



Departamento de Biomedicina. Área de Medicina Preventiva y Salud Pública  
Facultad de CC de la Salud  
Universidad de León

# EL PAPEL DE EZH2 EN EL CÁNCER COLORRECTAL

*The role of EZH2 in colorectal cancer*

Programa de doctorado en Biomedicina

## DIRECTORES DE LA TESIS

Antonio José Molina de la Torre

Cristina Díez Tascón

---

Laura Vilorio Marqués

León, 2017

---



La presente Tesis Doctoral ha sido posible gracias a la financiación de los proyectos LE22A10-2 y BIO/LE09/13 de la Junta de Castilla y León, y PI08/1770, PS09/01286 y PI15/01032 del Instituto de Salud Carlos III. El trabajo ha sido desarrollado en el Área de Medicina Preventiva y Salud Pública de la Universidad de León y en el servicio de Anatomía Patológica del Complejo Asistencial Universitario de León.

Las muestras biológicas utilizadas en este estudio han sido gestionadas con la colaboración del BEOCYL (Biobanco en Red de Enfermedades Oncológicas de Castilla y León) , perteneciente a la Plataforma Nacional de Biobancos (ISCIII).



# ***AGRADECIMIENTOS***



Esta Tesis Doctoral es el resultado de varios años de trabajo, y quisiera agradecer a todas las personas que han hecho posible su realización con su ayuda y apoyo:

Al Dr. Antonio José Molina de la Torre, quiero agradecerte el dirigirme este trabajo, por hacer que las cosas complicadas resulten sencillas. Gracias por la cercanía y la proximidad al trato, por el esfuerzo, la diplomacia, las horas, los ánimos, por ser apoyo y guía, muchas gracias.

A la Dra. Cristina Díez-Tascón, por dirigir este trabajo, si no te hubiera conocido, hoy no estaría redactando estos agradecimientos. Has sido una persona inspiradora, gracias por la motivación, el apoyo, la ayuda y la confianza.

Al Dr. Vicente Martín Sánchez por darme la oportunidad de trabajar en este grupo de investigación, por animarme a realizar este trabajo y apoyarme. Gracias por tu confianza, no hubiera sido posible sin tu ayuda.

A los compañeros del Área de Medicina Preventiva y Salud Pública, a Tania, Vero, Carmen, Lidia, Silvia, Laura, Andrea, Miriam, Jessy, Fernanda, a todas las personas que han pasado por el departamento, muchas gracias por compartir estos años, por el compañerismo, los ánimos, las comidas, los cumpleaños, no hubiera sido posible sin vosotros.

Al Servicio de Anatomía Patológica del CAULE, por facilitarme el desarrollo experimental de la tesis en sus laboratorios, en especial, muchas gracias a su Jefa de Servicio, la Dra. Teresa Ribas Ariño, por abrirme las puertas del Servicio Anatomía Patológica, darme todas las facilidades y hacerme sentir como en casa. Al Dr. Emiliano Honrado Franco, muchas gracias por la contribución a este trabajo con la realización e interpretación del "tissue array". Muchas gracias al resto del personal, a los técnicos, a los patólogos, extrañaré los agradables ratos del café en los que he aprendido tanto.

A los Servicios de Cirugía, de Oncología y de Farmacia del CAULE, por permitirme y ayudarme a realizar la parte clínica de este trabajo, gracias.

Muchas gracias al resto de investigadores que conforman el proyecto MCC-Spain y permitirme desarrollar esta Tesis Doctoral en el marco del proyecto, y por supuesto a los pacientes que han participado en el estudio, y que han dado una muestra enorme de generosidad en su enfermedad, para beneficio de los demás.

Muchas gracias a las Mariposas Rabiosas, porque una de las mejores experiencias de esta etapa, ha sido participar en este equipo de voleyball. Sin vosotras no hubiera sido lo mismo.

Muchas gracias a las compañeras femicletaccioneras, por seguir muchos años más pe(da)leando juntas.

A mis amigos, la familia que he escogido, sin vosotros no hubiera sido posible. En especial gracias a Bea, han sido momentos difíciles pero los hemos superado juntas. A Maite, a Lutxy, Patty, Mari, Desi, Vane, Gotzone, Bárbara, Diego, Alba, Vero... muchas gracias por la confianza.

Gracias a mi familia, a mis abuelos, a mis tíos, a mis primos, por el apoyo, la confianza, muchas gracias. Gracias a Oscar, que me ha apoyado siempre, tu ayuda ha sido inestimable.

Por último, gracias a mis padres, por servirme de ejemplo, ayudarme, apoyarme y confiar en mí. A mi padre Javi, por transmitirme el amor por la naturaleza y los libros, enseñarme a esforzarme y perseverar e intentar educarme como librepensadora, a mi madre Emi, por legarme su curiosidad insaciable y por depositar su confianza ciega en todo lo que hago.

Muchas gracias a todos, salud y resiliencia.

*La naturaleza me hace humilde y me prepara para soportar los golpes con valor.*

***(Ana Frank, 1945)***

*“Así es la vida”, a pesar de que sean pocas las veces en que viene descrita así: un implantarse, un barrer para adentro, un vampirizar el camino descendente de la energía, de su noble forma solar a la degradada de calor a baja temperatura. En su camino ascendente, que lleva al equilibrio y por ende a la muerte, la vida dibuja un asidero y anida en él.*

***(Carbono, Primo Levi, 1970)***

*No temas a las dificultades: lo mejor surge de ellas.*

***(Rita Levi-Montalcini, 1988)***



# ***ÍNDICE***



---

INTRODUCCIÓN .....	1
El cáncer colorrectal .....	3
Diagnóstico y clasificación del cáncer colorrectal .....	4
Epidemiología del cáncer colorrectal .....	9
Factores de riesgo del cáncer colorrectal .....	12
Bases moleculares del cáncer colorrectal.....	16
Cáncer colorrectal hereditario.....	16
Alteraciones genéticas en el cáncer colorrectal esporádico.....	20
Modelo “adenoma a adenocarcinoma” (modelo de Volgestein) .....	20
Epigenética del cáncer colorrectal.....	26
Complejos represivos del grupo Polycomb PcG .....	33
Complejo represivo Polycomb 1: PRC1.....	34
Complejo represivo Polycomb 2: PRC2.....	35
Estructura y función de EZH2 .....	37
Papel patológico de EZH2 .....	39
PLANTEAMIENTO Y OBJETIVOS .....	43
Planteamiento .....	45
Objetivos .....	46
DISEÑO EXPERIMENTAL.....	47
METODOLOGÍA.....	51
Revisión sistemática de la literatura.....	53
Estrategia de búsqueda .....	53

Selección e inclusión de estudios .....	53
Extracción de los datos cualitativos y cuantitativos .....	54
Análisis de la calidad metodológica de los estudios incluidos.....	54
Análisis estadístico de la revisión sistemática .....	54
Estudio de la cohorte de casos del proyecto MCC-Spain .....	55
Selección de pacientes .....	55
Obtención de variables epidemiológicas y clínico-patológicas .....	56
Obtención de muestras biológicas .....	58
Extracción de los ácidos nucleicos.....	59
Análisis de polimorfismos (SNPs) del gen EZH2.....	60
Análisis de expresión de EZH2 .....	61
RESULTADOS.....	67
Resultados revisión sistemática.....	69
Identificación de estudios relevantes/Proceso de búsqueda.....	69
Meta-análisis .....	70
Análisis de sesgos de publicación .....	71
Estudio de la cohorte de casos del proyecto MCC-Spain. ....	74
Análisis descriptivo de las características sociodemográficas y clínico-patológicas.....	74
Análisis descriptivo de los marcadores moleculares .....	77
Estudios de asociación de los marcadores moleculares con las características basales de los pacientes. ....	84
Análisis de supervivencia en función de los marcadores moleculares.....	98

---

DISCUSIÓN .....	107
Revisión sistemática .....	109
Estudio de la cohorte de casos del proyecto MCC-Spain .....	112
Análisis descriptivo de la muestra .....	112
Análisis de la expresión génica de EZH2 y su asociación con variables anatomopatológicas. ....	113
Análisis de la expresión proteica de EZH2 y su asociación con variables anatomopatológicas .....	114
Relación entre la expresión génica y proteica .....	115
Análisis de polimorfismos de EZH2.....	116
Análisis de la supervivencia en función de la expresión de EZH2.....	118
Análisis de supervivencia en función de los polimorfismos de EZH2 .....	121
Limitaciones y fortalezas .....	123
CONCLUSIONES .....	125
BIBLIOGRAFÍA .....	129
ANEXOS .....	147
ANEXO 1 .....	149
ANEXO 2 .....	151
ANEXO 3 .....	153
ANEXO 4 .....	155
ANEXO 5 .....	157
ANEXO 6 .....	159



## ***ÍNDICE DE FIGURAS***



Figura 1.- Tipos histológicos del cáncer colorrectal .....	7
Figura 2.- Mapa de incidencia mundial de cáncer colorrectal en ambos sexos.....	9
Figura 3.- Mapa de mortalidad mundial por cáncer colorrectal en ambos sexos.....	10
Figura 4.- Tasas de incidencia y mortalidad por CCR mundiales estandarizadas por edad, por cada 100,000 habitantes.....	11
Figura 5.- Fenotipo CIN: etapas de la carcinogénesis. ....	21
Figura 6.- Fenotipo MSI y Fenotipo CIMP. ....	23
Figura 7.- Mecanismo secuencial de acción de PRC2.....	36
Figura 8.- Interacción del PRC1 con el PRC2 provocando la condensación de la cromatina. ....	36
Figura 9.- EZH2 interactúa con otras enzimas como HDAC y DNMT cooperando para llevar a cabo la represión transcripcional de numerosos genes.....	38
Figura 10.- Algoritmo que representa el diseño experimental y la metodología llevada a cabo en este trabajo. ....	49
Figura 11.- Algoritmo de la revisión sistemática. ....	70
Figura 12.- Representación gráfica del meta-análisis. ....	71
Figura 13.- <i>Funnel Plot</i> realizado para comprobar la existencia de sesgos de publicación.....	71
Figura 14.- Distribución de segundos tumores primarios localizados durante el periodo de seguimiento. ....	77
Figura 15.- Ejemplo de expresión diferencial de una muestra tumoral con respecto a su pareada no tumoral. ....	78

Figura 16.- Dot plot de los valores de $\Delta C_t$ .....	79
Figura 17.- Niveles de expresión relativa de EZH2 en la cohorte de casos analizados. ....	79
Figura 18.- Imagen de un fragmento de la matriz de tejido “ <i>tissue array</i> ” .....	80
Figura 19.- Imágenes ampliadas de la matriz de tejido.....	81
Figura 20.- Curvas Kaplan Meier para el estudio de la OS en función de la sobreexpresión génica de EZH2.....	98
Figura 21.- Curvas Kaplan-Meier para el análisis de la OS en función de la expresión proteica de EZH2. ....	99
Figura 22.- Curvas Kaplan-Meier de la supervivencia global OS en función de los diferentes genotipos de los polimorfismos analizados. ....	101
Figura 23.- Curvas Kaplan-Meier para la representación de la supervivencia libre de enfermedad DFS en función de la sobreexpresión del gen EZH2. ....	102
Figura 24.- Curvas Kaplan-Meier para la representación de la supervivencia libre de enfermedad DFS en función de la-expresión proteica de EZH2.....	103
Figura 25.- Curvas Kaplan-Meier de la supervivencia libre de enfermedad DFS en función de los diferentes genotipos de los polimorfismos analizados.....	105

## ***ÍNDICE DE TABLAS***



Tabla 1.- Transcritos e isoformas proteicas del gen EZH2 descritas y sus respectivos tamaños	37
Tabla 2.- Posibles dianas de EZH2 y su efecto sobre la expresión génica. ....	39
Tabla 3.- Variables obtenidas de la historia clínica de la cohorte de casos .....	57
Tabla 4.- Secuencias de las sondas Taqman para la detección de SNPs en el gen EZH2, y ciclos de la reacción de PCR.....	61
Tabla 5.- Secuencias de los primers para la PCR cuantitativa de EZH2 y GADPH, tamaños de los amplicones y ciclos de la reacción de PCR. ....	63
Tabla 6.- Características de los estudios elegibles para la revisión sistemática.....	72
Tabla 7.- Valor pronóstico de la sobreexpresión de EZH2 en la OS del CCR. ....	73
Tabla 8.- Frecuencias de las características clinicopatológicas de la cohorte de casos. ....	75
Tabla 9.- Características de los SNPs, localización, variante, alelo ancestral y de menor frecuencia (MAF).....	82
Tabla 10.- Frecuencias genotípicas y alélicas de los SNPs de EZH2 y resultados del análisis del equilibrio Hardy-Weinberg. ....	84
Tabla 11.- Análisis de asociación de la expresión génica con las características sociodemográficas y anatomopatológicas. ....	86
Tabla 12.- Análisis de asociación de la expresión génica con las características sociodemográficas y anatomopatológicas.....	87
Tabla 13.- Asociación de la expresión proteica con las características sociodemográficas y anatomopatológicas. ....	89

Tabla 14.- Análisis de asociación de la expresión génica con las características sociodemográficas y anatomopatológicas. ....	90
Tabla 15.- Distribución de rs6958683 en función de las características anatomopatológicas. ..	95
Tabla 16.- Distribución de rs2302427 en función de las características anatomopatológicas. ..	96
Tabla 17.- Distribución de rs3757441 en función de las características anatomopatológicas. ..	97
Tabla 18.- Análisis crudo y ajustado por edad y estadio de la supervivencia global OS en función de los marcadores moleculares de EZH2. ....	100
Tabla 19.- Análisis crudo y ajustado por edad y estadio de la supervivencia libre de enfermedad (DFS) en función de los marcadores moleculares de EZH2. ....	104

## ***ABREVIATURAS***



---

AAPC	Poliposis Adenomatosa Colónica Atenuada
ADN	Ácido Desoxirribonucleico
ADNc	ADN complementario
ARN	Ácido Ribonucleico
ASL	Área de Salud de León
CAULE	Complejo Asistencial Universitario de León
CCR	Cáncer Colorrectal
CIBERESP	Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública
CIE-O	Clasificación Internacional de Enfermedades para Oncología
CIMP	CpG Island Methylator Phenotype
CIN	Chromosomal Instability
CRCSC	CRC Subtyping Consortium
DFS	Disease Free Survival
DNMT	DNA Methyl Transferase
DZNep	3-Deazaneplanocin A
EED	Embryonic Ectoderm Development
EZH2	Enhancer of Zeste Homolog 2
FRET	Fluorescence Resonance Energy Transfer
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GWAS	Genome-Wide Association Study
H	Histona
H3K27me2	Histona 3 dimetilada en el residuo de lisina 27
H3K27me3	Histona 3 trimetilada en el residuo de lisina 27
HAT	Histona Acetilasa
HDAC	Histona Deacetilasas
HDMT	Histona Demetil-Transferasa
HE	Hematoxilina-Eosina
HMT	Histona Metil-Transferasa
HNPCC	Cáncer Colorrectal Hereditario No Polipósico o Síndrome de Lynch
HR	Hazard Ratio
IHC	Inmunohistoquímica
lncARN	ARN largo no codificante

LTI	Laboratorio de Técnicas Instrumentales de la Universidad de León
MAP	Poliposis Asociada a MUTHY
MCC-Spain	Multicaso Control de los Cánceres poblacionales más frecuentes
MET	Metabolic Equivalents
miARN	Micro ARN
MMR	Mismatch Repair
MSI	Microsatellite Instability
OMS	Organización Mundial de la Salud
OR	Odds Ratio
OS	Overall Survival
PAF	Poliposis Adenomatosa Familiar
PCR	Polymerase Chain Reaction
PET	Tomografía por emisión de positrones
PgC	Polycomb Group Complex
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PTM	Post-Translational Modifications
qPCR	Real Time Polymerase Chain Reaction
REMARKS	Reporting Recommendations for Tumor Marker Prognostic Studies
RIN	RNA Integrity Number
RQI	RNA Quality Number
RT-PCR	Retro-Transcriptase Polymerase Chain Reaction
SAM	S-Adenosil-Metionina
SNP	Single Nucleotide Polymorphism
SUZ12	Supresor of Zeste 2
TC	Tomografía Computerizada
TNM	Tumor-Node-Metastasis
WOS	Web of Science

# ***INTRODUCCIÓN***



## ***El cáncer colorrectal***

El término cáncer agrupa un gran número de patologías que tienen en común células tumorales que presentan un crecimiento celular descontrolado y una diseminación a otros tejidos. En base al trabajo publicado por Hanahan y Weinberg <sup>[1,2]</sup>, se definieron varias anomalías distintivas que exhiben estas células tumorales: (a) autoestimulación mitótica o potencial replicativo ilimitado, (b) evasión de los inhibidores del crecimiento (c) resistencia a la muerte celular programada, (d) potencial replicativo ilimitado, (e) angiogénesis sostenida, (f) reprogramación del metabolismo energético celular, (g) evasión de la destrucción inmunitaria, (h) invasión tisular y metástasis, (i) inestabilidad genómica y (j) capacidad pro-inflamatoria. El proceso por el cual una célula normal se convierte en una célula cancerígena implica cambios dinámicos en el genoma, que desembocan en la reprogramación de la célula, provocando que se reproduzca de manera descontrolada, originando de esta forma una masa, en el caso de las neoplasias sólidas, conocida como tumor, que ha perdido la arquitectura tisular. La acumulación de mutaciones en las células tumorales, debido a la inestabilidad genómica, supone una ventaja adaptativa y les permite adquirir nuevas propiedades, como la capacidad de migrar a otros tejidos con el fin de establecerse en nuevos nichos y proliferar dando lugar a metástasis.

Los tipos de cáncer reciben el nombre del tejido o el órgano donde se originan<sup>[3]</sup>. Así, el cáncer colorrectal (CCR) es el tipo de cáncer que afecta a las células del colon y del recto, que conforman el intestino grueso, en todas las localizaciones anatómicas. El intestino grueso es el último tramo del tubo digestivo, tiene una longitud aproximada de 1.5 metros y se extiende desde el final del intestino delgado (píloro) hasta el ano. La primera porción del intestino grueso está situada en la parte inferior derecha del abdomen, es donde desemboca el intestino delgado y se llama ciego. Desde aquí el colon asciende (colon ascendente) hasta llegar al ángulo hepático y atraviesa el abdomen (colon transversal) hasta el ángulo esplénico. El colon se dirige posteriormente hacia abajo, denominándose colon descendente, hasta llegar a una zona denominada sigma que desemboca en el recto y finalmente en el ano que se abre al exterior por el esfínter anal a través del cual se eliminan las heces. En cualquiera de estas localizaciones anatómicas se puede originar un cáncer colorrectal. <sup>[3,4]</sup>

## ***Diagnóstico y clasificación del cáncer colorrectal***

### ***Diagnóstico clínico***

El diagnóstico clínico del CCR se inicia ante la aparición de síntomas tales como sangre en heces, roja y brillante y/o asociada a las heces, de color negro y se denominan melenas, anemia, alteraciones en el ritmo o forma de las deposiciones, tenesmo o sensación de evacuación incompleta, cansancio extremo, pérdida de peso sin causa aparente o dolor abdominal. Es una sintomatología inespecífica y puede manifestarse o no, por lo que en algunos casos el diagnóstico se realiza en estadios avanzados de la enfermedad<sup>[5-10]</sup>.

Las técnicas diagnósticas pretenden, en primer lugar, determinar la localización del tumor primario y su posterior estudio, mediante anamnesis, y exploración física del paciente. Técnicas de imagen como la resonancia magnética (RM), o la tomografía computerizada de abdomen (TC) y la tomografía por emisión de positrones (PET) ayudan a completar el estudio y valorar el grado de extensión de la enfermedad<sup>[5-10]</sup>.

### ***Diagnóstico anatomopatológico***

Aunque el desarrollo tecnológico de los últimos años ha permitido que se implementen nuevos análisis de biomarcadores moleculares, la base fundamental para establecer el diagnóstico de CCR sigue siendo el estudio histológico, junto con la integración de datos clínicos.

El diagnóstico anatomopatológico de las muestras de CCR, se realiza de acuerdo con un protocolo general estandarizado que figura en el Anexo 1.

En primer lugar, se realiza la observación macroscópica de la pieza quirúrgica que llega del quirófano al laboratorio de Anatomía patológica. El examen macroscópico de las piezas debe ser siempre exhaustivo, teniendo en cuenta la localización del tumor, sus dimensiones máximas y el estado de los márgenes de resección (si están libres o afectados por el tumor), describiendo las características macroscópicas, el tamaño, la presencia de áreas hemorrágicas, la presencia de necrosis, las distintas consistencias y la infiltración o no de tejidos adyacentes<sup>[5]</sup> (Anexos 2 y 3).

Después se valoran los dos componentes principales del tumor, el parénquima y el estroma, la densidad celular y el patrón de distribución principal de las células, y luego la morfología individualizada de dichas células, considerando el núcleo y el citoplasma, las mitosis, etc. El estroma nos muestra la estructura de la sustancia fundamental extracelular, los vasos acompañantes y el posible componente inflamatorio asociado.

La tinción básica de hematoxilina-eosina (HE) y las técnicas auxiliares de inmunohistoquímica (IHQ) permiten orientar el diagnóstico y encajar el tumor en un grupo u otro con unas primeras nociones hacia el pronóstico. El estudio molecular es cada vez más relevante, proporcionando información pronóstica y predictiva de respuesta al tratamiento, y permitiendo seleccionar a los pacientes que se pueden beneficiar de determinadas terapias en un acercamiento a lo que hoy se conoce como medicina de precisión.

Con todo ello se pueden ir delineando una serie de premisas básicas en patología tumoral e integrando información histológica, molecular y clínica cada vez más eficaces e individualizadas<sup>[6,7,13]</sup>.

### *Clasificación anatomopatológica*

La evaluación histológica de un tumor proporciona una guía útil para predecir su comportamiento y una base racional para plantear el tratamiento. La evaluación histológica debe establecer los siguientes aspectos<sup>[5-9]</sup>

- **Tipo histológico:** se basa en el tejido de origen del que procede el tumor (tejido epitelial, tejido conectivo, linfoide)
- **Grado de diferenciación:** se basa en las características siguientes:
  - Semejanza del tumor con el supuesto tejido de origen, tanto arquitectural como citológicamente (**diferenciación**)
  - Grado de variabilidad de la forma y el tamaño celular (**pleomorfismo**)
  - Proporción de células en mitosis

En base a esos dos criterios, en la 4ª clasificación histológica del CCR de la Organización Mundial de la Salud (OMS) se consolidan los siguientes tipos histológicos del CCR<sup>[8-13]</sup> que se pueden observar en la figura 1.

### *Adenocarcinoma*

Más del 90% de todos los carcinomas colorrectales son, desde el punto de vista histológico, adenocarcinomas originados a partir de las células epiteliales de la mucosa colorrectal. Este tipo de carcinoma epitelial exhibe un patrón de crecimiento glandular, que es la base para determinar el grado de diferenciación histológico.

En los adenocarcinomas bien diferenciados o tumores de bajo grado, más del 95 % del tumor está formado por el patrón glandular, los núcleos celulares son hiper cromáticos, con una relación núcleo-citoplasma alta, y se observan numerosas mitosis; en los moderadamente

diferenciados se observa este patrón celular en el 50-95%, y los tumores pobremente diferenciados o tumores de alto grado exhiben una tendencia mínima a la formación del patrón glandular y sus células son extremadamente pleomórficas<sup>[8,9]</sup>.

#### *Adenocarcinoma mucinoso*

Este tipo especial de CCR se caracteriza por presentar más del 50 % del volumen del tumor está formado por mucina extracelular. Hay tumores que exhiben una composición de mucina del 10 al 50% y se conocen como adenocarcinomas con características mucinosas o con diferenciación mucinosa.

El adenocarcinoma mucinoso exhibe grandes estructuras celulares (vacuolas) llenas de mucina extracelular. Muchos adenocarcinomas mucinosos aparecen en pacientes con cáncer colorrectal hereditario no-poliposo (HNPCC) o Síndrome de Lynch<sup>[8,9]</sup>.

