li G, Cai Q, Edwards T, Smalley WE, et al. Association of genetic variants for colorectal cancer differs by subtypes of polyps in the colorectum. Carcinogenesis. 2012; 33:2417–23. doi: 10.1093/carcin/bgs308 PMID: 23027627

25. Carvajal-Carmona LG, Zauber AG, Jones AM, Howarth K, Wang J, Cheng T. Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. Gastroenterology. 2013; 144:53–55. doi: 10.1053/j.gastro.2012.09.016 PMID: 22999960

26. Burnett-Hartman AN, Newcomb PA, Hutter CM, Peters U, Passarelli MN, Schwartz MR, et al. Variation in the association between colorectal cancer susceptibility loci and colorectal polyps by polyp type. Am J Epidemiol. 2014; 180:223–32. doi: 10.1093/aje/kwu114 PMID: 24875374

27. Cheng TH, Gorman M, Martin L, Barclay E, Casey G. Common colorectal cancer risk alleles contribute to the multiple colorectal adenoma phenotype, but do not influence colonic polyposis in FAP. Eur J Hum Genet. 2015; 23:260–3. doi: 10.1038/ejhg.2014.74 PMID: 24801760

28. Rex DK, Ahnen DJ, Baron JA, Batts KP, Burke CA, Burt RW, et al. Serrated lesions of the colorectum: review and recommendations from an expert panel. Am J Gastroenterol. 2012; 107:1315–29. doi: 10.1038/ajg.2012.161 PMID: 22710576

29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. PMID: 17701901

30. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76:887–93. PMID: 15789306

31. Benjamini Y and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B (Methodological). 1995; 57(1): 289–300.

32. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–13. PMID: 16415888

33. Fletcher O1, Houlston RS. Architecture of inherited susceptibility to common cancer. Nat Rev Cancer. 2010; 10(5):353–61. doi: 10.1038/nrc2840 PMID: 20414203

34. Kundu S, Aulchenko YS, van Duijn CM: PredictABEL: an R package for the assessment of risk prediction models. Eur J Epidemiol 2011; 26(4):261–4. doi: 10.1007/s10654-011-9567-4 PMID: 21431839

35. Berndt SI, Potter JD, Hazra A, Yeager M, Thomas G, Makar KW, et al. Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. Hum Mol Genet. 2008; 17:2665–72. doi: 10.1093/hmg/ddn166 PMID: 18535017

36. Abulí A, Bessa X, González JR, Ruiz-Ponte C, Cáceres A, Muñoz J, et al. Susceptibility genetic variants associated with colorectal cancer risk correlate with cancer phenotype. Gastroenterology. 2010; 139:788–96. doi: 10.1053/j.gastro.2010.05.072 PMID: 20638935

37. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet. 2009; 4:885–90.

38. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. Nat Rev Genet. 2009; 10:353–8. doi: 10.1038/nrg2574 PMID: 19434079