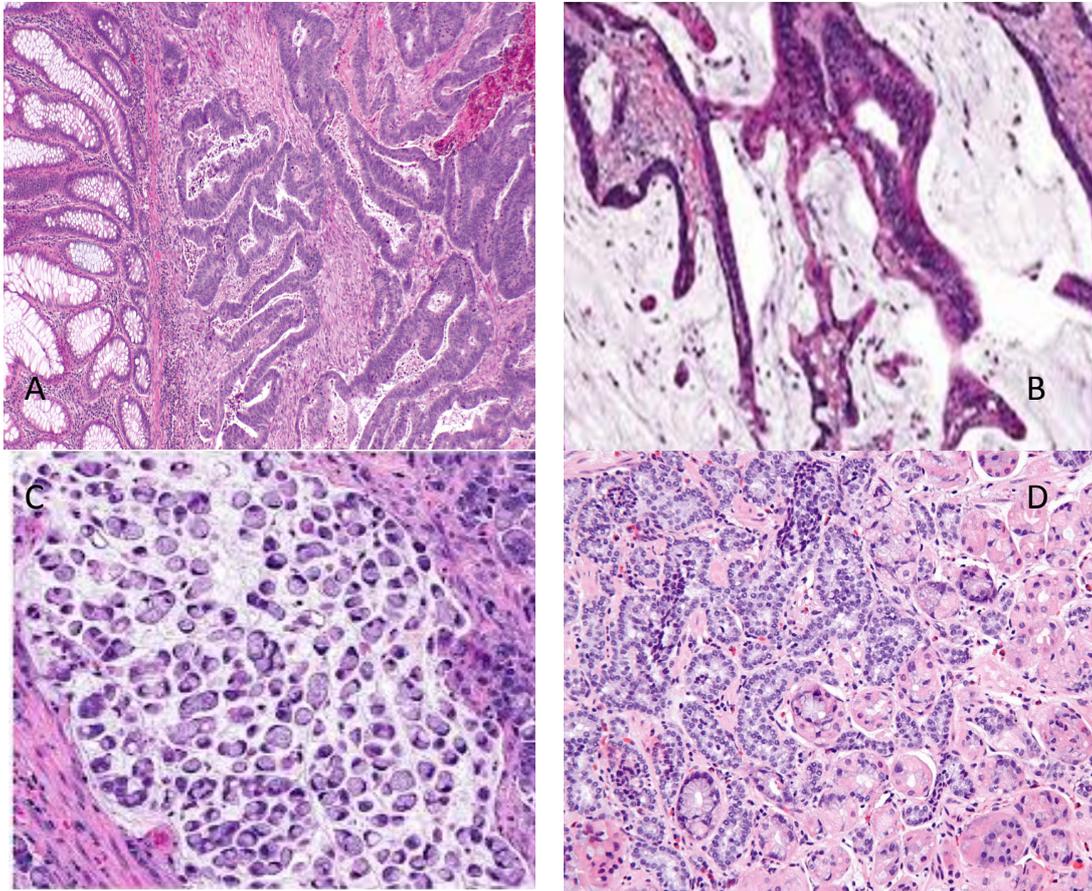
#### *Carcinoma de células en anillo de sello*

El carcinoma de células en anillo de sello es un tipo de tumor poco frecuente en el colon (en contraste con el cáncer gástrico en el que aparece como un tipo histológico frecuente), representando el 1% de todos los tumores colorrectales. Se presenta de manera similar al adenocarcinoma mucinoso, caracterizándose por la presencia de más del 50% de células en anillo de sello, que presentan una prominente vacuola de mucina intracitoplasmática que empuja el núcleo hacia la periferia. Las células en anillo de sello presentan un patrón de crecimiento e infiltración dentro del conjunto de la mucina extracelular. Por definición, este tipo histológico es pobremente diferenciado (tumores de alto grado) y comporta un peor pronóstico que el adenocarcinoma convencional<sup>[8,9]</sup>.

#### *Tumores neuroendocrinos*

Son tumores derivados de las células del sistema neuroendocrino que segregan hormonas polipeptídicas o aminas fisiológicamente activas. Los tumores neuroendocrinos pueden ser benignos o malignos, y se originan principalmente en el tracto gastrointestinal, el páncreas, las glándulas suprarrenales y el tiroides. Se aplica una clasificación funcional cuando es posible identificar el producto segregado, por ejemplo insulinoma, gastrinoma y glucagonoma. No obstante, muchos tumores neuroendocrinos no tienen productos secretores identificables. Por tradición, se denominan **tumores carcinoides** a los tumores de células endocrinas que segregan 5-hidroxitriptamina y pueden dar lugar al llamado síndrome carcinoide.

Los tumores neuroendocrinos comparten ciertas características histológicas: las células son relativamente pequeñas y uniformes, con núcleos prominentemente redondos y citoplasma granular característico debido a la presencia de gránulos secretores<sup>[8,9]</sup>.



**Figura 1.-** Tipos histológicos del cáncer colorrectal A) Adenocarcinoma colorrectal, B) Adenocarcinoma mucinosos, C) Carcinoma de células en anillo de sello, D) Tumor neuroendocrino gastrointestinal<sup>[14]</sup>

### Estadificación

El sistema de estadificación que se emplea con más frecuencia para el cáncer colorrectal es el sistema TNM del American Joint Committee on Cancer (AJCC)<sup>[15-19]</sup> que se basa en tres piezas clave de información:

- ✓ Categoría T = "Tumor": tamaño, extensión o profundidad de la penetración del tumor primario.
- ✓ Categoría N= "Nodo linfático": ausencia o presencia de afectación de la región del ganglio linfático.
- ✓ Categoría M= "Metástasis": ausencia o presencia de metástasis a distancia.

Las categorías T de cáncer colorrectal describen la extensión de la propagación a través de las capas que forman la pared del colon y del recto. Estas capas, de la interior a la exterior, incluyen:

- ✓ El revestimiento interno (mucosa), la capa en la que casi todos los cánceres colorrectales comienzan. Esto incluye una capa muscular delgada (muscularis mucosae)
- ✓ El tejido fibroso que se encuentra debajo de esta capa muscular (submucosa).
- ✓ Una capa muscular gruesa (muscularis propria).
- ✓ La capa delgada más externa de tejido conectivo (la subserosa y la serosa) que cubre la mayor parte del colon, pero no el recto.

Estos factores **T**, **N**, **M** se dividen en categorías que van de 0 a 4 indicando el incremento de severidad. En cada tumor se determina su categoría TNM y se asocia al estadio correspondiente. Las definiciones de cada estadio, siguiendo la clasificación TNM, son las siguientes: estadio I (el menos avanzado) (T1 o T2 N0 M0), estadio II (T3 o T4 N0 M0), estadio III (cualquier T N1 o N2 M0) y estadio IV (cualquier T o N M1 )<sup>[15,16]</sup>.

El sistema TNM se ha desarrollado bajo la influencia de los avances en el diagnóstico y el tratamiento. Las imágenes radiográficas han mejorado la exactitud de la clínica T, N, M y las clasificaciones. Los avances en el tratamiento han obligado a dividir más detalladamente algunas categorías de T4. El uso de técnicas de inmunohistoquímica se ha traducido en la clasificación de las células tumorales aisladas diferenciándolas de micrometástasis.<sup>[15,16]</sup> En la séptima edición de la clasificación TNM se han incorporado nuevas subdivisiones de categorías en función de estos avances<sup>[17-20]</sup>.

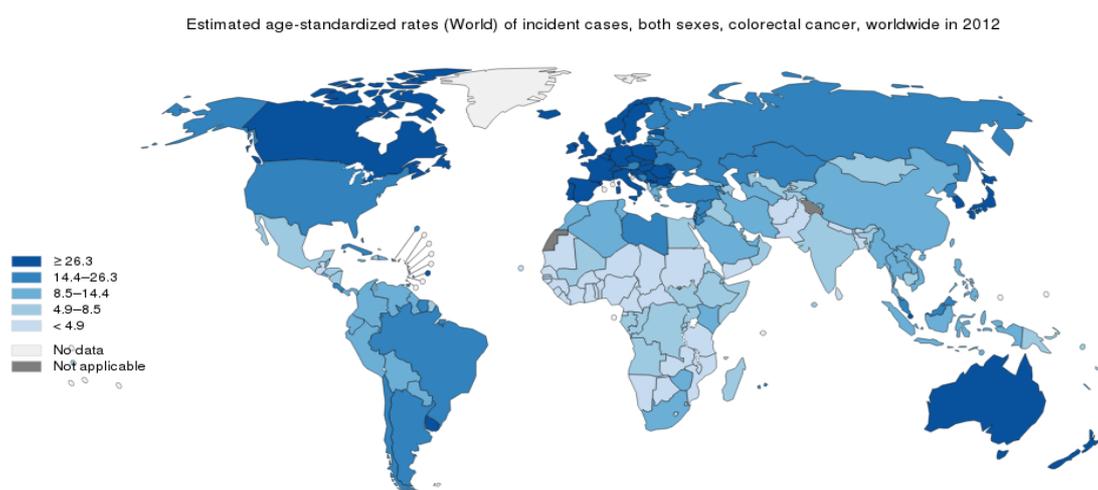
- ✓ La categoría T4 ha sido subdividida en dos categorías T4a y T4b (el tumor invade el peritoneo u otros órganos, o bien estructuras adyacentes, respectivamente).
- ✓ La categoría pN1 ha sido subdividida en N1a (metástasis en 1 ganglio linfático regional), N1b (metástasis en 2-3 ganglios) y N1c (depósitos tumorales en la subserosa, el mesenterio, o el tejido perirrectal, sin presencia de metástasis en los ganglios linfáticos regionales).
- ✓ La categoría pN2 ha sido subdividida en pN2a (metástasis en 4-6 ganglios linfáticos regionales) y en pN2b (metástasis en 7 o más ganglios linfáticos)
- ✓ En la categoría M, se contempla la subdivisión en M1a (metástasis en un solo órgano o lugar) y M1b (metástasis en más de un órgano o lugar o en el peritoneo)<sup>[17]</sup>

## Epidemiología del cáncer colorrectal

### Incidencia

El cáncer colorrectal a nivel mundial, ocupa la tercera posición en incidencia (1,360,000 nuevos casos anuales), siendo el tercer tipo de cáncer más frecuente en hombres y el segundo más frecuente en mujeres. En el año 2012, se estimó la incidencia en 746,000 nuevos casos en hombres, que es el 10% de incidencia de todos los cánceres en hombres y 614,000 nuevos casos en mujeres, lo que supone el 9.2% de todos los cánceres incidentes en mujeres<sup>[21]</sup>. El mapa reproducido en la figura 2, representa la incidencia del CCR mundial en función de las diferentes regiones geográficas.

Existe una variación geográfica importante en la incidencia de este cáncer a nivel mundial, observándose que un 55% de todos los casos se producen en países desarrollados, siendo los países que mostraron mayores tasas de incidencia estimadas (por 100,000 habitantes), en el año 2012 Australia y Nueva Zelanda (44.8 casos en hombres y 32.2 en mujeres), y los que mostraron las tasas más bajas los países del oeste de África (4.5 en hombres y 3.8 en mujeres). Este hecho, junto con estudios que han observado que individuos que han migrado de una zona de baja incidencia a un país con alta incidencia de la enfermedad adquieren el riesgo del país de acogida, sugiere que puede existir cierta influencia del ambiente, de los hábitos de vida y dietéticos, que actúen en conjunción con la interacción de factores genéticos, para determinar las diferentes incidencias de la enfermedad entre regiones geográficas<sup>[21]</sup>.



**Figura 2.-** Mapa de incidencia mundial de cáncer colorrectal en ambos sexos. <sup>[21]</sup>

En Europa, el CCR ocupa la tercera posición en incidencia, por detrás del cáncer de próstata y el de mama, pero en España ocupa la primera posición, se ha estimado en el año 2015 el número de nuevos casos en 41,441, siendo más frecuente en hombres (24,764) que en mujeres (16,677). En España, se ha observado un incremento de la incidencia de este tipo de tumor a partir de mediados de los años 90, principalmente en hombres. A pesar de la implantación del cribado y *screening* del CCR, la tendencia sigue en aumento, y este aumento en la incidencia puede deberse a la exposición a factores de riesgo, ya que se ha visto que el cribado no es la causa de este aumento de casos diagnosticados<sup>[22]</sup>.

En lo que se refiere al Área de Salud de León (ASL) se ha observado un incremento en la incidencia así como una distribución heterogénea, afectando en mayor medida a la capital y a las zonas central y sur del ASL. Los resultados de estos trabajos son coincidentes con los informes publicados en el “Atlas de Mortalidad Municipal por Cáncer y otras causas”, que reflejan una elevada mortalidad por CCR en la zona sur y oriental de la provincia de León<sup>[23-25]</sup>.

### Mortalidad

En cuanto a mortalidad a nivel mundial, las tasas son menores que las de incidencia, se estimaron los casos de muerte a consecuencia de esta enfermedad en 694,000, que supone el 8,5% de las muertes por cáncer al año. La mayor parte de las muertes por CCR, el 52% en 2012, sucedieron en las regiones del planeta menos desarrolladas, reflejando la menor supervivencia en esos lugares, en parte debido a las condiciones de pobreza y menor nivel socioeconómico, y el acceso al sistema de salud en dichas regiones.

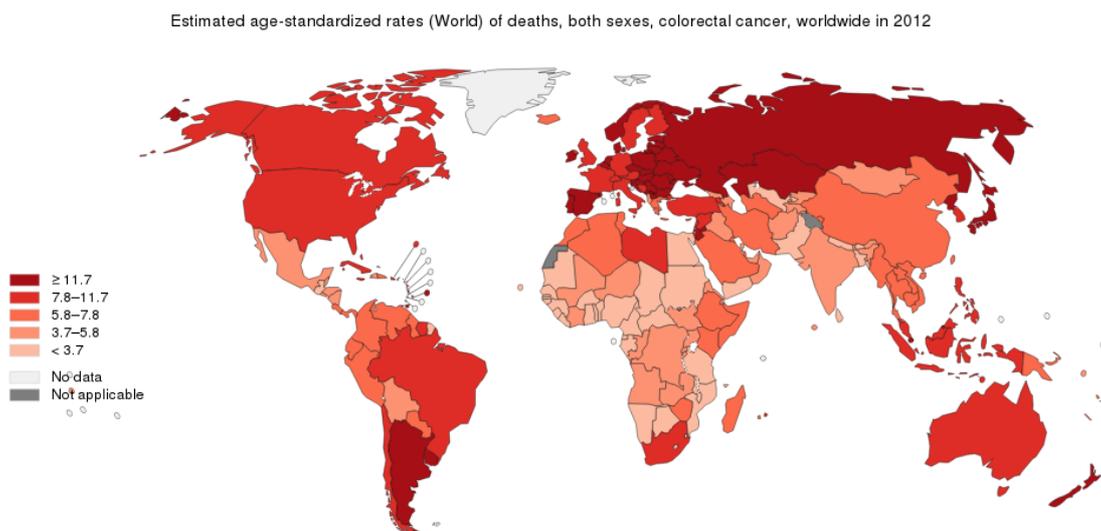
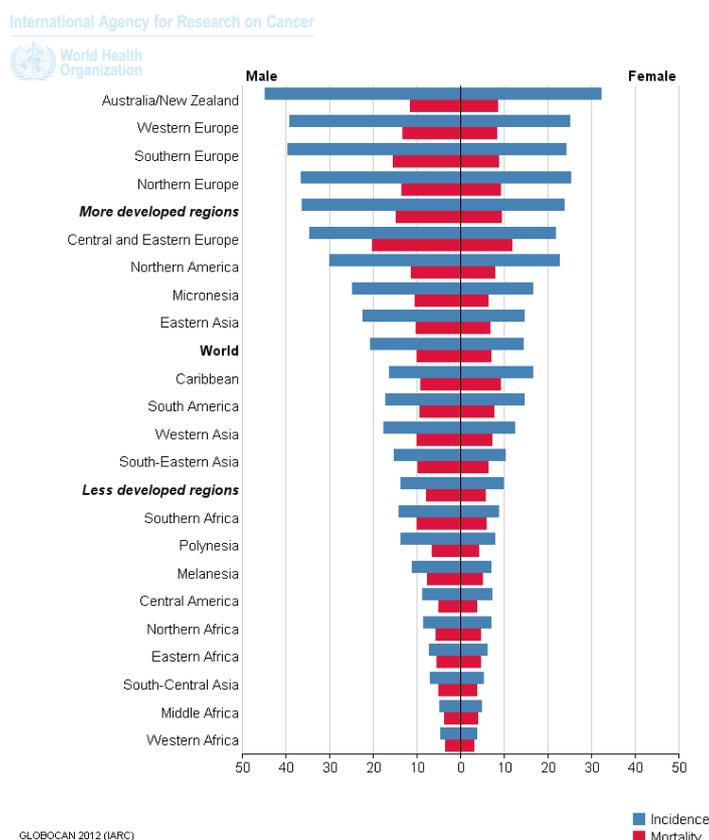


Figura 3.- Mapa de mortalidad mundial por cáncer colorrectal en ambos sexos.<sup>[21]</sup>

En ambos sexos las tasas de mortalidad estimadas más elevadas (por cada 100,000 habitantes) se localizaron en países de Centroeuropa y Europa del este (20.3 muertes en hombres y 11.7 en mujeres) y las menores tasas de mortalidad al igual que la incidencia en el oeste de África (3.5 muertes en hombres y 3.0 en mujeres) [21]. El mapa reproducido en la figura 3, representa la mortalidad por CCR mundial en función de las diferentes regiones geográficas y la figura 4 representa las tasas de incidencia y mortalidad de estos países estandarizadas por edad (por cada 100,000 habitantes).

En España las muertes por CCR durante el periodo de 1989-2008 fueron 224,305 (123,965 en hombres y 100,340 en mujeres) suponiendo un 14,5% de las muertes producidas por cáncer. El patrón geográfico de la tasa de mortalidad para el cáncer colorrectal no es muy pronunciado y presenta similitudes a la distribución entre sexos, aunque se ha observado en hombres y mujeres (en el primer y segundo periodo de 5 años de seguimiento respectivamente) unas elevadas tasas de mortalidad en ciudades de Cataluña y en localidades de la provincia de León [25].



**Figura 4.-** Tasas de incidencia y mortalidad por CCR mundiales estandarizadas por edad, por cada 100,000 habitantes. [21]

La supervivencia del CCR está determinada por el estadio de la enfermedad al diagnóstico. Si la enfermedad está localizada y puede ser resecada quirúrgicamente presenta una supervivencia a los 5 años del 89.9%. Los pacientes que presentan la enfermedad avanzada regionalmente, con afectación de los ganglios linfáticos, muestran una supervivencia a los cinco años de 71.3%. Sin embargo, si la enfermedad está diseminada a órganos distantes, la supervivencia a los cinco años es del 13.9%. La supervivencia relativa global a nivel mundial del CCR es del 64.9%<sup>[20]</sup>.

### ***Factores de riesgo del cáncer colorrectal***

Existen factores de riesgo que influyen en los procesos de desarrollo del cáncer colorrectal y están relacionados con las modificaciones en las tasas de incidencia de esta enfermedad. Entre ellos se encuentran, además de la edad, la alimentación, el consumo de alcohol y tabaco, la actividad física y el sedentarismo o el uso de determinados fármacos. No obstante, la exposición a un factor, o incluso a varios factores de riesgo, no significa que una persona padecerá la enfermedad. Por otra parte, es posible llegar a padecerla sin haber estado expuesto a ningún factor de riesgo conocido.

El cáncer colorrectal es un importante y emergente problema de salud pública en España debido al aumento y evolución de la incidencia y la elevada prevalencia de los factores de riesgo conocidos (dieta, sobrepeso, estilo de vida sedentario, etc.), ya que más del 90% de los casos diagnosticados en este país ocurren de forma esporádica por interacción con la exposición ambiental, y sólo unos pocos se corresponden con síndromes hereditarios asociados a la predisposición a desarrollar la enfermedad, de los que hablaremos más adelante en este trabajo.

Existen dos tipos de factores de riesgo para el cáncer colorrectal, los intrínsecos o no modificables como la edad, el sexo, la predisposición genética o enfermedad colónica previa y los factores de riesgo modificables, que vienen influidos por los hábitos y el estilo de vida o la exposición a tóxicos como el alcohol o el tabaco<sup>[20,26]</sup>.

### ***Factores de riesgo modificables***

#### *Obesidad*

El sobrepeso aumenta el riesgo de cáncer de colon tanto en los hombres como en las mujeres, aunque esta asociación parece ser mayor entre los hombres<sup>[20]</sup>. Se ha determinado que la grasa corporal, y sobre todo la grasa abdominal, son factores causales de CCR<sup>[26]</sup>. La obesidad es un factor de riesgo para desarrollar carcinogénesis de cualquier tipo, pero en el

caso del CCR esta relación se hace más estrecha todavía. Esto se debe a que los pacientes con obesidad muestran hiperinsulinemia y resistencia a la insulina, con un incremento en el bioactivo IGF-1. Tanto la insulina como el IGF-1 producen un incremento en la proliferación celular anormal y reducen la apoptosis. Además, el consumo de grasas animales incrementa la producción de prostaglandinas, lo cual induce al desarrollo de clones celulares tumorales. Por otro lado, produce un incremento la producción de esteroides y provoca cambios en la composición y actividad de la flora bacteriana intestinal<sup>[27,28]</sup>.

### *Actividad física*

La actividad física y el ejercicio físico reducen la incidencia y la mortalidad por CCR. En trabajos realizados para estudiar el efecto de la actividad física en el cáncer de colon, y en los que se ha empleado como medida del ejercicio físico los MET·horas globales estimados a partir de las actividades realizadas, se ha observado que existe una mayor protección frente al CCR en los individuos que más actividad física han realizado. Un elevado nivel de actividad física total está asociado significativamente con un menor riesgo de CCR<sup>[30]</sup>. Esta protección proporcionada por la práctica de ejercicio físico puede deberse a la reducción de la inflamación, que promueve la tumorigénesis y a un mayor aumento de la motilidad intestinal, que reduce el tiempo de exposición de la mucosa del colon a sustancias potencialmente carcinógenas<sup>[31,32]</sup>. Por el contrario, el sedentarismo, entendiendo como tal el tiempo que se pasa sentado, es un factor de riesgo asociado al desarrollo del CCR, tal y como ha mostrado un estudio reciente que asoció el tiempo que pasaba su muestra viendo la televisión con un incremento del 54% del riesgo de padecer CCR y el tiempo que pasaba sentado en el trabajo con aumentos del 24% de dicho riesgo<sup>[29]</sup>.

### *Dieta*

La relación entre la dieta y el CCR se ha estudiado de diferentes formas, analizando alimentos, analizando nutrientes y analizando patrones dietéticos. Diversos estudios relacionan un elevado consumo de carne procesada y roja con el riesgo de desarrollar CCR, mientras que un consumo de vegetales y pescado tiene un efecto protector<sup>[33]</sup>.

Se han analizado nutrientes específicos de los alimentos y se ha observado que un elevado consumo de fibra reduce significativamente el riesgo de adenoma y CCR<sup>[34-36]</sup>, porque la ingesta de fibra aumenta el tránsito intestinal y reduce la inflamación<sup>[35]</sup>. El papel del ácido fólico tiene efectos contrarios dependiendo del estadio en el que se encuentre la enfermedad: si no existe enfermedad, el bajo consumo de folato promueve la carcinogénesis; sin embargo si la enfermedad está presente, un bajo consumo de folato protege de la progresión<sup>[36]</sup>.

Se han realizado estudios que incluyen patrones de nutrientes, en los que se ha observado como el patrón de vitaminas y minerales y el patrón caracterizado por vitamina B12, riboflavina, calcio y colesterol, y proteínas y fósforo se asocian con un menor riesgo de desarrollar CCR en contraposición a los patrones de nutrientes caracterizados por alta ingesta de grasas saturadas de procedencia animal, que se relacionan con un mayor riesgo de padecer CCR<sup>[37]</sup>.

Analizar cómo influyen los patrones dietéticos en el riesgo del CCR aporta información acerca de las asociaciones entre grupos de alimentos y nutrientes, ya que los alimentos interactúan y presentan sinergias entre sí. Según algunos trabajos de reciente publicación en los que se estudian los patrones de las dietas occidental y prudente, se concluyó que la dieta prudente, que contiene alimentos como frutas, vegetales, pescado, legumbres, cereales y soja es protectora frente al cáncer, mientras que el patrón de dieta occidental, caracterizado por una ingesta elevada de carne roja y alimentos procesados, azúcares añadidos y granos refinados conlleva un mayor riesgo de desarrollar CCR<sup>[38,39]</sup>.

### ***Factores de riesgo no modificables***

#### *Edad*

El cáncer es una enfermedad asociada al envejecimiento. El envejecimiento implica acumulación de errores en los procesos de división celular y la falta de reparación de estos errores. La incidencia de CCR aumenta a partir de los 50 años, edad a partir de la cual se diagnostican en España más del 90% de los casos<sup>[22]</sup>.

#### *Neoplasia colorrectal previa*

Una historia de CCR previa favorece e incrementa el riesgo de desarrollar tumores metacrónicos. El riesgo se ve incrementado entre un 6 y un 25%. Paralelamente, una historia previa de pólipos adenomatosos puede incrementar el riesgo de desarrollar adenomas que pueden malignizarse y llegar a convertirse en CCR<sup>[40]</sup>.

#### *Enfermedades inflamatorias intestinales*

La enfermedad inflamatoria intestinal es un factor de riesgo para el desarrollo del CCR. La prevalencia de CCR en pacientes con colitis ulcerosa se ve incrementada del 3.7 al 5.4% y el efecto es acumulativo al pasar el tiempo con esta enfermedad, siendo el riesgo acumulado de desarrollar CCR aproximadamente del 2% a los 10 años del diagnóstico de colitis ulcerosa, del 8% a los 20 años y del 18% a los 30 años de padecer esta enfermedad. La incidencia y el riesgo

de CCR para los pacientes de enfermedad de Crohn son similares a los reportados para los pacientes de colitis ulcerosa<sup>[40]</sup>.

### *Historia familiar*

Si se excluyen los síndromes de cáncer hereditarios, que describiremos más adelante, los individuos con antecedentes familiares de CCR o adenomas colorrectales tienen un mayor riesgo de padecer CCR.<sup>[40,41]</sup>

### *Síndromes hereditarios*

Existe la evidencia de que hay determinados síndromes genéticos hereditarios que predisponen a los individuos a padecer CCR. Estos síndromes serán abordados más adelante en el apartado de CCR hereditario<sup>[41,42]</sup>.

### *Polimorfismos genéticos de tipo SNP*

Los SNPs (del inglés, *single nucleotide polymorphisms*) son variaciones genéticas que afectan a un único nucleótido de la secuencia de ADN. Presentan dos alelos, que pueden estar en homocigosis o en heterocigosis, y cuyas frecuencias pueden ser variables en distintas poblaciones. Estas variaciones definen genéticamente a cada individuo y pueden ser marcadores de susceptibilidad a enfermedades o de éxito/fracaso de un tratamiento farmacológico. En el contexto del cáncer, la presencia de un alelo u otro puede conferir una mayor o menor susceptibilidad de padecerlo. En CCR los estudios de GWAS (*Genome-wide association study*) han identificado aproximadamente 40 *loci* comunes para el CCR esporádico. Se ha observado en estos *loci* cómo las variaciones genéticas o SNPs confieren un efecto acumulativo y creciente para el CCR. Identificar estas variantes ayuda a predecir la susceptibilidad de los individuos a desarrollar la enfermedad además de orientar el pronóstico. En otras palabras, la detección y caracterización de SNPs está en la base de la llamada medicina personalizada<sup>[43,44]</sup>.

## ***Bases moleculares del cáncer colorrectal***

El cáncer de colon es un proceso multietapa, caracterizado por la acumulación de eventos moleculares. En el modelo simple de progresión tumoral de adenoma a carcinoma incluye mutaciones o sobreexpresión de diferentes oncogenes y genes supresores tumorales que juegan un papel crucial en la tumorigénesis colorrectal.<sup>[45]</sup>

Un análisis molecular más detallado del cáncer de colon muestra una gran heterogeneidad de alteraciones moleculares, lo que implica poder clasificar los procesos neoplásicos en base a sus alteraciones en fenotipos tumorales específicos<sup>[45]</sup>. Avances recientes en el conocimiento de la biología del tumor han hecho posible el descubrimiento de marcadores moleculares nuevos y potencialmente útiles que pueden convertirse en prometedores biomarcadores pronósticos y/o predictivos<sup>[46]</sup>.

Actualmente se conoce que el CCR es una enfermedad en la que están involucrados cambios genéticos y epigenéticos. Los cambios genéticos pueden ser somáticos o germinales, definiendo los dos grandes patrones observados: los síndromes hereditarios y los casos esporádicos, respectivamente.

### ***Cáncer colorrectal hereditario***

El cáncer colorrectal hereditario representa el 10% de todos los casos de CCR. Se caracteriza por mutaciones en las células germinales o gametos, en genes supresores tumorales.

Knudson<sup>[47]</sup> postuló su hipótesis del doble impacto que explica la mayor susceptibilidad al desarrollo del cáncer en los pacientes con determinados síndromes hereditarios. Según esta hipótesis, el cáncer sólo se desarrolla cuando los dos alelos de ciertos genes con herencia autosómica dominante están inactivos. En los casos hereditarios, el individuo nace con un alelo inactivado y sólo precisaría una segunda mutación en el otro alelo del gen para desarrollar la enfermedad, mientras que en los casos esporádicos se precisaría la mutación en ambos alelos. Más reciente es la hipótesis de los genes guardianes<sup>[48]</sup>, que postula que en cada tipo de tejido existe un gen particular que desempeña un papel esencial en el mantenimiento de un número constante de células.

Distintos síndromes hereditarios son causados por mutaciones germinales en genes supresores tumorales, o que regulan su expresión, o bien en genes reparadores del ADN. De entre ellos, destacan los síndromes de poliposis adenomatosa familiar (PAF), de Cowden y de Lynch.

### ***Síndrome de poliposis adenomatosa familiar PAF.***

Se caracteriza por la presencia de cientos y hasta miles de pólipos a través del epitelio colónico. Es un síndrome con una herencia de tipo autosómico dominante y con una penetrancia aproximada del 100%. Si los pólipos adenomatosos de la PAF no son tratados se malignizarán y se transformarán en CCR. La media de edad de aparición del CCR en los pacientes en los que no se han tratado los adenomas se sitúa entre los 39 y 42 años. Es un síndrome que se produce por una mutación del gen APC y que se caracteriza por la presencia de cientos de pólipos que aparecen en una edad temprana. El producto del gen APC participa en la adhesión celular y en la organización del citoesqueleto. La simple posesión de un alelo mutante da lugar a numerosos pólipos benignos (adenomas) en la mucosa del colon. Estos adenomas progresan a adenocarcinomas tras la inactivación del segundo alelo o, como sucede generalmente, la delección del mismo<sup>[49-51]</sup>.

El gen APC es un gen supresor tumoral que codifica para una proteína citoplasmática que regula a la  $\beta$ -catenina de la vía de señal WNT, que tiene un papel fundamental en el control del destino celular, la adhesión y la polaridad celular durante el desarrollo embrionario. APC es un componente de WNT que señala a través de una familia de receptores de superficie celular llamados *frizzled* (FRZ) y estimula varias vías, de las cuales la central implica a la  $\beta$ -catenina y a APC. La función de APC en esta vía consiste en modular negativamente a la  $\beta$ -catenina. En ausencia de señales WNT, APC causa la degradación de  $\beta$ -catenina, evitando su acumulación en el citoplasma. Para ello se forma un complejo macromolecular con  $\beta$ -catenina, axina y GSK3b, el cual conduce a la fosforilación y finalmente ubiquitinación de  $\beta$ -catenina y su consecuente destrucción por el proteasoma<sup>[49]</sup>.

Las señales de WNT bloquean el complejo de destrucción APC-axina-GSK3b, permitiendo que se transloque la  $\beta$ -catenina del citoplasma al núcleo, donde forma un complejo con un factor de transcripción (TF). Este factor de transcripción regula positivamente la proliferación celular mediante el aumento de transcripción de c-MYC, ciclina D1 y otros genes. Si se altera el gen APC, se altera el complejo de destrucción y la  $\beta$ -catenina sobrevive y se transloca al núcleo, donde coopera con dicho TFN. Así, las células con pérdida de APC se comportan como si estuvieran bajo señal continua de WNT. La importancia de esta vía se evidencia por el hecho de que los tumores de colon, aunque no se limita a éstos, que tienen genes APC normales albergan mutaciones de  $\beta$ -catenina que impiden su destrucción por APC, lo que permite que la proteína mutante se acumule en el núcleo y coopere con el TFN<sup>[49,50]</sup>.

A continuación se produce la activación del gen *K-ras* con actividad GTPasa y de *P53*. Por último, el gen DCC (del inglés, *deleted in colorectal carcinoma*) desempeña una función en la

apoptosis, y su delección da lugar a modificaciones en la superficie celular y a mutaciones adicionales que provocan metástasis. Las mutaciones específicas aquí descritas son características de la enfermedad, pero pueden ocurrir en cualquier orden<sup>[49-51]</sup>.

Existen variantes del síndrome PAF, las mejor descritas son las siguientes:

- ✓ Poliposis adenomatosa colónica atenuada (AAPC): se trata de una variante fenotípica de la PAF, en la que los pólipos aparecen en menor número y frecuencia, y si éstos no han sido tratados los pacientes desarrollarán el CCR en edades próximas a los 50 años. Se retrasa una década el CCR en los pacientes en los que los adenomas no han recibido tratamiento<sup>[49]</sup>.
- ✓ Síndrome de Gardner: es una variante de la PAF en la que aparecen tumores benignos en la piel, tejido blando y huesos<sup>[49]</sup>.
- ✓ Poliposis asociada a MUTYH (MAP): es un síndrome autosómico recesivo. Representa el 1% de los casos de CCR<sup>[49]</sup>.

### ***Síndrome hereditario no polipósico (HNPCC) o síndrome de Lynch***

Se trata de un síndrome autosómico dominante caracterizado por la aparición, a una edad temprana, de un cáncer colorrectal que se desarrolla con más frecuencia en el colon derecho, y que exhibe neoplasias metacrónicas y sincrónicas en el colon así como neoplasias extracolónicas como cáncer de endometrio, cáncer gástrico (intestino delgado) cáncer renal y de pelvis y lesiones de piel. En el caso de las mujeres que padecen este síndrome, presentan un elevado riesgo de desarrollar cáncer de endometrio. Las lesiones precursoras aparecen pero en menor número que en los casos de PAF que aparecen cientos de pólipos en la mucosa colónica.

En la mayoría de los casos se produce este desorden por alteraciones que inactivan las proteínas reparadoras del ADN del sistema “mismatch-repair” (MMR). Este sistema se encarga de reparar el daño que se produce tras las divisiones celulares y de corregir los errores que comete la polimerasa durante la replicación celular. La inactivación sucede más frecuentemente en los genes *MLH1* y *MSH2*, pero también en los genes *MSH6* y *PMS2*. Los pólipos suelen aparecer como lesiones precursoras que posteriormente malignizan, si bien su número es muy inferior al de los casos de FAP en los que aparecen cientos de lesiones<sup>[49,51,52]</sup>.

### ***Síndrome de Cowden***

El síndrome de Cowden, también llamado síndrome de hamartomas múltiples es una enfermedad de origen genético que se transmite según un patrón autosómico dominante. Se

caracteriza por la aparición en diferentes órganos de una serie de tumores benignos que se llaman hamartomas. Las principales localizaciones son piel, tiroides, mama, tracto gastrointestinal, cerebro y útero. Este síndrome se produce por mutaciones en las células germinales que afectan al gen *PTEN*. (homólogo de la tensina y fosfatasa) es un gen supresor tumoral que participa en el control del ciclo celular, eliminando grupos fosfatos de una clase de lípidos conocida como fosfoinosítidos (inositol fosfato), que resultan en la inhibición de la vía de señalización Akt/PKB, un mecanismo que potencia factores de crecimiento que promueven la proliferación celular. Las mutaciones en *PTEN* implican una expresión constitutiva de estos factores de crecimiento y la aceleración del ciclo celular, promoviendo la supervivencia celular y la evasión de la apoptosis<sup>[53,54]</sup>.

### ***Síndrome de Peutz-Jeghers***

Es un síndrome con herencia autosómica dominante que se caracteriza por la aparición de lesiones hamartomatosas a lo largo del tracto gastrointestinal. Los portadores presentan un mayor riesgo que otros individuos de desarrollar cáncer tanto en el tracto gastrointestinal, donde estas lesiones precancerosas aparecen en bajo número tendiendo a concentrarse en el intestino delgado, como en órganos sexuales, pulmones o hígado.

El gen asociado a este síndrome (*STK11/LKB1*), codifica para la Serina/Threonina Kinasa 11 (Serine/threonine kinase 11), también conocida como Kinasa Hepática B1 (Liver Kinase B1) y funciona como supresor tumoral. La mutación heredada de alguno de los alelos contribuye a acelerar la carcinogénesis ya que este gen fosforila a otras kinasas incluyendo a la proteína 5'adenosina monofosfato-activada kinasas AMPK que controla la glucosa celular y el metabolismo celular<sup>[55,56]</sup>.

### ***Síndrome de Li-Fraumeni***

Es una enfermedad hereditaria con transmisión autosómica dominante de alta penetrancia que está debida, en el 70% de los casos, a mutaciones germinales en el gen TP53. TP53 Las personas con este síndrome pueden desarrollar múltiples tipos de cáncer y varios tumores entre los que se incluyen cáncer colorrectal, cáncer adrenocortical, tumores cerebrales, sarcomas de tejidos blandos, osteosarcoma, rhabdomyosarcoma, cáncer de mama, leucemia, linfoma y glioblastoma entre otros<sup>[57]</sup>.

### ***Alteraciones genéticas en el cáncer colorrectal esporádico***

En los casos considerados como esporádicos, el CCR es una enfermedad causada también por alteraciones genéticas, siendo las que han sido mejor descritas, una serie de mutaciones puntuales que confieren una mayor o menor susceptibilidad de padecer la enfermedad a la persona portadora. La mayoría de estas alteraciones provocan que el producto del gen cambie la secuencia aminoacídica de la proteína o alteren la expresión o cantidad de proteína final. Por otra parte, diferentes variantes genéticas como SNPs o polimorfismos de inserción-delección se han asociado a una mayor o menor susceptibilidad de desarrollar la enfermedad<sup>[45,46,49]</sup>

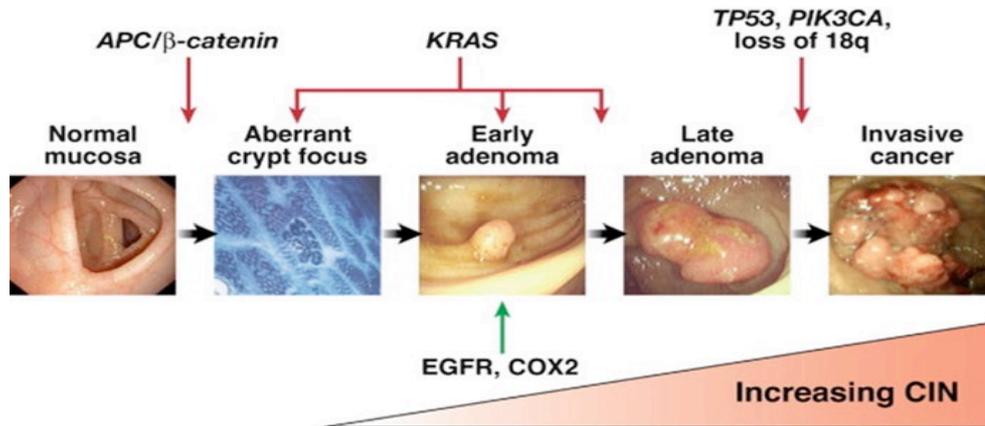
### ***Modelo “adenoma a adenocarcinoma” (modelo de Volgestein)***

En 1990, Fearon y Vogelstein propusieron un modelo genético de tumorigénesis como proceso secuencial, al que designaron “secuencia adenoma-carcinoma”, sugiriendo que la acumulación de mutaciones germinales y somáticas condicionan las características del tumor. Este modelo establece que el CCR es el resultado de mutaciones en genes con importantes funciones, como la regulación de la proliferación celular o la reparación del daño al ADN, que son necesarias mutaciones en más de un gen y que la acumulación secuencial de las mutaciones es importante para determinar la progresión del CCR.<sup>[50]</sup>

Dependiendo del tipo de alteración que ocurra, en el CCR se describieron dos vías principales por las que se producen estos sucesos moleculares durante la progresión tumoral: la vía CIN, o de inestabilidad cromosómica y la vía MSI o de inestabilidad de microsatélites que serán descritas a continuación<sup>[58]</sup>.

### ***Vía de inestabilidad cromosómica: fenotipo CIN***

El fenotipo CIN aparece en el 85% de los tumores esporádicos de colon y en el síndrome familiar PAF, está caracterizado por aneuploidía y múltiples cambios en el cromosoma, las alteraciones genéticas que se producen conllevan sobreexpresión o mutación de genes (oncogenes (KRAS) y genes supresores tumorales (P53, SMAD2, SMAD4)<sup>[45]</sup> que actúan como puntos de control durante el ciclo celular. En consecuencia aparecen defectos en numerosos procesos como la expresión aberrante o la mutación de genes de control de la mitosis, defectos en el microtúbulo y disfunción en el telómero. Los tumores que presentan este fenotipo están relacionados con un pronóstico desfavorable<sup>[45,46,58,59,60]</sup>.



**Figura 5.-** Fenotipo CIN: etapas de la carcinogénesis. El primer paso en la carcinogénesis colorrectal es la formación de criptas aberrantes. La hiperactivación de la vía de señalización de Wnt se produce en este paso con la consecuencia de la inactivación del gen supresor tumoral APC. La progresión de adenoma a carcinoma es inducida por mutaciones activadoras del oncogen KRAS y la pérdida del supresor tumoral TP53. Por último se produce la pérdida del gen DCC, su delección da lugar a modificaciones en la superficie celular y a mutaciones adicionales que provocan metástasis. Imagen reproducida de Pino et al 2010.<sup>[60]</sup>

Los cánceres colorrectales esporádicos que muestran el fenotipo CIN exhiben ganancias o pérdidas de fragmentos grandes de cromosomas de lo que resulta una gran variabilidad en el cariotipo celular. Esto se ha definido como inestabilidad cromosómica y aparece como vía predominante en el CCR esporádico<sup>[61]</sup>. La inestabilidad cromosómica incluye alteraciones estructurales, como reordenamientos cromosómicos alteraciones numéricas, defectos en la segregación de cromosomas o pérdidas de fragmentos cromosómicos grandes. A nivel cromosómico se han encontrado tumores colorrectales con reordenamientos en diferentes genes, identificados tanto a nivel somático como germinal<sup>[62]</sup>.

Entre los mecanismos propuestos por los investigadores que darían lugar a la inestabilidad cromosómica en el CCR podemos reseñar los siguientes:

#### *Defectos en la segregación de los cromosomas*

La falta de segregación conduce a la acumulación de defectos en puntos clave de la mitosis que pueden llevar a la aneuploidía, la cual desestabiliza el genoma, induciendo la acumulación de mutaciones y resultando cariotipos heterogéneos. Este proceso puede aumentar las pérdidas de genes supresores tumorales o resultar en la duplicación de oncogenes<sup>[63]</sup>.

### *Disfunción del centrómero*

Otra de las causas posibles del fenotipo CIN, sería una función anormal del centrómero. En las células de CCR se ha observado centrosomas extra, que llevan a la formación de múltiples polos del huso, resultando una distribución desigual de los cromosomas<sup>[64]</sup>.

### *Disfunción del telómero*

Otro mecanismo propuesto que conduciría al fenotipo CIN en el CCR es la disfunción de los telómeros. Los telómeros son repeticiones hexaméricas de ADN, concretamente de la secuencia TTAGGG en el caso de los humanos que protegen el final de los cromosomas de la fusión y rotura durante la segregación. Un trozo de ADN telomérico se pierde después de cada ronda de replicación del ADN. Las células que presentan unos telómeros lo suficientemente cortos entran en senescencia y apoptosis. Las células activan una proteína; la telomerasa, que es capaz de elongar los telómeros e impedir que la célula muera. En diversos estudios en CCR, se observó que entre el 77-90% de los tumores poseían telómeros cortos pero con la telomerasa sobreexpresada<sup>[65]</sup>.

### *Pérdida de heterocigosidad LOH*

La pérdida de heterocigosidad (LOH) es una de las características o mecanismos que presenta el fenotipo CIN. Aproximadamente entre el 25-30% de los alelos se pierden en los tumores. La no disyunción mitótica, la recombinación entre cromosomas homólogos y la deleción de fragmentos de cromosomas son los mecanismos que provocan la LOH<sup>[61]</sup>.

### *Deficiencias en la respuesta al daño sobre el ADN.*

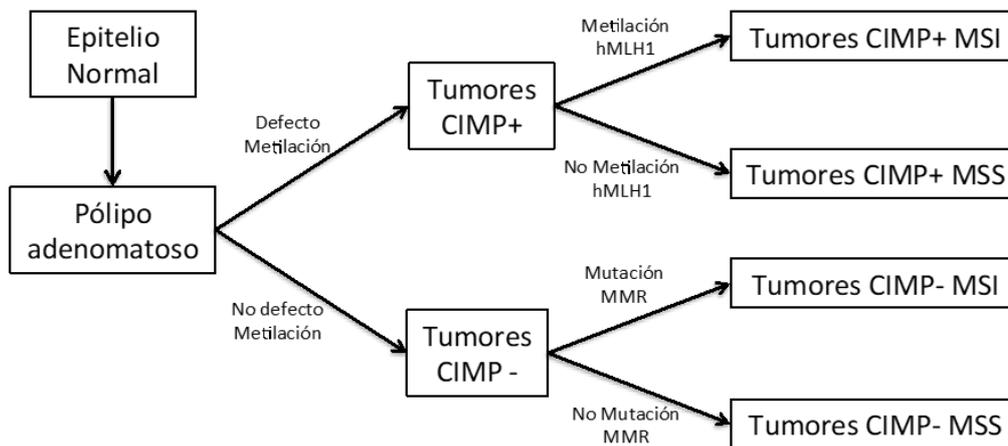
La pérdida de actividad de genes supresores tumorales que responden al daño sobre el ADN, tales como las mutaciones en el gen TP53 o en el gen APC que inactivan su función como supresor tumoral, es otro de los mecanismos característicos de la vía de inestabilidad cromosómica<sup>[66]</sup>.

## ***Vía de inestabilidad de microsatélites en CCR: Fenotipo MSI y Fenotipo CIMP***

El fenotipo MSI aparece en el 15% de los tumores esporádicos y en el síndrome familiar de Lynch (HNPCC)<sup>[45,46,49]</sup>. Los tumores que presentan el fenotipo MSI, son consecuencia de defectos en la maquinaria de reparación de genes *Mismatch repair* MMR relacionadas con un silenciamiento génico (MLH1) y acumulación de mutaciones somáticas en la línea germinal en genes del complejo MMR (MLH1, MSH2, MSH6 y PMS2). Debido a que los microsatélites son secuencias cortas repetidas en tándem, tienen especial tendencia a acumular pequeñas inserciones-deleciones, que pueden ser detectadas como marcadores de este fenotipo. Un

sistema de reparación deficiente está relacionado con la pérdida de función en genes supresores tumorales (TGFB2, IGF2R y PTEN) y sobreexpresión de oncogenes como BRAF. [46,49]

El sistema MMR está controlado fundamentalmente por las proteínas MLH1, MSH2, MSH6 y PMS2. Entre ellas forman complejos heterodiméricos. El más abundante y primero en actuar está formado por MSH2-MSH6. A este complejo se le une posteriormente otro formado por MLH1-PMS2. El complejo final escinde el error del ADN y lo repara. Si ambos alelos de alguno de estos genes se inactivan por cualquier mecanismo ya sea delección, mutación o metilación, los errores de replicación no se pueden reparar y se acumularán mutaciones en microsatélites, produciéndose IMS y desarrollándose el CCR. La inactivación de ambos alelos tiene como consecuencia la pérdida de expresión de la proteína, lo cual puede detectarse mediante inmunohistoquímica<sup>[49,52,57,58]</sup>.



**Figura 6.-** Fenotipo MSI y Fenotipo CIMP. El fenotipo CIMP se presenta en los tumores con inestabilidad de microsatélites cuando se produce la hipermetilación de MLH1. Si no existe metilación del gen y se producen mutaciones en el mismo se trata de tumores con inestabilidad de microsatélites CIMP-.

El fenotipo CIMP aparece en un 75% de los tumores que presentan fenotipo MSI, generalmente por hipermetilación del promotor de MLH1 en tumores con MSI-H, o de alta frecuencia que presentan MSI más de un 40% de los genes estudiados, o del promotor del gen MGMT en tumores con MSI-L, o de baja frecuencia en los que aproximadamente aparece MSI en un 10% de los genes estudiados a su vez correlacionado con la mutación del oncogén KRAS. Los tumores que presentan fenotipo MSI-H están relacionados con un pronóstico favorable. En la figura 6 se simplifica el proceso por el que se produce el fenotipo CIMP en los tumores con inestabilidad de microsatélites MSI<sup>[67,68]</sup>.

Los tumores esporádicos MMR+ se originan debido a inactivación por hipermetilación del promotor del gen MLH1. En estos casos siempre se observa pérdida de expresión inmunohistoquímica de MLH1. La pérdida de expresión de MLH1 generalmente se asocia a pérdida secundaria de PMS2. De la misma forma, la pérdida de MSH2 suele asociarse a pérdida secundaria de MSH6. Las pérdidas aisladas de PMS2 o MSH6, indicativas de mutación en estos genes, son poco frecuentes y sólo deben considerarse tras haber observado expresión de MLH1 y MSH2<sup>[67,68]</sup>.

### ***Uso de múltiples marcadores moleculares en la clasificación del CCR: subtipos moleculares del CCR***

El desarrollo, en los últimos años, de nuevas tecnologías como la transcriptómica, la genómica o la epigenómica, que implican la obtención y análisis masivo de datos mediante secuenciación masiva, está arrojando luz sobre las numerosas alteraciones que suceden durante el proceso de carcinógenesis colorrectal. El uso conjunto en base a modelos matemáticos y analíticos de toda la información obtenida va a permitir mejorar la clasificación del CCR, a efectos del pronóstico y predicción, respecto a las clasificaciones tradicionales descritas hasta la fecha<sup>[69,70]</sup>.

Por ello, a partir de estas premisas se creó el consorcio de subtipificación del CCR “*CRC Subtyping Consortium*”(CRCSC)<sup>[69]</sup> para determinar la presencia o ausencia de patrones utilizando algoritmos basados en los datos de expresión génica del CCR. Uno de los objetivos de este consorcio fue la integración y comparación con todas las fuentes de datos posibles; mutaciones, número de copias, metilación, microARN y proteómica, para establecer los subtipos moleculares del CCR en función de los parámetros relacionados con el seguimiento y la supervivencia de los pacientes. Este proceso permitió discriminar cuatro subtipos moleculares diferentes para el CCR, que fueron validados con datos de expresión procedentes de otras plataformas tecnológicas (Affymetrix, Agilent y RNA-sequencing) demostrando que la tecnología empleada fue robusta. Esta integración de datos permite la translación de los subtipos moleculares en subtipos clínicos, definiéndose los siguientes cuatro subtipos:

#### ***Subtipo 1: CMS1-MSI inmune***

Este subtipo comprende el 14 % de los tumores de CCR e incluye a la mayor parte de aquellos que presentan fenotipo MSI. Son, por tanto, tumores con inestabilidad de microsatélites que, además, presentan alta prevalencia de fenotipo CIMP y baja prevalencia de alteraciones somáticas en el número de copias (SCNA). En este subtipo, las proteínas del sistema de reparación MMR se encuentran sobreexpresadas. Su nombre, MSI inmune, hace

referencia a su relación con la vía inflamatoria y a la presencia de infiltración de células inmunológicas. Esta infiltración inmunológica va acompañada del incremento de la expresión de genes asociados a la respuesta Th1 y a las células T. Por otra parte, en este subtipo el gen BRAF se encuentra mutado con una elevada prevalencia, lo que conlleva una hiperactivación de la vía de señalización de las MAPKinasas. Se caracteriza por el incremento de la expresión de genes asociados a la infiltración inmunitaria principalmente compuesta por TH1 y las células T.

Respecto al comportamiento clínico se ha observado que este subtipo es diagnosticado de forma más frecuente en mujeres, tratándose de tumores con alto grado de diferenciación y localización preferente en el colon proximal. Los tumores con este subtipo, son los que presentan un peor pronóstico, mayor recurrencia y menor supervivencia después del tratamiento, sobre todo en los casos en los que se presentan mutación en BRAF<sup>[69]</sup>.

#### *Subtipo 2: CMS2-Canonical*

El perfil molecular presenta un fenotipo CIN, con inestabilidad cromosómica, con una elevada prevalencia de SCNA con un patrón de ganancia de copias en oncogenes y pérdida de copias en genes supresores tumorales. Se correspondería con una alta prevalencia de la mutación en KRAS y una sobreexpresión de este gen en este perfil molecular.

Los tumores CMS2 presentan sobrerregulación de la actividad de las vías de señalización de *wnt* y MYC, y de sus dianas. Estos tumores se localizan con mayor frecuencia en el colon distal y se asocian a un mejor pronóstico y una mayor supervivencia que en los otros subtipos moleculares<sup>[69]</sup>.

#### *Subtipo 3: CMS3-Metabolic*

El perfil molecular presenta características propias del fenotipo CIN pero también del fenotipo MSI. Se caracteriza por presentar una baja prevalencia de alteraciones somáticas en el número de copias(SCNA), por presentar fenotipo CIMP-low, y un estatus de hipermetilación intermedio, enriquecido por múltiples “firmas” metabólicas que ocurren en línea con la activación constitutiva del gen KRAS, cuyas mutaciones se ha descrito, pueden inducir una adaptación metabólica. Los pacientes con este perfil molecular responden de manera similar a los que presentan el segundo subtipo CMS2 asociándose a un buen pronóstico<sup>[69]</sup>.

#### *Subtipo 4: CMS4: Mesenchymal*

El perfil molecular presenta un fenotipo CIN con una elevada prevalencia de SCNA y con un patrón de oncogenes sobreexpresados y genes supresores tumorales inhibidos similar al

observado en el subtipo2. Los tumores de subtipo CMS4 muestran un claro aumento de la actividad de los genes implicados en la transición epitelio-mesénquima y un aumento de la actividad del factor de crecimiento TGF- $\beta$ , de las vías de señalización implicadas en la angiogénesis, de los mecanismos de la remodelación de la matriz y de los genes implicados en el proceso inflamatorio. Este subtipo tumoral tiende a diagnosticarse en estados avanzados de la enfermedad asociándose a un peor pronóstico y una menor supervivencia global y libre de enfermedad<sup>[69]</sup>.

### ***Epigenética del cáncer colorrectal***

Desde los primeros años de la revolución de la Biología Molecular, la investigación del cáncer se ha centrado principalmente en los cambios genéticos, y este conocimiento ha sido crucial en la comprensión de la patogénesis y de la biología del cáncer. No obstante, no todas las alteraciones implicadas en la tumorigénesis conllevan cambios en la secuencia del ADN. Ya en 1942 Conrad H. Waddington<sup>[71]</sup> acuñó el término “epigenética” para definir las “las interacciones causales entre los genes y sus productos, de los cuales resulta el fenotipo”. Desde el conocimiento actual podemos definir epigenética como “la herencia de la actividad del ADN que no depende de la secuencia estricta del mismo”. Se trata de modificaciones bioquímicas que regulan la expresión de los genes y son heredables, tanto en las mitosis de las células somáticas como en las meiosis de las células germinales.<sup>[67,-71]</sup> Las modificaciones epigenéticas, se han reconocido por la comunidad científica como mecanismos importantes implicados en la formación del cáncer, del mismo modo que las alteraciones genéticas.

La epigenética es uno de los mecanismos que ayuda a explicar las diferencias entre los individuos de una misma especie. Un caso claro de influencia de la epigenética está representado por las diferencias entre los gemelos monozigóticos, que teniendo idéntico genotipo van adquiriendo durante la vida diferencias fenotípicas así como diferencias en la susceptibilidad a enfermedades<sup>[67]</sup>. Una de las características del epigenoma es que no es estático y puede modificarse. Además es diferente en los distintos tejidos y tipos celulares del organismo. Las marcas epigenéticas son reversibles, se nace con un epigenoma que se adquiere durante el desarrollo embrionario, la exposición a lo largo de la vida a diferentes factores ambientales como la dieta, la actividad física o la exposición a tóxicos, modifican las marcas epigenéticas de nuestro genoma, induciendo la aparición de enfermedades como el cáncer<sup>[72,73]</sup>.

El papel de la epigenética en la patogénesis del CCR ofrece una vía alternativa para elucidar los mecanismos del desarrollo del cáncer<sup>[73,74]</sup>. Existen varios niveles de modificación

epigenética, como la metilación del ADN, la modificación de las histonas, la acción de los Complejos remodeladores de la cromatina, y los ARN no codificantes (MicroARN, Long non-coding RNA), que se describen a continuación.

### **Metilación del ADN**

El mecanismo epigenético mejor caracterizado es la metilación del ADN. Dicho mecanismo es crucial para el control de la actividad génica, el mantenimiento de la arquitectura del núcleo de las células e imprescindible para el control de la diferenciación celular<sup>[74,75]</sup>.

Este tipo de modificación epigenética consiste en la adición de grupos metilo sobre el ADN. La metilación de la región promotora de un gen es uno de los mecanismos que pueden silenciar la expresión génica, impidiendo el acceso a factores de transcripción que interactúan con el ADN<sup>[76]</sup>.

El proceso de metilación está mediado por unas enzimas denominadas DNA metil transferasas (DNMTs) que catalizan la adición covalente de un grupo metilo al carbono 5' de la citosina creando 5-metil-citosina. Hay dos tipos de metil transferasas: las metilasas de novo (DNMT3a y DNMT3b) y las de mantenimiento (DNMT1). Estas proteínas requieren como cofactor donador del grupo metilo a la molécula S-adenosil-metionina (SAM), que añade el grupo metilo a los residuos de citosina<sup>[77]</sup>. Las metilasas de novo metilan el ADN en una nueva posición, mientras que las metilasas de mantenimiento ejercen su acción sobre sitios metilados de forma constitutiva para mantener el silenciamiento génico.<sup>[77]</sup>

En humanos la metilación del ADN ocurre en las citosinas que preceden a las guaninas; estos dinucleótidos se denominan CpGs, que suelen estar distribuidos en grupos denominados islas CpGs que se concentran en los extremos 5' de las regiones reguladoras de los genes. La mayoría de las islas CpGs no están metiladas en las células normales, exceptuando en los alelos silenciados por impronta genómica y en el cromosoma X. Mientras las islas CpG de la región promotora del gen no estén metiladas, el gen se encontrará activo y estará expresándose. La mayoría de los lugares CpG mantienen un estado de metilación, especialmente aquellos que se encuentran fuera de la región promotora del gen<sup>[78,79]</sup>.

En el cáncer se modifican estos patrones de metilación apareciendo hipermetilación en las islas CpG de genes supresores tumorales, y pérdida de la metilación en genes silenciados implicados en la proliferación celular. En el CCR existen ejemplos de hipermetilación e hipometilación aberrante.

### *Hipometilación del ADN en el CCR*

La hipometilación del ADN representa una modificación epigenética que se relaciona con el envejecimiento, debido a la adquisición gradual de errores durante la metilación llevada a cabo por metiltransferasas constitutivas o “de mantenimiento” (DNMT1) y que se correlaciona con el daño genómico. La pérdida de la metilación de genes induce errores durante la recombinación y una duplicación de los cromosomas inadecuada<sup>[80]</sup>. La hipometilación del ADN favorece la formación de cáncer al fomentar la inestabilidad cromosómica, la reactivación de trasposones y la pérdida de la impronta genética<sup>[72]</sup>.

En el cáncer se ha observado una hipometilación global del genoma, caracterizada por la depleción amplia y gradual de las bases de citosinas metiladas. Las zonas hipometiladas incluyen regiones codificantes de genes pero también regiones no codificantes, como intrones y secuencias repetitivas de ADN<sup>[81]</sup>.

En el CCR se ha observado este patrón de hipometilación global del genoma, en el que aparecen regiones largas no codificantes LINE-1 de-metiladas asociándose a un aumento de la inestabilidad genómica y una mayor capacidad para penetrar en los tejidos circundantes y metastatizar, un fenómeno que es más habitual en el cáncer colorrectal con fenotipo CIN<sup>[77]</sup>. Otro mecanismo relacionado con el desarrollo y promoción de la carcinogénesis colorrectal atribuido a la hipometilación, es la regulación transcripcional positiva de oncogenes<sup>[82]</sup>.

### *Hipermetilación del ADN en el CCR*

Se ha establecido una asociación entre metilación aberrante de las islas CpG localizadas próximas a la región promotora del gen y una inhibición anormal de la transcripción, con un silenciamiento de genes supresores tumorales, reguladores del ciclo celular y genes reparadores del ADN<sup>[83,84]</sup>.

La hipermetilación aberrante de las islas CpG conlleva el silenciamiento de los genes, normalmente genes supresores tumorales, esto puede darse por acción de las DNMTs y además, a través de la cooperación de las DNMTs con otros complejos represores que modifican las histonas con los que interactúan para provocar el silenciamiento génico y la represión transcripcional de los genes<sup>[85]</sup>. Las metilaciones aberrantes son específicas según el tipo de cáncer, es decir, en el cáncer de mama el perfil de hipermetilación es diferente al de glioma o al de cáncer colorrectal<sup>[72]</sup>.

Se ha demostrado que la hipermetilación del promotor de APC se correlaciona con CCR metastásico y con el fenotipo CIN mientras que la inactivación del promotor de MLH1 una de las proteínas reparadoras del ADN del complejo MMR (mismatch-repair) se ha observado a

menudo con frecuencia en CCR esporádico de fenotipo de inestabilidad de microsatélites o MSI.[Van engeladn y Roy citas arriba] En CCR con fenotipo MSI además se ha observado que la metilación de MLH1 no sólo afecta al tejido tumoral sino que aparece en mucosa normal adyacente al tumor. Además MGMT otra proteína reparadora, aparece silenciada, con su promotor hipermetilado en el CCR, y dicho silenciamiento parece favorecer la aparición de mutaciones en KRAS y BRAF en CCR<sup>[86]</sup>.

Estos datos sugieren que la hipermetilación de los promotores de ciertos genes, incluyendo las proteínas reparadoras del ADN, podría preceder a la formación del tumor y su detección en el epitelio normal de la mucosa del colon podría ser un indicativo de lesiones premalignas que ocurren en etapas tempranas del desarrollo del CCR.

### ***Modificación de las Histonas***

La modificación de las histonas es otro tipo de modificación epigenética e incluye las siguientes modificaciones bioquímicas como la acetilación, metilación, fosforilación, ubiquitinación, sumoilación, ADP ribosilación, desaminación e isomerización de la prolina, de residuos específicos del core de las histonas (H2A, H2B, H3 y H4). Se trata de cambios epigenéticos que regulan la transcripción coordinando la accesibilidad al ADN a través del control del estado estructural de la cromatina. Este balance en la compactación de la cromatina, entre el estado activo, eucromatina, y el inactivo, heterocromatina, va a determinar el acceso de la maquinaria de transcripción al ADN favoreciendo o reprimiendo la expresión génica<sup>[82,87]</sup>.

Estas modificaciones de las histonas tienen una contribución esencial al desarrollo de diversos procesos fisiológicos, pero también son responsables de la aparición de perfiles epigenéticos y genéticos aberrantes, que aparecen en numerosos tipos de tumores sólidos como próstata, mama, pulmón y/o colorrectal, así como en patologías hematológicas<sup>[87]</sup>.

La acetilación y metilación de las histonas son el tipo de modificaciones epigenéticas más estudiadas en el CCR. El residuo que es modificado, la localización en la histona y la categoría de la alteración definen el pronóstico<sup>[84]</sup>.

#### ***Acetilación de las histonas***

La adición post-traslacional de grupos acetilo interfiere en la unión electrostática entre el ADN cargado negativamente y las histonas, mediante la descondensación de la cromatina facilitando la transición de heterocromatina a eucromatina. La acetilación y desacetilación de los residuos de lisina de la cola de las histonas es regulado por las histonas acetiltransferasas

(HATs) las enzimas encargadas de colocar esta marca sobre las histonas y las histonas deacetilasas (HDACs) son las enzimas que se encargan de eliminar estas marcas de activación de las histonas<sup>[87]</sup>.

Las histonas acetiltransferasas (HATs) que al colocar marcas de acetilo promueven la activación transcripcional de los genes, están divididas en dos categorías básicas: (a) HATs tipo A nuclear que acetilan las histonas que forman el nucleosoma y (b) HATs tipo B citoplásmico que acetilan histonas libres<sup>[88]</sup>. Su papel en el CCR es complicado ya que aparecen implicados complejos que tejen redes afectando a varias vías de señalización celular como wnt / $\beta$ -catenina<sup>[89]</sup>.

La familia de histonas deacetilasas (HDACs) que tienen un papel represor, contiene 18 miembros subdivididos en 4 categorías: (a) categoría I (HDAC 1-3, 8), (b) categoría II (HDAC 4-7, 9 y 10), (c) categoría III (Sirtuinas 1-7) y (d) IV (HDAC 11) Las clases I, II y IV comprenden la familia de HDAC clásica, dependientes de Zn, mientras que la clase III es NAD<sup>+</sup>-dependiente<sup>[90]</sup>. Además de con las histonas, las HDACs interactúan con otras proteínas, para mantener la represión transcripcional. Se ha observado sobrerregulación de estas enzimas en CCR<sup>[91]</sup>.

### *Metilación de las histonas*

La metilación de las histonas es un procedimiento dinámico y flexible que ocurre en los residuos de lisina, arginina e histidina, siendo la lisina el residuo más frecuente metilado. Los residuos de lisina pueden estar mono-, di- o tri-metilados, siendo característico de las regiones silenciadas las marcas de tri-metilación como las de la histona H3, en el residuo 9 de lisina o en el 27, H3K9me3 o H3K27me3, junto con una ausencia de acetilación<sup>[84]</sup>. A diferencia de la acetilación, la metilación no provoca cambios en las propiedades electrostáticas de la estructura, sino que está relacionada con el silenciamiento de la región de ADN porque provoca un estado de compactación de la cromatina que no permite el acceso a los factores de transcripción.<sup>[72]</sup>

Las histonas metiltransferasas (HMTs) y las histonas demetilinasas (HDMTs) son las proteínas encargadas de mantener los perfiles específicos de metilación de las histonas. [Wang Y. Degrees make all difference... Epigenetics 2009]. Las histonas metiltransferasas (HMT) se caracterizan por poseer un dominio SET responsable de la función enzimática, mientras que las HDMTs son grandes complejos correpresores.<sup>[92]</sup> Por ejemplo la LSD1 *Lysine specific demethylase 1* desmetila los residuos de metilo de la histona H3, en H3K4me2 y H3K9me. La desmetilación puede resultar patológica al asociarse con varios mecanismos moleculares como la proliferación y diferenciación celular eliminando marcas de silenciamiento de oncogenes.

Además, en concreto esta desmetiltransferasa LSD1 interacciona con otras moléculas como metil transferasas DNMT1 presentando un posible papel en el silenciamiento de p53 promoviendo la tumorigénesis<sup>[93]</sup>.

Por otro lado, las marcas de metilación de histonas que aparecen de forma más frecuente asociadas a la tumorigénesis colorrectal son las marcas de trimetilación de la histona H3, H3K9me3 y H3K27me3 asociadas a la promoción de factores que regulan la expresión el ciclo celular y la angiogénesis<sup>[94]</sup>.

### *Ubiquitinación de las histonas*

La ubiquitinización es un proceso enzimático de modificación proteica post-traducciona (PTM) en el cual el ácido carboxílico de la glicina terminal forma un enlace amida con el grupo amino de la lisina en la proteína modificada. El marcaje de ubiquitina dirige las proteínas al proteosoma, gran complejo de proteínas que degrada y recicla proteínas innecesarias. Las histonas normalmente se encuentran mono-ubiquitinadas. La ubiquitinación puede verse asociada tanto a una activación como a un silenciamiento de los genes que participan en el desarrollo y progresión del cáncer. Este proceso es llevado a cabo por un balance entre proteínas ubiquitinas ligasas y desubiquitinas ligasas. La desregulación de ambas se asocia a tumores sólidos que incluyen el CCR. La pérdida de la monoubiquitinación de la H2B se asocia a tumorigénesis<sup>[95]</sup>.

### *Complejos remodeladores de la cromatina*

Los complejos remodeladores de la cromatina se encargan de mantener un equilibrio dinámico, entre activación y represión, en muchos procesos fisiológicos tales como la proliferación, auto-renovación y diferenciación celular<sup>[96]</sup>. Utilizando la energía derivada del ATP, estos complejos manipulan la arquitectura del nucleosoma, a través tanto de la movilización dinámica por inserción o eliminación de los nucleosomas, como de la configuración nucleosomal del ADN y los octámeros de histonas. Además, los remodeladores de cromatina pueden comportarse como supresores tumorales o promotores, en función del contexto de señales celulares, reclutando HDACs, HDMTs y numerosas proteínas celulares para construir esta arquitectura y participar en una red compleja de factores de transcripción<sup>[96,97]</sup>.

La sobreexpresión o la mutación de algunos miembros importantes de la maquinaria de remodelación de la cromatina han sido identificados en un amplio rango de enfermedades hematológicas y tumores sólidos<sup>[96,97]</sup>.

### ***ARNs no codificantes (Micro ARNs, long non coding RNAs)***

Avances en el conocimiento y en la comprensión de la regulación de la expresión génica que han sido puestos de manifiesto por el consorcio ENCODE (Encyclopedia of DNA Elements Consortium) con el proyecto del transcriptoma humano, que ha revelado que los genes que codifican para proteínas representan menos del 2% del total del genoma y aproximadamente el 80% del genoma se transcribe activamente en ARNs no codificantes (ncRNA) <sup>[98]</sup>.

Aunque se pensaba que estos ARN no codificantes formaban parte del “ruido de background transcripcional” relacionado con “ADN basura”, se han acumulado evidencias que los ARN no codificantes juegan un papel en muchos procesos biológicos que incluyen la regulación de oncogenes y genes supresores tumorales en cáncer<sup>[99]</sup>. Los ARNs no codificantes se diferencian en dos grupos en función de su tamaño:

- Pequeños ARNs no codificantes: con tamaños menores de <200 nucleótidos. Este grupo lo componen los microARNs (miARN), piwi interacting RNAs y pequeños ARNs nucleolares (snoRNAs).
- Long non coding RNAs (lncRNA) (Largos ARN no codificantes) <sup>[100]</sup>.

Estudios recientes han arrojado luz sobre el papel que desempeñan estos ARN no codificantes en particular sobre el papel de los microARN y su contribución biológica a la patogénesis del cáncer.

#### ***MicroARNs***

Los microARNs (miARN) constituyen un mecanismo complementario del control de la expresión génica que implica a una maquinaria que interacciona con la maquinaria epigenética regulándola y siendo a su vez regulada por ella.

Los microARN son secuencias cortas de aproximadamente 22 nucleótidos que no se traducen e hibridan con ARNm silenciado su traducción. Los microARN pueden estar poco expresados en cáncer por hipermetilación de su extremo 5'. Por eso se considera que los microARN pueden tener funciones de “genes supresores tumorales”.

Durante la biogénesis, los miARN premaduros se exportan del núcleo al citoplasma. Se procesan estos miARN premaduros generando miARN maduros, que se unen al extremo 3'-UTR de ARNm, silenciándolos, todo este proceso se ve regulado por el complejo que induce el silenciamiento del ARN o RNA-induced silencing complex (RISC). La unión del micro-ARN al ARNm puede implicar por un lado la degradación del ARNm o la inhibición de su traducción en proteína. Se ha demostrado que su patrón de comportamiento puede ser modulado

dependiendo de los estímulos del micro ambiente tumoral<sup>[101]</sup>. Hasta la fecha, se ha observado que cientos de micro-ARN aparecen desregulados en muchos tipos de cáncer se ha señalado alguno de ellos como relevante en la tumorigénesis colorrectal, regulando la expresión de genes supresores tumorales y oncogenes<sup>[102]</sup>.

#### *Long-non coding RNAs (ARN largos no codificantes)*

Estos ARNs funcionales que no se traducen en proteína, interaccionan con la maquinaria epigenética, regulando la expresión de numerosos genes. En los últimos años se ha observado como los ARN largos no codificantes tienen un papel en procesos fisiopatológicos como el cáncer. En el CCR se ha demostrado como dos de estos ARN no codificantes, lincRNA-p21 y el ARN transcrito antisentido intergénico Hox (HOTAIR), aparecen sobreexpresados en el CCR correlacionándose con estadios avanzados y con enfermedad metastásica<sup>[103]</sup>.

En concreto en esta Tesis Doctoral abordaremos el papel de la modificación de las histonas en la carcinogénesis colorrectal, estudiando los mecanismos represores de la histona metil transferasa EZH2 *Enhancer of zeste homolog 2*, responsable de la actividad del complejo represivo polycomb 2 PRC2. A continuación, en la sección tres de esta introducción, estudiaremos con mayor profundidad los antecedentes de esta proteína y su posible papel como biomarcador del CCR.

### ***Complejos represivos del grupo Polycomb PcG***

Como ya se ha señalado, las modificaciones post-traduccionales que suceden en las colas de las histonas, proteínas fundamentales del nucleosoma, controlan el estado estructural de la cromatina, determinando el acceso de la maquinaria de transcripción al ADN y favoreciendo o reprimiendo la expresión génica y son de naturaleza reversible. Las proteínas del grupo Polycomb (PcG) son importantes reguladores epigenéticos que actúan a través de la modificación de las histonas. Entre sus funciones más importantes se encuentra la regulación del desarrollo embrionario y el control del ciclo celular<sup>[104]</sup>.

Los complejos del grupo *polycomb* fueron analizados inicialmente en *Drosophila* y se describieron como un conjunto de genes implicados en el control de la segmentación del cuerpo en *Drosophila*, a través de la represión del *cluster* de genes *HOX* (*homeobox genes*) determinando de este modo la activación apropiada de genes homeostáticos<sup>[105]</sup>. Estos complejos que participan en el silenciamiento génico se encuentran altamente conservados desde *Drosophila* a humanos incluyendo levaduras, plantas y animales, habiéndose descrito en

todos los mamíferos en que se ha estudiado. Los genes ortólogos al PcG en mamíferos comenzaron a describirse en los años 90, empezando por la identificación de *Bmi1* (Psc en *D.melanogaster*) y su implicación en el desarrollo del cáncer como oncogen cooperante junto a *c-myc* en un modelo murino con linfoma inducido<sup>[106,107]</sup>. Estas observaciones fomentaron el interés por los factores del PcG, lo que llevó a la identificación de numerosos ortólogos en mamíferos.

En la última década, estudios de mapeo genómico de los componentes de PcG en diferentes especies, revelaron como las proteínas PcG se unían a más genes que los genes *HOX*. Los genes identificados comprenden principalmente factores de transcripción implicados en funciones celulares diversas<sup>[108,109]</sup>, incluyendo el mantenimiento de las células madre embrionarias y la diferenciación celular<sup>[108-110]</sup>.

Todos estos estudios caracterizaron el PcG a nivel bioquímico y funcional y determinaron que las proteínas PcG se presentaban como complejos multiproteicos nucleares bien diferenciados. En concreto, existen dos complejos que han sido los más estudiados y mejor caracterizados, que fueron designados como complejos represivos polycomb 1 y 2, PRC1 y PRC2, que participan y controlan el silenciamiento génico mediante modificaciones post-traduccionales de las histonas. Entre sus funciones más importantes se encuentra la regulación del desarrollo embrionario, diferenciación celular, control del ciclo celular e inactivación del cromosoma X.

Las proteínas Polycomb se diferencian de los factores de transcripción en que carecen de capacidad de unión a una secuencia específica del ADN. Los estudios a nivel genómico han demostrado que PRC1 y PRC2 tienden a localizarse en islas CpG de genes que en condiciones fisiológicas están transcripcionalmente inactivos y la actividad de las proteínas del complejo depende del ambiente de la cromatina en mayor medida, pero también es regulada por factores de transcripción<sup>[111]</sup>.

### ***Complejo represivo Polycomb 1: PRC1***

El complejo represivo *polycomb 1*, del grupo PcG es el complejo con mayor número de subunidades y componentes descrito. Se han publicado estudios recientes sobre la existencia de al menos 5 sub-complejos diferenciados bioquímicamente y con funciones biológicas diferenciadas. PRC1 está compuesto por las subunidades BMI1 "*B cell-specific Moloney murine leukemia virus integration site 1*", proteínas RING1 "*Ring finger protein 1*" (RING1A y RING1B también conocidas como RING1 y RNF2 respectivamente), CBX4 "*Chromobox homolog 4*" y PHC "*Polyhomeotic-like protein*". Las proteínas RING1 poseen actividad E3-ligasa y catalizan la

mono-ubiquitinación del residuo de lisina 119 de la histona H2A (H2AK119ub1) conduciendo a la compactación y represión de la cromatina que conlleva el silenciamiento génico<sup>[112]</sup>.

Bmi1 fue la primera proteína del grupo Polycomb que apareció implicada en cáncer, se identificó en un estudio de *screening* de factores que pudieran cooperar con el oncogen *c-myc* en un modelo murino en linfoma<sup>[113]</sup>.

### **Complejo represivo Polycomb 2: PRC2**

El core del complejo represivo *polycomb 2*, PRC2, lo conforman 4 subunidades: EZH2 “*enhancer of zeste homolog 2*”, SUZ12 “*supresor of zeste 2*”, EED “*embryonic ectoderm development*” y RBBP4/7 “*retinoblastoma (Rb)-associated protein 46/48*”. EZH2 es la subunidad catalítica del complejo y presenta un dominio SET con actividad metiltransferasa, SUZ12 y EED son otras dos proteínas que conforman el *core* y son estrictamente necesarias para la actividad enzimática del complejo. Se piensa que RBBP4/7 contribuye a la unión al nucleosoma, así como a otros complejos reguladores de la cromatina<sup>[114,115]</sup>.

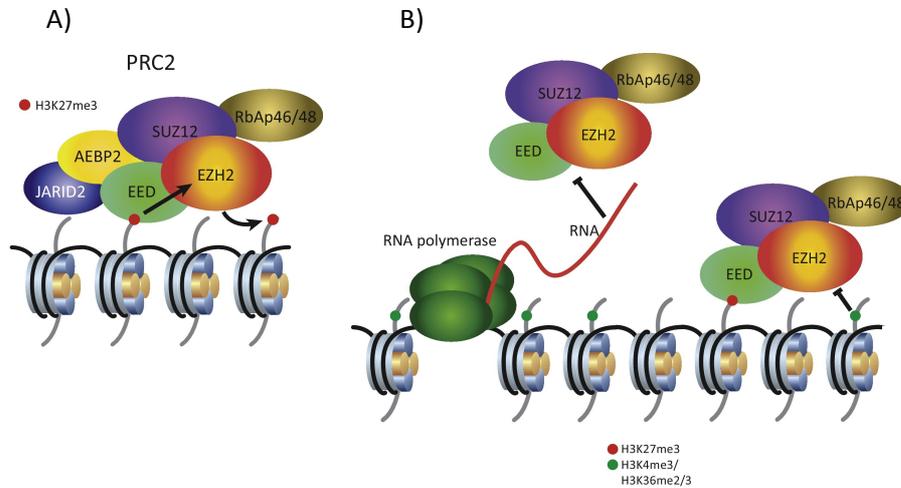
Además de los componentes del core se encuentran una gran variedad de cofactores como JARID2 “*jumonji at rich interactive domain 2*”, factor de transcripción implicado en la formación del tubo neural y el desarrollo del hígado, bazo, timo y el sistema cardiovascular, AEBP2 “*Adipocyte Enhancer-Binding Protein*” factor de transcripción que regula la formación de la cresta neural, y que forman parte del PRC2 de manera temporal y pueden modular la actividad del complejo<sup>[114,115]</sup>.

La función fisiológica de PRC2 en células normales, es el mantenimiento de la represión transcripcional de numerosos genes indispensables para la identidad y diferenciación celular. La actividad metiltransferasa coordinada de las proteínas del complejo PRC2 resulta esencial para la mono-, la di- y la tri-metilación del residuo de lisina 27 de las histona H3 (H3K27me2/3). Estas marcas de metilación de la histona se asocian con la compactación de la cromatina de su estado activado a su estado inactivo en forma de heterocromatina facultativa. La H3K27me3 es una marca asociada a la represión transcripcional y silenciamiento de la expresión de genes reguladores del ciclo celular, senescencia y apoptosis. Mientras la marca H3K27me2 se encuentra ampliamente distribuida, H3K27me3 es más específica<sup>[114-116]</sup>.

Para llevar a cabo la represión transcripcional se han descrito varios mecanismos:

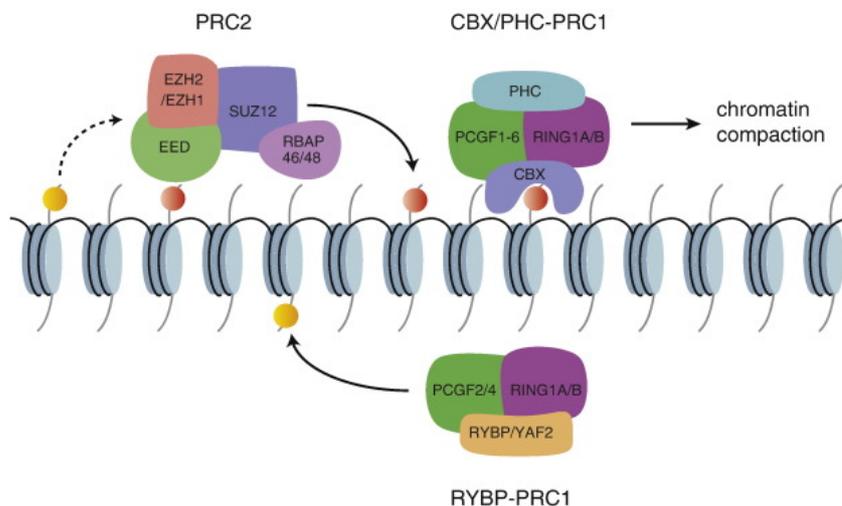
Por un lado se ha planteado que el PRC2 se comporta como una holoenzima en la que todos los componentes actúan de forma conjunta para interactuar con la cromatina de forma secuencial. Los pasos propuestos, que componen este mecanismo secuencial son: (a)

interacción de Jarid2 y AEBP2 con el ADN (b) interacción de RbAp46/48 con las histonas H3 y H4 (c) interacción de EED con H3K27me3 y (d) interacción de las subunidades de PRC2 con ARN largos no codificantes, tal y como se muestran en la figura 7<sup>[114,115]</sup>.



**Figura 7.-** Mecanismo secuencial de acción de PRC2. A) Interacción de Jarid2 y AEBP2 con el ADN y metilación de la H3 B) Regulación de PRC2 por elementos específicos de la cromatina como ARN largos no codificantes. Modificado de Wassef et al 2017<sup>[115]</sup>.

Además la represión transcripcional puede darse por acción conjunta de los dos complejos, PRC1 y PRC2, la interacción de ambos, provocaría en el estado de condensación de la cromatina como se representa en la figura 8<sup>[114,115]</sup>.



**Figura 8.-** Interacción del PRC1 con el PRC2 provocando la condensación de la cromatina. Modificado de Wassef et al 2016<sup>[114]</sup>.

Por otro lado, además de participar junto con los otros componentes del PRC2, EZH2 presenta actividad independiente, comportándose como un regulador de la transcripción<sup>[116,117]</sup>. La desregulación tanto de EZH2 como del complejo PRC2 se ha asociado al desarrollo y progresión de muchos tipos de cáncer<sup>[114-120]</sup>.

## Estructura y función de EZH2

Como hemos señalado en el apartado anterior, EZH2 es la subunidad catalítica del complejo, es una histona metil transferasa que es codificada por el gen *EZH2* que se encuentra en la región 36 del brazo largo del cromosoma 7, en la localización 7q36.1. Se han descrito hasta la fecha 5 transcritos de ARNm y 5 isoformas proteicas codificadas por el gen *EZH2*. Las características y el tamaño de las mismos se detallan a continuación en la tabla 1<sup>[120,121]</sup>.

Transcritos ARNm	UniProtBD	ARNm NCBI	ARNm (nt)	Región codificante	ADNc (nt)	Proteína (aa)
1	15910-2	NM_004456.4	2723	194-2449	2256	751
2	15910-3	NM_152998.2	2591	194-2317	2124	707
3	15910-1	NM_001203247.1	2708	194-2434	2241	746
4	15910-4	NM_001203248.1	2681	194-2407	2212	737
5	15910-5	NM_001203249.1	2682	194-2408	2088	695

**Tabla 1.-** Transcritos e isoformas proteicas del gen *EZH2* descritas y sus respectivos tamaños (nt= nucleótido, aa= aminoácido).<sup>[121]</sup>

La proteína EZH2 presenta un dominio SET conservado en el extremo C terminal cuya función es añadir el grupo metilo CH<sub>3</sub>- a las cadenas de lisina de las histonas. Además del dominio SET, contiene otros dominios funcionales como CXC, dominio rico en cisteína y ncRBD, dominio de unión al ARN no codificante y al ADN, que son imprescindibles para su interacción con los otros miembros del complejo PRC2 y con otras proteínas<sup>[119,120]</sup>.

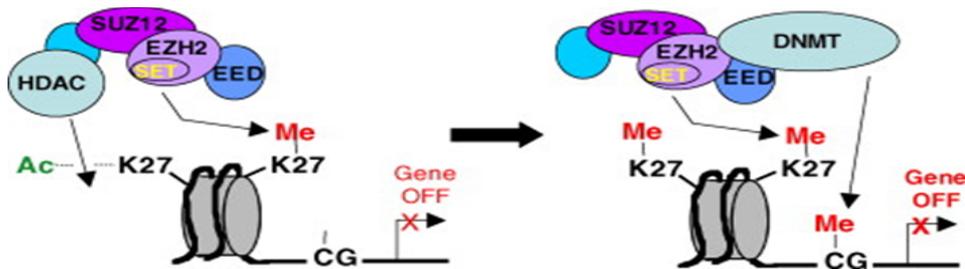
Las metiltransferasas SET dependen para su actividad catalítica de S-Adenosil metionina (SAM) como cofactor que actúa donando grupos metilo. Las proteínas con dominio SET se diferencian de otras metil-transferasas dependientes de SAM en que se unen al sustrato y al cofactor SAM en el lugar opuesto al sitio activo de la enzima. Esta orientación del sustrato y el cofactor permite a SAM separarse del sustrato sin romperse y puede dar lugar a múltiples rondas de metilación de la lisina<sup>[119,120]</sup>.

EZH2 es responsable de la metilación de los residuos de la Lys-9 (H3K9me) y de la Lys-27 (H3K27me) en determinados genes, lo que conlleva la represión transcripcional de los mismos. En el caso del residuo de Lys-27, la metilación puede ser simple (mono) H3K27me<sub>1</sub>, doble (di) H3K27me<sub>2</sub> o triple, (tri) H3K27me<sub>3</sub>. La actividad catalítica sería S-adenosil-L-metionina + L-lisina-[histona] = S-adenosil-L-homocisteína + N(6)-metil-L-lisina-[histona]. [Fuente uniprot]

Por un lado EZH2 actúa cooperando con otros miembros del complejo PRC2, para llevar a cabo la represión transcripcional pero por otro lado presenta actividad independiente de PRC2

como represor de genes concretos metilando directamente sus promotores<sup>[120]</sup> y participando tanto como inductor o activador de la transcripción empleando su segundo dominio CXC e interactuando con otros complejos<sup>[114,115]</sup>.

Además de estas funciones, EZH2 interactúa con otros modificadores epigenéticos, proporcionando una plataforma de reclutamiento de las DNMT transferasas(43) e interacciona con histonas deacetilasas HDAC para eliminar marcas de activación y de esta forma presenta varios sistemas diferentes de represión transcripcional<sup>[120,121]</sup> Este mecanismo de acción aparece representado en la Figura 9.



**Figura 9.-** EZH2 interactúa con otras enzimas como HDAC y DNMT cooperando para llevar a cabo la represión transcripcional de numerosos genes. Reproducido de Simon et al, 2009<sup>[186]</sup>.

Los genes diana cuyas histonas aparecen metiladas por el complejo EZH2/PRC2 incluyen a genes HOX como HOXC8 (Homeobox protein Hox-C8), que participa en el desarrollo del sistema nervioso (cartílago tb),HOXA9 (Homeobox protein Hox-A9) que es un importante regulador de la hematopoiesis y los linajes de los eritrocitos y participa en la diferenciación y formación del sistema circulatorio, MYT1 (Myelin transcription factor 1) gen que codifica para una proteína involucrada en el desarrollo del sistema nervioso, CDKN2A (cyclin-dependent kinase Inhibitor 2A) gen que codifica para dos proteínas p16INK4a y p14arf. En ambas formas proteicas se comporta como un represor tumoral y aparece involucrado, en rutas de señalización y control del ciclo celular, entre otros como los genes diana del ácido retinoico, que también resultan silenciados por EZH2<sup>[122]</sup>.

EZH2 además puede metilar proteínas no histonas como el factor de transcripción GATA4 (Transcription factor GATA-4), una proteína que participa en la embriogénesis y en la diferenciación y función del miocardio, o el receptor nuclear RORA (RAR-related orphan receptor alpha (ROR $\alpha$ ), que participa en la regulación transcripcional de genes circadianos, metilando sus promotores. Los genes candidatos más importantes sobre los que actúa EZH2 que se han descrito hasta la fecha y el efecto que causa EZH2 en cada uno de ellos, se resumen en la tabla 2<sup>[122]</sup>.

Diana	Localización subcelular	Efecto provocado por EZH2	Papel	Referencia
Histona H3	núcleo	Tri-metilación de la H3 en el residuo Lys27 (H3K27me3)	Silenciamiento	[121]
GATA4	núcleo	Metilación GATA4	Silenciamiento	[123]
ROR $\alpha$	núcleo	Metilación dependiente de ubiquitinación de ROR $\alpha$	Silenciamiento	[124]
ER $\alpha$ / $\beta$ -catenina	citoplasma/ núcleo	Interacción con ER $\alpha$ / $\beta$ -catenina para activación de Wnt y receptor de estrógeno	Activación	[125]
PAF	núcleo	Interacción con PAF para activar vía de Wnt	Activación	[126]
AKT	citoplasma/ núcleo	Fosforilación de EZH2 en residuo Ser 21	Activación	[127]
RelA/RelB	citoplasma/ núcleo	Interacción con RelA/RelB para activar NF- $\kappa$ B	Activación	[128]
STAT3	citoplasma/ núcleo	Co-activador de STAT3	Activación	[129]
AR	citoplasma/ núcleo	Co-activador de AR	Activación	[127]

**Tabla 2.-** Posibles dianas de EZH2 y su efecto sobre la expresión génica.

## ***Papel patológico de EZH2***

### ***Mutaciones en EZH2: Síndrome de Weaver***

En humanos, una mutación en el gen EZH2 en la línea germinal se ha asociado con un síndrome autosómico dominante, conocido como Síndrome de Weaver y descrito en 1974. El síndrome lo causa la presencia de una variante patogénica del gen *EZH2* y en algunos casos de otro componente del PRC2, el gen *EED*. Se trata de un trastorno de sobrecrecimiento que cubre un espectro clínico muy amplio: algunos individuos afectados tan sólo presentan estatura elevada mientras otros presentan otra sintomatología más severa, en función de las variantes patogénicas. En el 90% de los individuos con la mutación se ha observado gran estatura y en la mayoría de pacientes afectados ya al nacimiento se ha observado un peso y una talla por encima de lo normal. Los rasgos faciales característicos que pueden observarse en la primera infancia son; hipertelorismo ocular, orejas grandes y retrognatia con el aspecto de mentón "aplastado" asociado a un pliegue cutáneo horizontal. Además alrededor del 80% de los pacientes tiene discapacidad intelectual leve y un reducido número presenta retraso

intelectual moderado con un mayor impacto en la autonomía personal, mientras que los déficits graves son poco frecuentes<sup>[130]</sup>.

Otros rasgos variables incluyen macrocefalia, laxitud articular, escoliosis leve a grave, pectus excavatum, hipo y/o hipertensión, poca coordinación, piel suave, clinodactilia, camptodactilia de dedos de manos y/o pies, hernia umbilical, y llanto ronco de tono bajo en la infancia. Se han descrito dos mutaciones dentro del síndrome, E745K y A682T en las que los individuos con el síndrome desarrollaban enfermedades hematológicas; linfoma y leucemia linfoblástica para cada una de las mutaciones respectivamente<sup>[130]</sup>.

### ***EZH2 y su papel en el cáncer***

El papel de EZH2 en el cáncer es un tema en estudio en la actualidad. En los últimos años se han estudiado *in vitro* los mecanismos por los que participa en el cáncer y se ha valorado sobre pacientes oncológicos los efectos de la expresión de EZH2 o de sus variantes genéticas, como posibles marcadores del pronóstico en el cáncer.

En muchos estudios en tumores se ha encontrado que la expresión de EZH2 aparece elevada en tejido tumoral respecto de tejido sano, asociándose a características clínicas de mal pronóstico en concreto en tumores sólidos de mama y próstata, y en enfermedades hematológicas.

La primera consecuencia funcional que se describió de la sobreexpresión de EZH2 en el cáncer fue el silenciamiento de genes que promueven la diferenciación celular y genes que inhiben el ciclo celular, por lo que el resultado de este silenciamiento sería un aumento de la proliferación y la evasión de la señales de parada o arresto del ciclo celular. Además se asociado la marca de la trimetilación de la H3 (H3K27me3), consecuencia o diana de la actividad de EZH2, como una marca de represión que aparece en genes supresores tumorales implicados en diferentes tipos de cáncer<sup>[114,115,117,118]</sup>.

Se han publicado muchos estudios que profundizan en la relación entre la expresión de EZH2 y el cáncer encontrando asociación entre un incremento en la expresión de EZH2 con estadios más avanzados del cáncer, enfermedad metastásica y menor supervivencia de la enfermedad, reafirmando su papel pro-oncogénico en el cáncer<sup>[132-135]</sup>. Estas características han hecho de EZH2 un buen candidato para ser una diana terapéutica, y en la actualidad se están desarrollando ensayos farmacológicos para valorar el efecto de moléculas capaces de inhibir o bloquear la actividad de EZH2, como por ejemplo el fármaco 3-Deazaneplanocin A que inhibe la expresión de EZH2 e induce la apoptosis de las células tumorales en estudios *in vitro*<sup>[131]</sup>.

Sin embargo, recientemente se han descrito nuevos comportamientos e implicaciones del gen *EZH2* en la tumorigénesis. Se ha demostrado como la pérdida prolongada o inhibición de la expresión de *EZH2* en modelos *in vitro* y modelos animales de cáncer de mama, próstata y glioblastoma, no sólo no previene del desarrollo del cáncer sino que incluso promueve fenotipos más agresivos de la enfermedad<sup>[114,117,118]</sup>. Esto implica que se desconocen todos los mecanismos de actuación del gen en el proceso tumoral. Además se ha observado que *EZH2* se comporta diferente en función del tipo de tumor que se esté analizando, por lo que antes de proponerlo como una diana terapéutica o un marcador de pronóstico, se debe estudiar en más profundidad en el contexto de cada una de las enfermedades en las que puede ser de interés.

### ***Papel de EZH2 en el cancer colorrectal***

En esta línea, se ha observado como el papel de la expresión de *EZH2* en el CCR es controvertido. *EZH2* aparece sobreexpresado en células tumorales del CCR en estudios *in vitro* y en muestras de tejido tumoral de pacientes de CCR<sup>[136,137]</sup>. Sin embargo en los estudios realizados sobre tejido tumoral se observaron dos tendencias contradictorias, por un lado existen trabajos en los que la expresión se asociaba a características de mal pronóstico de la enfermedad como estadios avanzados, enfermedad metastásica y menor supervivencia del CCR<sup>[136,137]</sup> y por otro lado, se han presentado estudios en los que la sobreexpresión de *EZH2* se asociado a características de buen pronóstico y una mayor supervivencia del CCR<sup>[138-141]</sup>. El papel que desempeña la expresión de *EZH2* en el CCR y en la supervivencia de otros cánceres está actualmente en debate. Aunque se han estudiado algunas variantes génicas de *EZH2* existen muy pocos artículos que aborden el papel de estos polimorfismos SNPs siendo insuficientes para predecir el efecto de estas variantes en el pronóstico<sup>[138]</sup>.

El hecho de que se desconocen todos los mecanismos en los que participa *EZH2* en el CCR, y que muchos de los estudios publicados ofrecen resultados contradictorios sobre el papel como marcador predictivo de *EZH2* de la supervivencia de la enfermedad, resulta muy interesante el planteamiento de realizar un estudio que profundice e incremente la comprensión de la relación entre *EZH2* y el pronóstico del CCR. El potencial uso de *EZH2* como biomarcador del CCR podría ayudar a clasificar a pacientes que tuvieran un mayor riesgo de desarrollar CCR y que pudiesen beneficiarse de un seguimiento más de cerca y/o con más intervenciones de tipo preventivo, dada la naturaleza reversible de las modificaciones epigenéticas. Por este motivo en este trabajo trataremos de analizar el papel de *EZH2* en el CCR, siguiendo el planteamiento descrito en el siguiente apartado.



# ***PLANTEAMIENTO Y OBJETIVOS***



## **Planteamiento**

El cáncer colorrectal es el tumor maligno más frecuente en nuestro país y el más frecuente de nuestra provincia lo que supone un problema importante de salud pública. En los últimos años y con la aparición de nuevas técnicas en biología molecular se han descubierto distintos marcadores que han sido incorporados a la práctica clínica, que presentan valor diagnóstico y/o predictivo a los diferentes tratamientos. El hecho, sin embargo, de que la incidencia de ésta enfermedad vaya en aumento genera la necesidad de desarrollar nuevos biomarcadores para el CCR que puedan contribuir a explicar y orientar cómo va a progresar la enfermedad o cómo el paciente va a responder a la terapia.

En los últimos años se ha descubierto que los mecanismos epigenéticos contribuyen a la carcinogénesis de la misma forma que las alteraciones genéticas, pero dada su naturaleza reversible les convierte en potenciales dianas terapéuticas. Una de las modificaciones epigenéticas más relevantes que aparece implicada en el cáncer es la modificación de las histonas, controlando la transcripción génica y el estado estructural de la cromatina.

EZH2 es una histona metil transferasa responsable de la represión transcripcional de numerosos genes. Se ha observado que se relaciona con mal pronóstico y menor supervivencia en otras neoplasias, pero el papel que desempeña en el CCR resulta contradictorio y aparecen muy pocos estudios publicados. Por ello, nos planteamos que sería interesante tratar de dilucidar el papel que desempeña EZH2 en el CCR, su relación con la supervivencia y comprobar en nuestra población su valor como marcador del pronóstico.

Esta Tesis Doctoral se ha desarrollado en el marco del proyecto MCC-Spain, “Estudio multi-caso control de base poblacional de los cánceres más frecuentes”<sup>[142]</sup>, en el que participa el Área de Medicina Preventiva y Salud Pública de la Universidad de León y el Complejo Asistencial Universitario de León (CAULE). La colaboración de los servicios de Anatomía Patológica, Cirugía, Digestivo y Oncología del CAULE ha facilitado el contacto con los pacientes del Área de Salud de León (ASL) que aceptaron colaborar en el estudio, así como la obtención, el procesamiento y el almacenamiento de sus muestras en el biobanco, que pertenece a la Plataforma Nacional de Biobancos, y que se gestiona desde el propio Servicio de Anatomía Patológica.

## **Objetivos**

Se plantea como objetivo general de esta Tesis Doctoral estudiar el papel del gen *EZH2*, su expresión génica y proteica y sus variantes genéticas y en la aparición y supervivencia del cáncer colorrectal, dentro del marco del proyecto MCC-Spain.

Para ello, se establecieron los siguientes objetivos específicos

1. Realizar una revisión sistemática sobre la evidencia previa del papel de *EZH2* en el CCR y un meta-análisis cuantitativo.
2. Realizar una caracterización anatomopatológica y clínico-epidemiológica de las muestras de los pacientes participantes en el proyecto MCC-Spain.
3. Estudiar el efecto de algunas variantes génicas de *EZH2* en la aparición y supervivencia del CCR.
4. Determinar la expresión génica de *EZH2*, a nivel de ARNm y estudiar su asociación con la aparición y supervivencia del CCR.
5. Analizar la expresión proteica de *EZH2* y estudiar su asociación con la aparición y progresión del CCR.
6. Preservar y almacenar el ADN y el ARN de tejido tumoral y tejido sano obtenidos en la presente Tesis Doctoral, con el fin de ser utilizados en futuras líneas de investigación del proyecto MCC-Spain.

# ***DISEÑO EXPERIMENTAL***

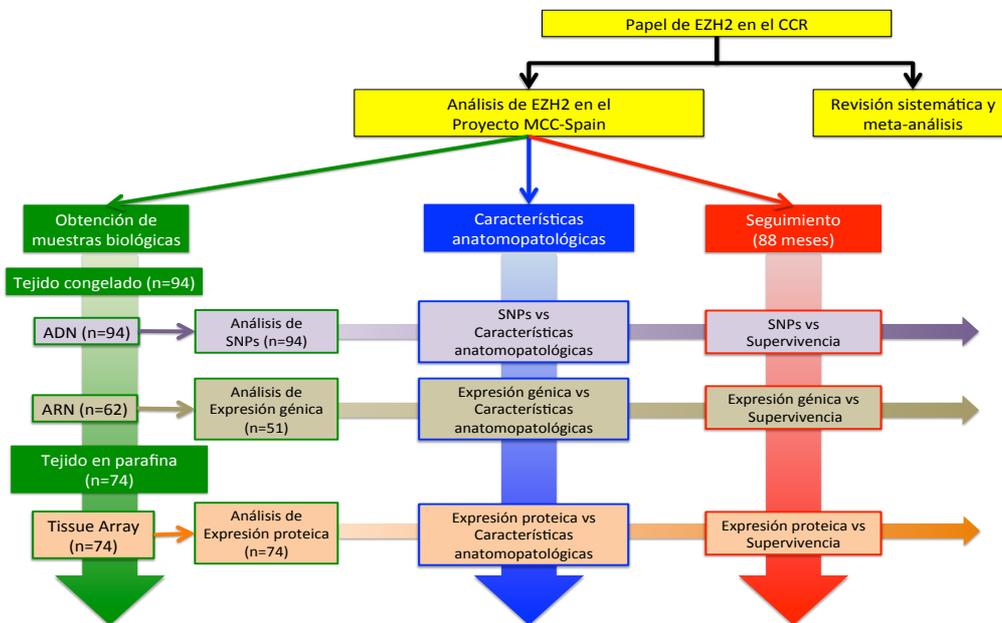


De acuerdo al planteamiento descrito y para abordar los objetivos propuestos se desarrolló el diseño experimental que se explica a continuación y que aparece representado en el esquema de la figura 10.

Por un lado, se realizó una revisión sistemática de la literatura, con objeto de valorar el papel del gen *EZH2* en el pronóstico del CCR; por otro, se estudió la expresión del gen y algunas de sus variantes génicas (SNPs) en una cohorte de casos, en el contexto del proyecto MCC Spain.

El estudio de la cohorte de casos comenzó con la selección de los pacientes de CCR diagnosticados en el CAULE. A partir de sus muestras biológicas de tejido congelado, se extrajeron ácidos nucleicos para los estudios de expresión génica y el genotipado de SNPs, y a partir de tejido en parafina se construyó una matriz de tejido para el análisis de la expresión proteica. Simultáneamente, se obtuvieron datos clínico-patológicos, de manera retrospectiva, a partir de las historias clínicas de los pacientes. Esta información permitió la realización de estudios de asociación con los datos moleculares.

La última fase consistió en la realización de un estudio del seguimiento de los eventos acontecidos durante un periodo máximo de 88 meses desde el diagnóstico y el análisis de la supervivencia de la población en estudio en función del gen *EZH2*.



**Figura 10.-** Algoritmo que representa el diseño experimental y la metodología llevada a cabo en este trabajo. En cada color se representa cada una de las líneas de trabajo seguidas.



# ***METODOLOGÍA***



## **Revisión sistemática de la literatura**

### **Estrategia de búsqueda**

Con el fin de estudiar el papel de EZH2 en la supervivencia del cáncer colorrectal, se realizó una revisión sistemática de la literatura científica. Como variable de exposición principal se consideró la expresión de EZH2 y, como variable de efecto, la supervivencia al CCR. La revisión se realizó de acuerdo a las recomendaciones de la guía REMARKS (*Reporting Recommendations for Tumor Marker Prognostic Studies*) para meta-análisis y revisiones sistemáticas de estudios sobre marcadores de tumores con valor pronóstico<sup>[143]</sup>.

En primer lugar, se realizó una búsqueda de artículos originales en Pubmed, Scopus y WOS (*Web of Science*) publicados hasta Agosto de 2016, usando los términos “EZH2” o “PCR2” y “colorectal” o “colon” y “cancer” o “tumor” o “neoplasm” o “carcinoma”. En una segunda fase, se realizó una búsqueda en racimo, a partir de las citas bibliográficas de los trabajos obtenidos en la búsqueda inicial, con la finalidad de identificar trabajos que no hubiesen sido incluidos previamente.

### **Selección e inclusión de estudios**

Los criterios de inclusión que se definieron para la aceptación de los trabajos obtenidos en la búsqueda bibliográfica fueron los siguientes: (a) los estudios deben estar realizados sobre series de casos, casos y controles o cohortes de pacientes con diagnóstico histológico de CCR; (b) los estudios deben incluir la determinación de la expresión de EZH2 en tejido humano mediante técnica inmunohistoquímica (IHC) o PCR en tiempo real (qPCR); (c) los estudios deben contener suficiente información para la estimación del *Hazard ratio* (HR) relativo a la supervivencia (OS y/o DFS) con un intervalo de confianza del 95%.

Fueron excluidos aquellos artículos basados en estudios “in vitro”/ “ex vivo”, así como cartas, revisiones narrativas, resúmenes de conferencias, o estudios relativos a otras patologías neoplásicas. La búsqueda inicial se realizó sin limitaciones por idioma, pero para la lectura de los trabajos completos y su posterior evaluación, sólo se revisaron artículos publicados en inglés.

### ***Extracción de los datos cualitativos y cuantitativos***

Los artículos fueron revisados de forma independiente por dos investigadores. La información recogida para cada estudio incluyó: nombre del primer autor, nombre de la revista en la que se publicó el estudio, año de publicación, tipo de estudio, número de casos, tipo de análisis de expresión de EZH2, método de detección de la expresión de EZH2 y datos de seguimiento. Los datos seleccionados para el estudio se resumieron y transfirieron a una tabla que se muestra en resultados.

### ***Análisis de la calidad metodológica de los estudios incluidos***

La metodología empleada en los artículos fue evaluada de forma independiente por tres investigadores diferentes de aquellos que seleccionaron las publicaciones. Para cada trabajo, se revisaron los puntos de evaluación considerados en la guía REMARKS, que incluye 4 categorías con un total de 20 ítems, que abordan aspectos clave de la calidad del diseño del estudio, el análisis de los resultados y la reproducibilidad de los mismos. Las 4 categorías que comprende la guía son relativas a la introducción, la metodología, el análisis de los resultados y la discusión (Anexo 4). Además, con el fin de proporcionar una puntuación o “score” a cada estudio, se decidió asignar un valor de “1” o “0”, en función del cumplimiento o no cumplimiento de los criterios de calidad de cada ítem evaluado.

Por otra parte, se realizó una segunda baremación establecida en un trabajo previo (*European Lung Cancer Working Party quality scale for biological prognostic factors for lung cancer*)<sup>[144]</sup> que, si bien se refiere específicamente a cáncer de pulmón, se ha empleado para baremar otros tipos de cáncer incluido el CCR<sup>[145]</sup>. Esta guía incluye cuatro categorías principales con varios ítems por categoría que abordan aspectos de calidad en el diseño, la metodología y el análisis, puntuando cada ítem con 2 puntos si está bien definido, 1 punto si no es claro o está incompleto y 0 puntos si no está definido. Aceptando los criterios de Chen et al<sup>[143]</sup>, se consideraron de alta calidad metodológica los trabajos puntuados por encima del 80%<sup>[140,143]</sup> (Anexo 5).

### ***Análisis estadístico de la revisión sistemática***

Los HR y sus intervalos de confianza se obtuvieron directamente de los artículos cuando fueron reportados por los autores, y cuando no se disponía de dicha información fueron estimados mediante aproximación visual a partir de las curvas de supervivencia de Kaplan-Meier, extrayendo los datos de supervivencia y los puntos en el tiempo de las curvas, asumiendo que la tasa de datos censurados durante el seguimiento fue constante<sup>[146]</sup>.

Para estimar el papel de la sobreexpresión de EZH2 en la supervivencia del CCR, se realizó un meta-análisis utilizando el modelo de efectos aleatorios de DerSimonian y Laird <sup>[147]</sup> y se consideró que  $HR > 1$  implicaría una peor supervivencia asociada a la sobreexpresión de EZH2. Además se evaluó la heterogeneidad de la información ( $I^2$ ) y se utilizó para su estimación el modelo de Mantel Haenszel. La  $I^2$  mide el grado de inconsistencia/incompatibilidad entre los diferentes estudios, y los valores elevados de  $I^2$  indican una elevada heterogeneidad. Se realizó el análisis de sesgos de publicación empleando los métodos de Egger et al <sup>[148]</sup> y Begg et al <sup>[149]</sup>. Para todos los análisis se utilizó el paquete estadístico Stata 13.

## ***Estudio de la cohorte de casos del proyecto MCC-Spain***

### ***Selección de pacientes***

En 2008, el CIBERSP (Centro de Investigación Biomédica en Red Epidemiología y Salud Pública) del Instituto de Salud Carlos III (ISCIII) pone en marcha un estudio multicaso control para investigar la influencia de factores ambientales y su interacción con factores genéticos en tumores muy frecuentes o con características epidemiológicas peculiares en nuestro país, en las que los factores ambientales implicados no son suficientemente conocidos. El proyecto MCC-Spain es, por tanto, un estudio multicaso-control de base poblacional que incluye casos de cáncer colorrectal, cáncer de mama, cáncer gástrico, cáncer de próstata y leucemia linfática crónica (LLC) <sup>[142]</sup>.

La primera fase del proyecto, cuyo objetivo era el reclutamiento de casos y controles, se llevó a cabo entre septiembre de 2008 y diciembre de 2013. Para ello, participaron 23 hospitales ubicados en 12 provincias españolas (Asturias, Barcelona, Cantabria, Girona, Granada, Guipúzcoa, Huelva, León, Madrid, Murcia, Navarra y Valencia). El trabajo, que se realizó en estos centros de manera simultánea, consistió en contactar con casos nuevos de cáncer diagnosticados en el hospital correspondiente, así como controles seleccionados al azar de los registros administrativos de los centros de salud de atención primaria del área de influencia de estos hospitales.

En lo que se refiere al Complejo Asistencial Universitario de León (CAULE), se reclutaron casos de cáncer colorrectal, cáncer gástrico y cáncer de mama diagnosticados en el Hospital Universitario de León y se contó con la participación de controles del Área de Salud de León (ASL). El periodo de captación fue desde el mes de marzo del año 2009 al mes de diciembre del año 2011. El estudio se llevó a cabo siguiendo lo establecido en la Declaración de Helsinki y

fue aprobado por el Comité Ético del CAULE y por el Comité Ético de la Universidad de León. Con la presente Tesis Doctoral se aborda específicamente la serie de casos de CCR reclutados en el contexto de este proyecto.

Los criterios de inclusión definidos en el proyecto MCC para cáncer CCR fueron los siguientes: (a) paciente diagnosticado de cáncer de colon o recto, (b) diagnóstico histológicamente confirmado, (c) edad comprendida entre los 20 y 85 años y (d) residencia dentro del área de influencia del hospital, definida para cada centro participante, como mínimo 6 meses antes de la selección (en el caso del CAULE comprende el Área de Salud de León -ASL-).

Para considerar un diagnóstico de CCR se aceptaron los siguientes tipos de CCR definidos por la "Clasificación Internacional de Enfermedades para Oncología (CIE-O, 2003) que incluye los siguientes tipos; C18: neoplasias malignas de colon, C19: neoplasias malignas de la unión rectosigmoidea , C20: neoplasias malignas del recto y D01.0-D01.2: neoplasias colorrectales *in situ*<sup>[150]</sup>.

En cuanto a los criterios de exclusión, fueron los siguientes: (a) impedimentos para la comunicación por parte del paciente (b) incapacidad funcional física para participar en el estudio (c) diagnóstico previo de cáncer en la misma localización objeto de estudio es decir colon o recto.

### ***Obtención de variables epidemiológicas y clínico-patológicas***

Los pacientes dieron su consentimiento informado para la consulta de su historia clínica y para la realización de un cuestionario sobre hábitos de vida. Además, sus muestras biológicas fueron donadas para fines de investigación y gestionadas a través del biobanco del CAULE, previa firma del consentimiento informado correspondiente. (Anexo 6)

La información contenida en la historia clínica fue recogida a partir de los informes oncológicos, anatomopatológicos y radiológicos de tomografía axial computerizada (TAC), scanner y resonancia magnética (RM). La tabla 3 resume las variables extraídas, que se agrupan en cuatro tipos: (a) variables demográficas (b) variables anatomopatológicas y características propias del tumor (c) variables clínicas o el tipo de tratamiento que ha recibido el paciente y (d) variables de seguimiento, recogiendo a lo largo del período de seguimiento los eventos de recidiva o muerte por enfermedad u otra causa. El tiempo de seguimiento máximo fue de siete años y medio (88 meses), desde la fecha de diagnóstico hasta el 1 de Febrero de 2017.

<b>Variables demográficas</b>	<b>Variables anatomopatológicas</b>	<b>Variables clínicas: tratamiento</b>	<b>Variables de seguimiento</b>
Sexo	Tipo histológico	Cirugía	Aparición de segundos tumores primarios
Edad	Localización tumoral	Radioterapia	Recidiva
	Grado de diferenciación	Quimioterapia	Progresión
	Grado de extensión del tumor		Muerte
	Presencia de metástasis a distancia		Vivo con enfermedad
	Estadio TNM		Vivo sin enfermedad

**Tabla 3.-** Variables obtenidas de la historia clínica de la cohorte de casos

### ***Recodificación de las variables anatomopatológicas***

Las variables recogidas a partir de la historia clínica fueron recodificadas y agrupadas en categorías que facilitasen el análisis estadístico posterior.

En primer lugar, de las variables socio demográficas se recodificó la variable edad en dos grupos: (a) menores de 50 años y (b) mayores de 50 años.

A continuación, se recodificaron las variables anatomopatológicas del siguiente modo:

**Tipo histológico:** Los tipos histológicos más frecuentes del CCR son adenocarcinoma, adenocarcinoma mucinoso, carcinoma de células en anillo de sello, carcinoma escamoso y tumor gastrointestinal de tipo neuroendocrino. Dada la escasa frecuencia de otro tipo histológico distinto de adenocarcinoma en nuestra población, se decidió recodificar esta variable en dos categorías: a) “adenocarcinoma” y b) “otros”, categoría que comprende al adenocarcinomas mucinoso y al carcinoma de células en anillo de sello, agrupados en base al componente coloide y mucinoso que comparten.

**Localización tumoral:** En relación a esta variable, las localizaciones anatómicas de los tumores fueron ciego, colon ascendente, ángulo hepático, colon transverso, ángulo esplénico, colon descendente, sigma, recto-sigma y recto. Se recodificó esta variable en tres categorías: (a) “colon derecho” que agrupa ciego, colon ascendente, ángulo hepático y colon transverso, (b) “colon izquierdo” que comprende ángulo esplénico, colon descendente y sigma y (c) “recto” que incluye recto-sigma y recto.

**Grado de diferenciación:** La diferenciación celular de un tumor es el grado en el que las células cancerosas se asemejan a las células no cancerosas de las que proceden. Así, un tumor

"bien diferenciado" mantendrá semejanzas con las células y los tejidos normales, al contrario que un tumor "pobrementemente diferenciado", que habrá perdido la arquitectura tisular característica. En el caso del CCR, según el grado de diferenciación los tumores pueden ser de tres tipos: bien diferenciados, moderadamente diferenciados y pobremente diferenciados.

Se recodificó esta variable en dos categorías: (a) "tumor de bajo grado", que agrupa los tumores bien y moderadamente diferenciados y (b) "tumor de alto grado", que incluye a los tumores pobremente diferenciados.

**Estadio TNM:** El sistema de estadificación TNM <sup>[16-18]</sup> se basa en tres categorías: la categoría T, evalúa la extensión y penetración del tumor primario dentro de la pared del intestino y si se ha expandido hacia las áreas adyacentes; la categoría N valora si existe afectación de los ganglios linfáticos adyacentes o afectación regional de la enfermedad; por último, la categoría M examina si la enfermedad se ha propagado a otros órganos del cuerpo lo que se conoce como metástasis.

Finalmente, para las variables de seguimiento se obtuvieron dos categorías sobre el estatus del paciente al final del seguimiento: (a) "Exitus" y (b) "Recidiva" tanto local como a un órgano distante.

### ***Obtención de muestras biológicas***

Las muestras biológicas proceden de piezas quirúrgicas recibidas en el Servicio de Anatomía Patológica del CAULE, de acuerdo al procedimiento asistencial rutinario. Para su obtención se aplicó el protocolo habitual del biobanco, ubicado en el propio Servicio de Anatomía Patológica. La legislación vigente (Ley 14/2007 y RD 1761/2011) prioriza el empleo de la muestra para el correcto diagnóstico anatomopatológico de la enfermedad que sufre el paciente. Por tanto, solo se obtuvieron muestras de aquellos casos en los que existía tejido excedente del proceso diagnóstico.

En el Anexo aparece detallado el protocolo de obtención de muestras para el proyecto MCC, referido específicamente a CCR. Las muestras fueron diseccionadas de la pieza quirúrgica bajo la supervisión de un patólogo, y consistieron en tejido tumoral y su homólogo no tumoral. La conservación del tejido se realizó en dos formatos: (a) congelado en criomoldes con un criopreservante (OCT) y mantenidos a -80°C o (b) incluido en bloques de parafina tras su fijación en formol, y almacenados a temperatura ambiente.

### ***Extracción de los ácidos nucleicos***

Tanto el ADN como el ARN fueron extraídos a partir del tejido congelado en criomoldes. Para ello, se realizaron cortes de 20  $\mu\text{m}$  de espesor mediante un criostato. A continuación se detallan las técnicas empleadas en cada caso.

#### ***Extracción del ADN***

Se empleó el sistema comercial “Wizard Genomic DNA Purification kit” (Promega) de acuerdo con las condiciones recomendadas por el fabricante. Entre 5 y 10 cortes de tejido o el equivalente a 20-40  $\mu\text{g}$  fueron introducidos en un tampón de lisis en presencia de proteinasa K. El ADN fue posteriormente purificado mediante un tampón denominado por el fabricante “Protein Precipitation Solution” que se basa en el método “Salting out” o de precipitación de las proteínas por sales. Seguidamente, se precipitó el ADN con isopropanol, se lavó con etanol al 70% y se rehidrató en 100  $\mu\text{l}$  con un tampón “DNA rehydration solution”. Las muestras de ADN fueron congeladas a  $-20^{\circ}\text{C}$  hasta el momento de su utilización.

#### ***Extracción del ARN***

El sistema comercial empleado fue “miRCURY RNA Isolation Kit -Tissue” (EXIQON). El equivalente a 20  $\mu\text{g}$  de tejido (5-6 cortes) fue sometido a la acción de un tampón de lisis con  $\beta$ -mercaptoetanol y proteinasa K. La purificación del ARN se llevó a cabo mediante columnas de afinidad. Tras la aplicación de una solución de lavado o “Wash solution” se eluyó en 50  $\mu\text{l}$  del tampón “Elution Buffer”. Las muestras de ARN fueron alicuotadas y congeladas a  $-80^{\circ}\text{C}$  hasta el momento de su utilización.

### ***Cuantificación y control de calidad de los ácidos nucleicos***

La cantidad y calidad de los ácidos nucleicos extraídos fueron estimadas con la ayuda de un espectrofotómetro (Picodrop Microliter UV/Vis Spectrophotometer), a partir de las densidades ópticas (DO) a 260 nm y el ratio 260 nm/280 nm, respectivamente.

En el caso del ARN, se realizaron controles de calidad adicionales para asegurar la integridad de las muestras. Por un lado, se verificó en gel de agarosa al 0,8% la presencia de las bandas de RNA ribosómico 28S y 18S. Por otro, se analizó la integridad y la pureza de las muestras, mediante electroforesis en el sistema “Experion automated electrophoresis system” (BIORAD)<sup>[151]</sup> disponible en el Laboratorio de Técnicas Instrumentales (LTI) de la Universidad de León. Este sistema integra la separación electroforética de los ácidos nucleicos, la detección de fluorescencia y el análisis de datos. Para ello, emplea un chip microfluídico (LabChip), sobre el cual se aplica la muestra de ARN (1  $\mu\text{l}$ ), que se separa como en una electroforesis

convencional, pero con mayor velocidad y resolución. Un lector de fluorescencia detecta el patrón de bandas correspondiente a las distintas subunidades del ARN, y un software lo representa gráficamente en forma de electroferograma. El sistema proporciona un indicador de la calidad del ARN o “RNA quality indicator” (RQI), un valor de integridad de la muestra, “RNA integrity number” (RIN), los ratios del ARN ribosomal y la concentración a la que se encuentra la muestra.

El valor de RIN nos permitió seleccionar las muestras aptas para la conversión en cADN y el posterior experimento de PCR en tiempo real. Se consideraron válidas todas aquellas muestras que tuvieran una RIN superior a 6.

### ***Análisis de polimorfismos (SNPs) del gen EZH2***

#### ***Estrategia de búsqueda de polimorfismos***

Los polimorfismos del gen EZH2 analizados en este estudio fueron obtenidos a partir de las bases de datos Ensembl (EMBL) <sup>[152,153]</sup> y SNP database (NCBI) <sup>[154,157]</sup> que recogen la información de genotipado de los proyectos internacionales “1000genomes” <sup>[155]</sup> o “HapMap” <sup>[156]</sup>. En función de las posibles consecuencias funcionales e implicaciones clínicas, se seleccionaron tres SNPs rs2302427, rs3757441, rs6958683.

#### ***Genotipado de las muestras***

Para el análisis de SNPs se empleó la metodología “SNP’s Genotyping Assays-Taqman Technologies” mediante un equipo de PCR en tiempo real “StepOnePlus 7200” (Applied Biosystems). Las reacciones de PCR se realizaron con 100ng de ADN extraído de cada tejido tumoral, en un volumen de 25 µl. Se realizaron dos réplicas por cada muestra y como control negativo de amplificación se empleó agua destilada de biología molecular. Las secuencias de los SNPs analizados y las condiciones de PCR empleadas para cada análisis se representan en la tabla 4.

Este sistema comercial permite detectar en una muestra dos variantes de la secuencia, mediante el empleo de sondas marcadas con fluoróforos. En nuestro caso, se emplearon sondas Taqman, oligonucleótidos lineales de secuencia complementaria a la estudiada, marcadas en su extremo 5’ con los fluoróforos VIC o FAM “reporter” y en su extremo 3’ con un fluoróforo apantallador o “quencher” (Tabla 4). El funcionamiento de las sondas Taqman se basa en el fenómeno FRET (*Fluorescence Resonance Energy Transfer*). El “reporter” emite la fluorescencia que se desea cuantificar mientras que el “quencher” captura la fluorescencia del anterior, emitiendo a su vez fluorescencia a una longitud de onda no relevante para el análisis,

de modo que cuando el “reporter” y el “quencher” se encuentran unidos, la fluorescencia del primero no es detectable. Una vez la sonda Taqman ha hibridado—con la secuencia complementaria, la enzima Taq polimerasa hidroliza la sonda, liberando a la molécula “quencher”, lo que permite que la fluorescencia emitida por el “reporter” sea detectable y cuantificable.

El genotipo de las muestras o la discriminación alélica se determinó mediante el programa “StepOne plus”, en función de que la fluorescencia fuese emitida por VIC, por FAM o por ambos. De este modo, las muestras fueron catalogadas como homocigotas para el alelo 1, homocigotas para el alelo 2 o heterocigotas.

Nombre	Secuencia	Fluorocromo	Ciclos PCR
<b>rs2302427 (C/G)</b>	GGATCGTCTCCATCATCATCATCGT[C/G]ATC	VIC-C	95° 10 min X1
	ATCATTATATTGACCAAGGGCA	FAM_G	95° 15 seg 60° 1 min X40
<b>rs3757441 (T/C)</b>	TTGTTATTTTCTCTTTTTGAAATA[C/T]GCTT	VIC-T	95° 10 min X1
	CTTAGAAATTGCTATTTACC	FAM-C	95° 15 seg 60° 90seg X 40
<b>rs6958683 (C/G)</b>	5'GGCATCGAGCAGTGTCTGA 3'	VIC-C	95° 10 min X1
	3' TCCTCTCAGC/GGTAGATT 5'	FAM-G	95° 15 seg 60° 90seg X 40

**Tabla 4.-** Secuencias de las sondas Taqman para la detección de SNPs en el gen EZH2, y ciclos de la reacción de PCR.

### ***Análisis de expresión de EZH2***

En este trabajo se realizaron dos tipos de análisis de expresión de EZH2: por un lado, se estudió su expresión génica mediante experimentos de PCR en tiempo real y, por otro, su expresión proteica mediante la construcción de una matriz de tejido (“tissue array”).

### ***Análisis de expresión génica de EZH2***

Los experimentos de PCR en tiempo real permitieron determinar la expresión génica diferencial en el tejido tumoral con respecto al tejido no tumoral, utilizado como referencia. Además, permitieron cuantificar el nivel de expresión del gen en cada muestra para estudiar posteriormente su asociación con variables clínico-patológicas.

El proceso constó de las siguientes fases:

#### *Transcripción reversa*

El ARN extraído de los tejidos tumoral y no tumoral fue convertido a ADNc mediante transcripción reversa. Con objeto de reducir la variabilidad, se partió de 1µg de ARN total de cada muestra. En este trabajo solo se utilizaron aquellos ARN extraídos que tuvieran una integridad superior a una RIN de 6. El sistema comercial utilizado, “dART 2-Step RT-PCR kit” (EURx), emplea “*Random hexamers*” hexámeros que se unen a todo el ARN al azar e incorpora dNTPs para sintetizar la totalidad de la secuencia complementaria. Según recomendación del fabricante, las reacciones se realizaron en un volumen final de 20 µl, en presencia de dNTPs, DTT, y el enzima “dART reverse transcriptase” (Anexo protocolos). Además, el sistema incluye un inhibidor de RNAsas. El proceso se llevó a cabo en un termociclador modelo PTC200 (MJ Research). El ADNc fue conservado a -20 °C hasta su utilización.

#### *PCR cuantitativa en tiempo real*

Las reacciones de PCR se realizaron mediante el sistema comercial “SG qPCR Master Mix (2x), plus ROX Solution”, compatible con el termociclador “StepOne7200”, en un volumen final de 20 µl, según recomendaciones del fabricante. Este sistema comercial contiene una Taq polimerasa modificada, solución tampón, dNTPs (dUTPs) UNG (*N-uracilo glicosilasa*) para prevenir la contaminación entre reacciones y SYBR Green I, fluoróforo que se asocia a la hendidura menor de la hebra de ADN. La solución ROX que incluye este sistema sirve como referencia para la normalización de la señal de fluorescencia.

Tanto para EZH2 como para GAPDH, empleado como gen normalizador, la amplificación se realizó utilizando *primers* que solapasen con la unión de dos exones a nivel del ADNc. En concreto, el amplicón analizado para del gen EZH2 es un fragmento de 219 pb localizado entre los exones 17 y 19. El correspondiente a GAPDH, es un fragmento de 180 pares de bases entre el exón 1 y el exón 3.

Cada PCR fue optimizada para la obtención de un producto específico y evitando la formación de *primer-dimers*. La temperatura óptima de anillamiento se determinó realizando un gradiente de temperaturas de 52°C a 65°C en una PCR convencional. La tabla 5 refleja las secuencias empleadas y los ciclos de temperatura seguidos.

Las reacciones de PCR en tiempo real se realizaron por triplicado para cada muestra de ADNc y para cada uno de los genes. De acuerdo con las recomendaciones del fabricante, el volumen de las reacciones fue 25 µl, conteniendo 1 µl de ADNc, y una concentración final de 0.4 µM de cada uno de los primers (directo y reverso), 1X del *master mix* proporcionado y 0.5

$\mu\text{M}$  de ROX. Además, se realizó un pretratamiento con UNG durante 2 minutos a  $50^\circ\text{C}$  antes de la desnaturalización inicial. Al final de la PCR se añadió un análisis de fusión o *melting* para comprobar la calidad del amplicón generado, mediante la aplicación de un gradiente de temperaturas de  $95^\circ\text{C}$  a  $50^\circ\text{C}$  ( $-0.3^\circ$  cada 15 segundos)

Gen analizado	Secuencia de los primers	Tamaño del amplicón	Ciclos PCR
EZH2 <sup>[136]</sup>	D: 5'CCGCAAGGGTAACAAAAT3'	219	50°C 2 min x1
	R: 'GGTAGCAGATGTCAAGGGA3'		95°C 10 min x1 95°C 15 s x40
GAPDH	D:5'ATGGGGAAGGTGAAGGTCGG'3	180	55°C 30 s x40
	R: 5'GACGGTGCCATGGAATTTGC'3		72°C 30 s x40

**Tabla 5.-** Secuencias de los primers para la PCR cuantitativa de EZH2 y GAPDH, tamaños de los amplicones y ciclos de la reacción de PCR. D: *primer* directo R: *primer* reverso.

#### Cuantificación de la expresión génica

Para analizar la expresión génica se diseñó un experimento de "Cts comparativos"<sup>[158]</sup>. El valor de Ct o "ciclo *threshold*" es aquel en el cual la fluorescencia de la curva de PCR corta la línea umbral. Los valores de Ct, de acuerdo con este método, se emplearon para cuantificar la expresión relativa de EZH2 en la muestra tumoral con respecto a la muestra no tumoral. Los valores de Ct obtenidos fueron normalizados con los correspondientes a un gen calibrador, en nuestro caso GAPDH.

Previamente, se estimó la eficiencia de cada reacción realizando 5 diluciones seriadas de una muestra de ADNc por triplicado. Como control negativo ("Non template control"-NTC-) se empleó agua. Con los valores de Ct obtenidos, por regresión lineal se construyó una curva y se calculó el valor de la pendiente (m). La eficiencia (E) se obtuvo mediante el cálculo siguiente:

$$E = 10^{-1/m}$$

Conocidas las eficiencias, y dado que fueron similares para las reacciones en estudio, el método de cuantificación relativa de ADN empleado fue el diseñado por Livak et al 2001<sup>[158]</sup>. Este método determina que cuando las eficiencias muestran valores próximos, puede asumirse el valor máximo de  $E = 2$  (lo que equivale a un 100% de eficiencia).

De este modo podremos calcular el incremento relativo (IR), que es el número de transcritos de una reacción con respecto a otra

$$IR = 2^{-\Delta\Delta Ct}$$

dónde:

$$\Delta\Delta Ct = \Delta Ct \text{ muestra tumoral} - \Delta Ct \text{ muestra no tumoral}$$

$$\Delta Ct \text{ muestra tumoral} = Ct \text{ (EZH2) tumoral} - Ct \text{ (GAPDH) tumoral}$$

$$\Delta Ct \text{ muestra no tumoral} = Ct \text{ (EZH2) no tumoral} - Ct \text{ (GAPDH) no tumoral}$$

Se consideró que una muestra tenía sobreexpresión del gen cuando el incremento relativo era igual o superior a 2 y que existía inhibición de la expresión cuando el incremento relativo era inferior a 0.5<sup>[136]</sup>.

### ***Análisis de la expresión proteica***

La expresión proteica de EZH2 fue analizada mediante técnica inmunohistoquímica sobre una matriz de tejido o “tissue array” que incluyó las muestras tumorales de 74 casos de CCR como se describe a continuación.

#### *Construcción de una matriz de tejido (Tissue-array)*

Un patólogo experimentado seleccionó, sobre cortes histológicos de las biopsias en estudio teñidos con hematoxilina-eosina, áreas representativas de los tumores. De estas áreas se tomaron, a partir del bloque de parafina donante, dos cilindros de tejido de 1mm de diámetro, que se introdujeron en el hueco correspondiente practicado sobre un bloque de parafina receptor. El instrumento empleado (Manual Tissue Arrayer, Beecher Instruments, Silver Spring, MD) está dotado con dos agujas huecas que se desplazan en los dos ejes del plano mediante un sistema manual de rosca calibrado. De este modo, una de las agujas extrae el cilindro del bloque donante y la otra del bloque receptor<sup>[159]</sup>. Finalmente, el cilindro del bloque donante se introduce en el hueco practicado en el bloque receptor.

El “tissue array” diseñado, incluyó cada muestra de los tumores por duplicado, y también 4 controles positivos correspondientes a ganglio linfático.

#### *Técnica inmunohistoquímica*

Sobre cortes histológicos de la matriz de tejido (2,5 µm) se realizó la técnica inmunohistoquímica. Para ello se empleó un anticuerpo monoclonal anti-EZH2 de ratón con una dilución 1:100 (AM1836a, ABGENT), realizándose una inmunotinción con peroxidasa.

Tras la visualización al microscopio, se discriminó la positividad de las muestras en base a dos parámetros: intensidad de la tinción del núcleo y porcentaje de células positivas. El umbral definido para considerar sobreexpresión de EZH2 fue la detección de un 20% de células positivas.

### **Análisis estadístico**

En relación con la expresión génica, se valoró si las distribuciones de los datos obtenidos para los  $\Delta Ct$  de las muestras tumorales y no tumorales, así como para el cálculo de  $\Delta\Delta Ct$ , seguían una distribución normal mediante el test de Saphiro-Wilk. Tras ello, se empleó la prueba de rangos con signo de Wilcoxon para valorar si había diferencias estadísticamente significativas en la distribución del  $\Delta Ct$  entre ambos tipos de muestras.

El estudio de asociación de la expresión génica y proteica con las características sociodemográficas y anatomopatológicas, se estudió calculando el valor de OR (*odds ratio*) con sus intervalos de confianza al 95%, en un modelo de regresión logística y se realizaron pruebas de  $X^2$  para valorar la asociación, empleándose el test exacto de Fisher cuando los efectivos esperados en alguna de las categorías eran inferiores a 5. Todas estas pruebas se realizaron con el programa STATA 13, considerándose un error alfa de 0.05.

En el análisis descriptivo de los polimorfismos, se incluye el cálculo de las frecuencias alélicas y genotípicas, así como el cumplimiento del equilibrio Hardy-Weinberg para cada polimorfismo<sup>[160]</sup>. Se analizó la asociación de los polimorfismos con las diferentes características anatomopatológicas empleando los siguientes modelos de herencia: a) codominante que asume que los tres genotipos posibles tienen el mismo peso en la herencia; b) dominante en el que los genotipos homocigotos del alelo de mayor frecuencia se analizan frente a los heterocigotos y a los homocigotos del alelo de menor frecuencia; c) recesivo en el que se agrupan homocigotos del alelo de mayor frecuencia y heterocigotos frente a homocigotos del alelo de menor frecuencia;; d) sobredominante en el que se agrupan todos los homocigotos dominantes frente a heterocigotos; e) *log-additive* que considera el valor de tendencia por cada alelo de menor frecuencia presente. Todos estos análisis se hicieron empleando el programa SNPstats<sup>[160]</sup>.

### **Análisis de la supervivencia de CCR en función de los marcadores moleculares de EZH2.**

Se determinó la relación entre los marcadores moleculares estudiados, los tres SNPs de EZH2, su expresión génica y proteica, con la supervivencia global (OS, *Overall Survival*) y la supervivencia libre de enfermedad (DFS, *Disease Free Survival*).

La tasa de supervivencia global (OS) se refirió empleando evento el fallecimiento del paciente independientemente de la causa, y se empieza a contabilizar el período desde la fecha de diagnóstico de cada paciente. La tasa de supervivencia libre de enfermedad (DFS), que se refirió empleando como evento la aparición de recidiva o el fallecimiento del paciente si este sucede antes, el período de tiempo se considera desde la fecha de diagnóstico.

A partir de dichos datos, se elaboraron curvas de Kaplan-Meier para la representación gráfica de cada uno de los marcadores estudiados en función de las supervivencias OS y DFS. Además, mediante el modelo de regresión de riesgos proporcionales de Cox se calcularon los Hazard Ratio (HR) con su intervalo de confianza del 95%, tanto de manera cruda, como ajustando por edad y estadio recodificado en dos categorías: enfermedad inicial y enfermedad avanzada. Todas estas pruebas se realizaron con el programa STATA 13, considerándose un error alfa de 0.05.

## ***RESULTADOS***



## Resultados revisión sistemática

### Identificación de estudios relevantes/Proceso de búsqueda

El proceso seguido para la identificación de estudios relevantes, usando el método de búsqueda descrito en *Metodología*, aparece representado en el algoritmo de la figura 11. Inicialmente se identificaron 384 estudios, de los cuales se eliminaron 134 trabajos duplicados. Posteriormente, se realizó una primera lectura de los resúmenes con el objeto de identificar si los trabajos cumplían o no los criterios de inclusión. En esta fase, 237 estudios fueron excluidos. De ellos, 77 publicaciones eran revisiones narrativas, cartas al editor y resúmenes de conferencia, 89 artículos reportaban exclusivamente estudios *in vitro* y 3 estudios experimentales se basaban en un modelo murino. Además, se excluyeron 37 artículos por realizar el trabajo en otras patologías (neoplasias malignas diferentes al CCR), 24 por centrar el estudio en otros genes, y 6 por realizar análisis de polimorfismos (variantes genéticas) y no de expresión del gen. Por último, se excluyó un estudio por estar publicado en lengua china. Como resultado, 13 fueron los artículos que cumplieron los criterios de inclusión para los análisis posteriores.

Tras la lectura completa de los 13 estudios seleccionados 3 fueron excluidos, dado que no abordaban la asociación entre la expresión de EZH2 y parámetros de supervivencia, sino el estudio de la expresión de EZH2 en función de otros marcadores (ARN largos no codificantes *Long Non Coding RNA* –HOTAIR y micro ARN)<sup>[161,162]</sup> o el tratamiento con estatinas<sup>[163]</sup>.

Finalmente, 10 estudios fueron los elegibles para la revisión sistemática. Al realizarse la evaluación de la calidad metodológica se observó que las valoraciones variaban entre el 35.7% y el 90.9%, presentando más de la mitad de los artículos una valoración menor del 50%, y que sólo tres trabajos superaban una calidad metodológica del 80%. Los estudios empleaban diferentes medidas de supervivencia, estudiando 8 artículos la supervivencia global OS, *Overall survival*, y 3 la supervivencia libre de enfermedad DFS, *Disease free survival*. De los 10 estudios, 7 identifican la sobreexpresión de EZH2 como un indicador de mal pronóstico para la supervivencia del CCR<sup>[136,137,139,141,164,165,167]</sup> presentando uno de ellos resultados estadísticamente significativos<sup>[136]</sup>. Sin embargo, los 3 estudios<sup>[140,166,168]</sup> restantes encuentran, que la sobreexpresión de EZH2 podría ser un factor de buen pronóstico, mostrando dos de ellos resultados estadísticamente significativos<sup>[168]</sup>.

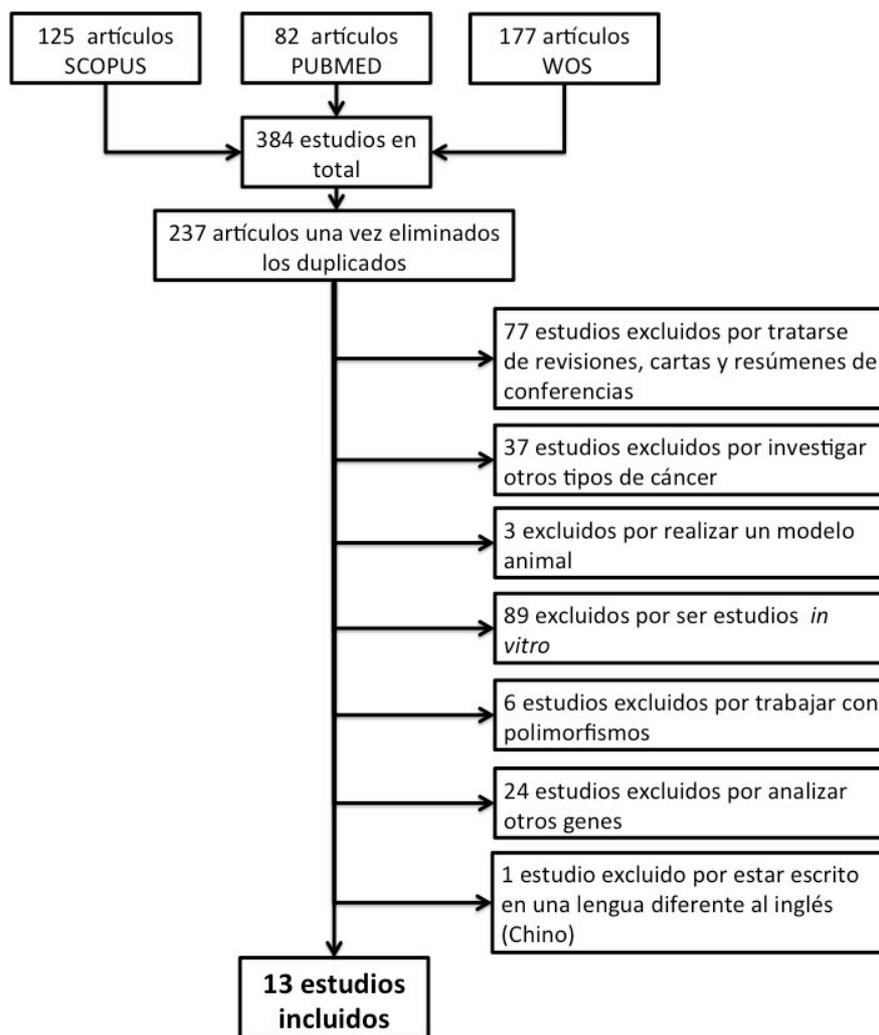
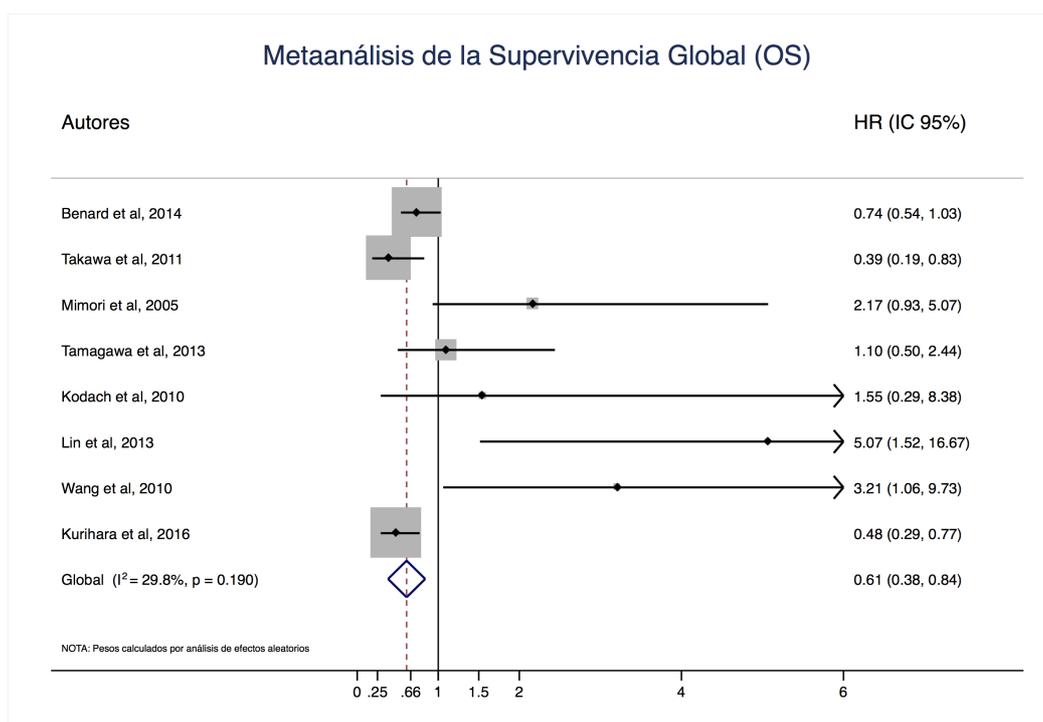


Figura 11.- Algoritmo de la revisión sistemática.

### Meta-análisis

Los 8 trabajos que estudiaban la supervivencia global fueron los incluidos en el meta-análisis. De ellos, 5 mostraban los datos de HR con sus respectivos intervalos de confianza al 95% (IC95%), mientras que en los tres restantes este valor se estimó a partir de los gráficos que representaban las curvas de supervivencia Kaplan-Meier. El valor de HR global que se obtuvo fue HR=0.61 con IC95%(0.38-0.84). Estos datos sugieren que se podría considerar la sobreexpresión de EZH2 en el CCR como un factor de buen pronóstico en el estudio de la supervivencia como se ve representado en la figura 12.

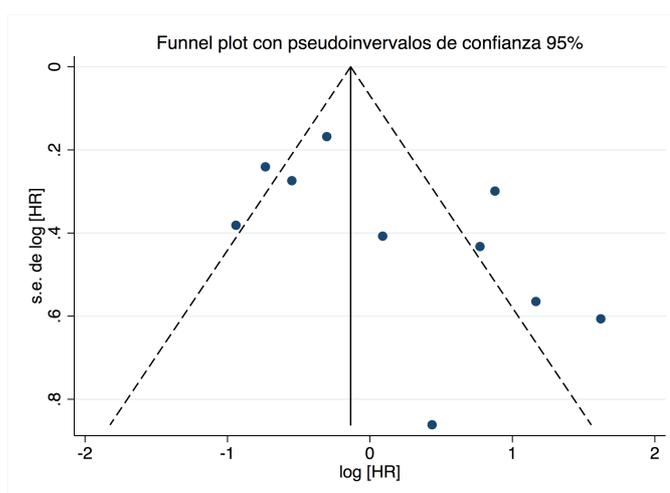
Respecto a la heterogeneidad de los estudios, el valor de  $I^2$  fue de 29.8%, sin significación estadística ( $p=0.190$ ). Esto indica que los estudios son lo suficientemente homogéneos para estudiarlos en conjunto. Además, se realizó un meta-análisis adicional agrupando los siete estudios que emplean técnica inmunohistoquímica (IHC) y analizando este sub-grupo se obtuvo un valor de HR=0.58 con IC95%(0.38-0.79).



**Figura 12.-** Representación gráfica del meta-análisis.

### ***Análisis de sesgos de publicación***

Para los 10 artículos seleccionados se realizó un análisis de sesgos de publicación empleando el test de Egger y el test de Begg. Aunque el *funnel-plot* (Figura 13) mostró alguna asimetría, ni el test de Egger ( $p=0.121$ ) ni el de Begg ( $p=0.072$ ) ofrecieron resultados significativos, lo que sugiere que no existen sesgos de publicación para la sobreexpresión de EZH2 en la supervivencia del CCR.



**Figura 13.-** *Funnel Plot* realizado para comprobar la existencia de sesgos de publicación. Los ocho estudios que investigan el papel de EZH2 en la supervivencia del CCR aparecen representados con puntos muestran alguna asimetría.

Nombre del primer autor	Año de publicación	Tipo de estudio	Método de detección	Número de casos	Medida de supervivencia	Puntuación de la calidad metodológica	Inclusión en el estudio
Kurihara H	2016	Estudio de cohorte	IHC	310	OS	90.90	SI
Liu YL	2014	Estudio de casos	qPCR	82	DFS	40.90	SI
Benard A	2014	Estudio de cohorte	IHC	247	OS/DFS/RFS	90.90	SI
Jinushi T	2014	Estudio de casos	-----	71	-----	-----	NO
Ishikawa S	2013	Estudio de casos	-----	742	-----	-----	NO
Tamagawa H	2013	Estudio de casos	IHC	61	OS	35.71	SI
Lin Y	2013	Estudio de casos	IHC	129	OS	43.18	SI
Takawa M	2011	Estudio de casos	IHC	172	OS	68.18	SI
Kogo R	2011	Estudio de casos	-----	100	-----	-----	NO
Kodach LL	2010	Estudio de casos	IHC	72	OS	40.90	SI
Wang CG	2010	Estudio de casos	IHC	119	DFS	63.63	SI
Fluge	2009	Estudio de cohort	IHC	409	RFS	84.09	SI
Mimori K	2005	Estudio de casos	q-PCR	61	OS	35.71	SI

**Tabla 6.-** Características de los estudios elegibles para la revisión sistemática.

Autor	Técnica	Punto de corte de sobreexpresión de EZH2	N	T (años)	OS_C IC 95%	OS_A IC 95%	DFS_C/RFS_C IC 95%	DFS_A/RFS_A IC 95%
Kurihara et al, 2016	IHC	- → < 80% Media positividad nuclear + → ≥80% Media positividad nuclear	301	4.4	<b>0.48</b> (0.29-0.77)			
Benard et al, 2014	IHC	- → ≤ mediana % de positividad nuclear + → > ≤ mediana % de positividad nuclear	247	8.6	<b>0.74</b> (0.54-1.03)	<b>0.84</b> (0.60-1.18)	<b>0.64</b> (0.42-0.99)	<b>0.67</b> (0.43-1.05)
Liu et al, 2014	q-PCR	- → ≤ 2 tumor/normal ratio + → > 2 tumor/normal ratio	82	3	---	---	<b>4.18</b> (2.08-8.36)	<b>2.52</b> (1.10 -5.73)
Tamagawa et al, 2013	IHC	- → ≤ mediana del score H EZH2 (0-300) + → > mediana del score H EZH2 (0-300)	54	8	<b>1.09</b> (0.50-2.44)	---	---	---
Lin et al, 2013	IHC	Extension X Intensidad de la tinción: - → Grados 0-3 / + → Grados 4-12	33	5.8	<b>5.07</b> (1.51-16.7)	---	---	---
Takawa et al, 2011	IHC	- → Sin tinción en células tumorales. + → Tinción marrón en el núcleo	172	5	<b>0.31</b> (0.18-0.82)	<b>0.42</b> (0.18-0.97)	---	---
Kodach et al, 2010	IHC	- → <70% células tumorales positivas + → ≥70% células tumorales positivas	72	6.6	<b>1.54</b> (0.29-8.38)	---	---	---
Wang et al, 2010	IHC	- → ≤ 30% células tumorales positivas + → > 30% células tumorales positivas	119	6.6	---	<b>3.21</b> (1.06-9.72)	---	---
Fluge et al, 2009	IHC	Extension X Intensidad de la tinción: - → Grados 0-3 / + → Grados 4-12	409	7	---	---	<b>1.17</b> (0.46-2.98)	---
Mimori et al, 2005	q-PCR	→ mediana ratio tumor/normal	61	8	<b>2.17</b> (0.93-5.07)	---	---	---

**Tabla 7.-** Valor pronóstico de la sobreexpresión de EZH2 en la OS del CCR. OS\_C= Tasa de supervivencia global, análisis crudo; OS\_A= Tasa de supervivencia global, análisis ajustado; DFS\_C= Tasa de supervivencia libre de enfermedad, análisis crudo; DFS\_A=Tasa de supervivencia libre de enfermedad, análisis ajustado; RFS\_C= Tasa de supervivencia libre de recurrencia, análisis crudo; RFS\_A= Tasa de supervivencia libre de recurrencia.

## ***Estudio de la cohorte de casos del proyecto MCC-Spain.***

### ***Análisis descriptivo de las características sociodemográficas y clínico-patológicas***

La tabla 8 muestra los datos referentes a las características sociodemográficas y clínico patológicas de los 94 pacientes de CCR que participaron en el estudio.

Dichos pacientes presentaban un rango de edad en el momento del diagnóstico comprendido entre los 46 y los 85 años y una media de edad de 71.73 años, con una desviación estándar de 11.05 años. El 67.71 % (63) de los tumores colorrectales analizados fueron diagnosticados en hombres mientras que el 32.29 % (31) fueron diagnosticados en mujeres, por lo que el *ratio* de la frecuencia en cuanto al sexo fue aproximadamente de 2:1. La media de edad para el diagnóstico de CCR en hombres fue de 72.66 años con desviación estándar de 7.60 y un rango comprendido entre 50.53 y 85.34 y en las mujeres la media de edad es de 69.79 con una desviación estándar de 16.07 y un rango entre los 46 y los 84.13 años en el momento del diagnóstico de la enfermedad.

En cuanto a las características clínico-patológicas, en primer lugar se identificó la localización anatómica del tumor. La localización más frecuente fue en sigma con 32 casos (34.04%). El resto de los tumores se localizaron 15 casos en ciego con (15.96%), 13 casos en colon ascendente (13.83%), 8 casos en ángulo hepático (8.51%), 3 casos en colon transversal (3.19%), 3 casos en ángulo esplénico (3.19%), 6 casos en colon descendente (6.38%), 6 casos en recto-sigma (6.38%) y, por último, 8 casos en recto (8.51%).

Según la recodificación en tres categorías se observó que la localización más frecuente fue en el colon izquierdo el 43,62%(41) de los casos y la menos frecuente en recto con el 14.89%(14). En el 41.49%(39) de los casos la localización fue en el colon derecho. Al recodificar en dos categorías se vió que el mayor porcentaje de casos de CCR aparecían en el colon 85.11% (80) frente a un 14.89% (14) que se desarrollaban en el recto.

El diagnóstico más frecuente fueron los adenocarcinomas, que comprendieron el 82.98% (78) de los tumores colorrectales analizados, el 13.82% fueron adenocarcinomas mucinosos (13), 1 (1.06%) caso se correspondió con diagnóstico de carcinoma de células en anillo de sello y los 2 (2.12%) restantes se corresponden con un diagnóstico de carcinoma y de carcinoma in situ respectivamente. De modo global y para mejorar la potencia estadística se consideraron 2 categorías para los análisis posteriores, adenocarcinoma con 78 casos, y otros tipos con un total de 16 casos.

VARIABLE	CATEGORÍAS	N (%)
Edad	<50 años	3(3.29)
	>50 años	91(96.71)
Género	Hombres	63(67.02)
	Mujeres	31(32.98)
Tipo histológico	Adenocarcinoma	78(82.98)
	Otros tipos	16(17.02)
Localización tumoral(x3)	Recto	14(14.89)
	Colon izquierdo	41(43.62)
	Colon derecho	39(41.49)
Localización tumoral(x2)	Recto	14(14.89)
	Colon	80(85.11)
Grado de diferenciación(x3)	Bien diferenciado	24(25.53)
	Moderadamente diferenciado	40(42.55)
	Pobrememente diferenciado	21(24.70)
Grado diferenciación(x2)	Bajo grado	64 (75.29)
	Alto grado	21(24.70)
pT	pT1	3(3.19)
	pT2	8(8.51)
	pT3	58(61.70)
	pT4	21(22.34)
pN	pN0	50(53.19)
	pN1	32(34.04)
	pN2	11(11.70)
Metástasis a distancia	M0	82(87.23)
	M1	12(12.77)
Estadio TNM	Estadio I	8(8.51)
	Estadio II	39 (41.49 )
	Estadio III	34(36.17)
	Estadio IV	12(12.77)
Estadio TNM(2x)	Estadios iniciales (I y II)	48(51.06)
	Estadios avanzados (III y IV)	46(48.94)
Segundos tumores primarios	Si	23 (24.47)

**Tabla 8.-** Frecuencias de las características clínico-patológicas de la cohorte de casos.

Por otro lado, según el grado de diferenciación de los tumores, en este estudio 24 casos (25.53%) presentaron tumores bien diferenciados o Grado I (bajo grado), 40 casos (42.55%), el grado moderadamente diferenciados o Grado II (bajo grado) y 21 casos (22.34%) que se

corresponden con tumores pobremente diferenciados o Grado III (alto grado). En 11 pacientes (11.70%) no se pudo determinar el grado de diferenciación. Se realizó una recodificación de esta categoría agrupando a los tumores bien y moderadamente diferenciados que se corresponderían con tumores de bajo grado, con un total de 64 casos (75.29%) y en tumores de alto grado, que se corresponde con los casos pobremente diferenciados 21(24.70%).

Finalmente se implementó la información de la anatomía patológica con el diagnóstico de extensión del tumor pT, la afectación de los ganglios pN y la presencia de metástasis a distancia M, para su posterior clasificación y estadificación de acuerdo a la clasificación TNM.

Según el grado de penetración y extensión del tumor T, en nuestra cohorte de pacientes hemos observado 1 caso (1.06%) donde el tumor es in situ Tis, (T0), en el que el tumor se encuentra en su etapa más temprana y no ha penetrado la capa interna (mucosa) del colon o del recto. Se han observado 2 casos (2.13%) en los que el cáncer ha crecido a través de la capa muscular de la mucosa (*muscularis mucosae*) hasta la submucosa (T1). En la categoría T2, el cáncer ha penetrado hasta la *muscularis propria*, se han detectado 8 casos (8.51%). La categoría T más frecuente es la T3, el tumor ha crecido hacia las capas más externas del colon o del recto, pero no ha atravesado estas capas con 58 casos (61.70%), en la categoría T4, el tumor ha atravesado las capas más externas del colon y recto, tenemos 21 casos (22.34%), y en 1 caso (1.06) no se ha podido determinar la T.

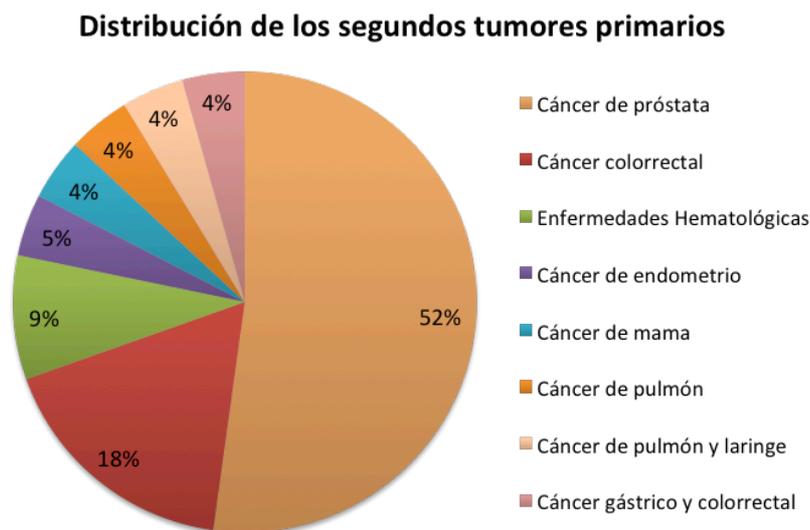
En cuanto a la afectación regional de los ganglios linfáticos, del total de pacientes 43 (45.74%) tenían afectación de los ganglios linfáticos; 32 casos (34.04%) existía propagación a menos de 3 ganglios linfáticos (N1a o N1b) o a la grasa adyacente a los mismos, sin afectación de los ganglios en sí (N1c) y en 11 casos (11.70%) a más de tres ganglios linfáticos (N2a y N2b). En 50 casos (53.19%) no existe afectación regional de los ganglios, y en 1 caso (1.06%) no se ha podido determinar.

Se complementó el estudio con información de la presencia o ausencia de metástasis a distancia a partir de imágenes radiológicas (no determinado en anatomía patológica) El 12.77% de los casos (12) presentan metástasis a distancia en el momento del diagnóstico, mientras que el 87.23% (82 casos) restante no.

Atendiendo a estos 3 criterios, la clasificación TNM mostró que los estadios más frecuentes en el momento del diagnóstico son los estadios II y III con 39 casos (41.49%) y 34 casos (36.17%), se observó 1 caso (1.06%) con estadio 0, 8 casos (8.51%) con Estadio I, y 12 casos (12.77%) con estadio IV. Se decidió para posteriores análisis recodificar esta categoría en estadio inicial de la enfermedad, que agrupó a los estadios 0, I y II, que representa el 51.06%

de los casos (48) y en estadio avanzado de la enfermedad que agrupa a los estadio III y IV, con una representación del 48.94% del total de los casos.

En estos pacientes se detectó durante el periodo de seguimiento la aparición de al menos un segundo tumor primario en un total de 23 pacientes (24.5 %). La distribución de los distintos tipos de segundos tumores primarios se muestra en la figura 14.

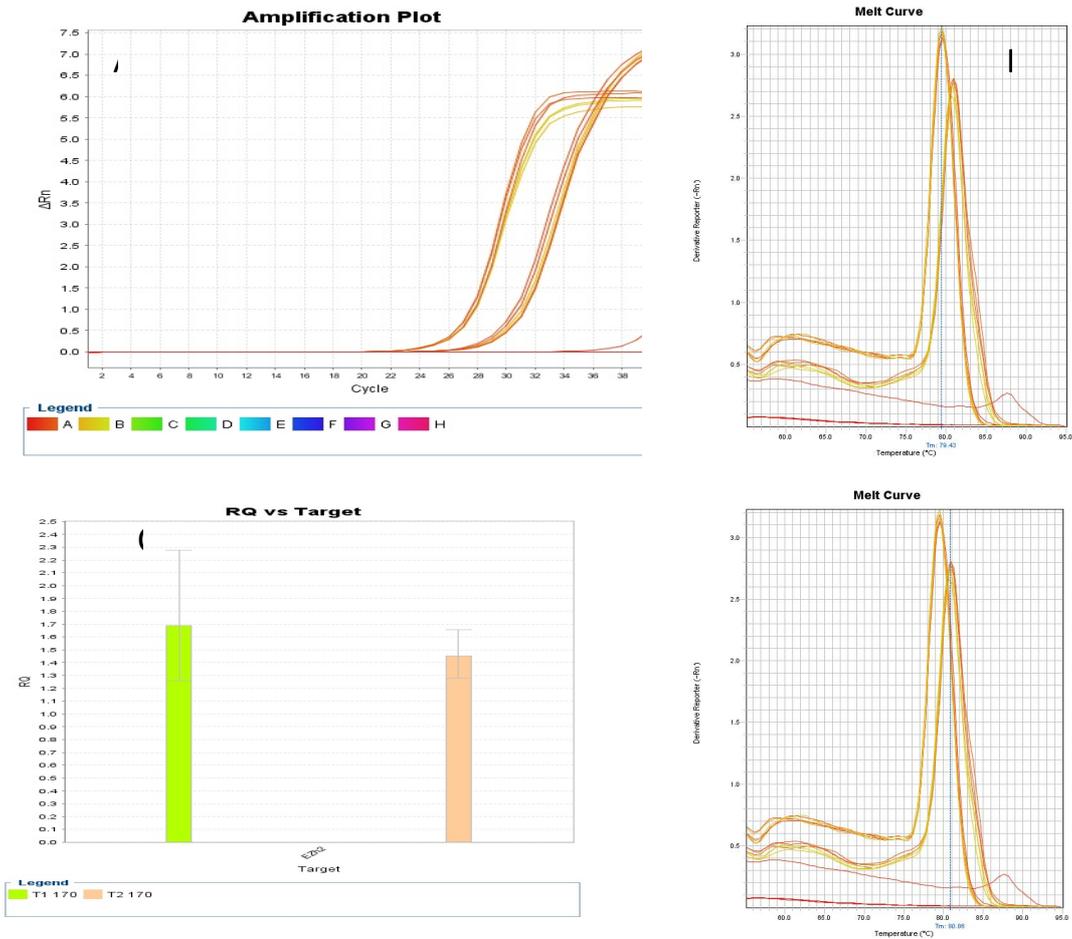


**Figura 14.-** Distribución de segundos tumores primarios localizados durante el periodo de seguimiento.

## ***Análisis descriptivo de los marcadores moleculares***

### ***Análisis de la expresión génica EZH2***

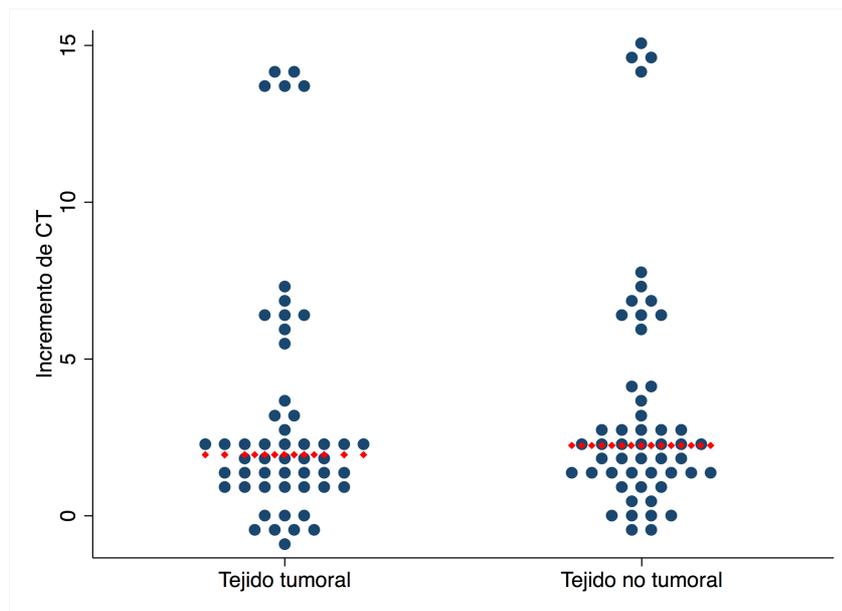
El análisis de la expresión génica diferencial para EZH2 se realizó a partir de ARN extraído de 51 muestras pareadas, tumoral-no tumoral seleccionadas de forma aleatoria de los 94 casos de CCR. Como ejemplo del procedimiento seguido, la figura 15 representa los resultados obtenidos para una muestra tumoral y su homóloga no tumoral. Los valores de Ct para EZH2 y para el gen normalizador GAPDH fueron estimados a partir del punto de corte de la línea umbral con la curva de PCR. La expresión diferencial entre muestras tumorales y no tumorales se obtuvo a partir del valor de  $\Delta\Delta Ct$ .



**Figura 15.-** Ejemplo de expresión diferencial de una muestra tumoral con respecto a su pareada no tumoral para los genes EZH2 (amarillo) y GAPDH (rojo). (A). Curvas de amplificación, realizadas por triplicado. (B). Análisis de fusión de los productos de PCR para EZH2 ( $T_m = 79.43$ ) y GAPDH ( $T_m = 80.86$ ). (C) Expresión diferencial de ambos genes en las muestras pareadas.

El test de Shapiro-Wilk para el estudio de la distribución de las variables determinó que tanto los valores de Ct, como los de  $\Delta\Delta Ct$  seguían una distribución no normal ( $Z$  Shapiro-Wilk=7.190 y  $p = 0.0000$ ). Consecuentemente, la expresión diferencial se valoró mediante el test de Wilcoxon obteniéndose un valor de  $z = -3.253$  y  $p = 0.0011$  indicando que existe expresión diferencial entre el tejido tumoral y el tejido no tumoral pareado. Se ha realizado una representación gráfica (dot-plot) del valor de Ct de las muestras tumorales y de las no tumorales (Fig. 16).

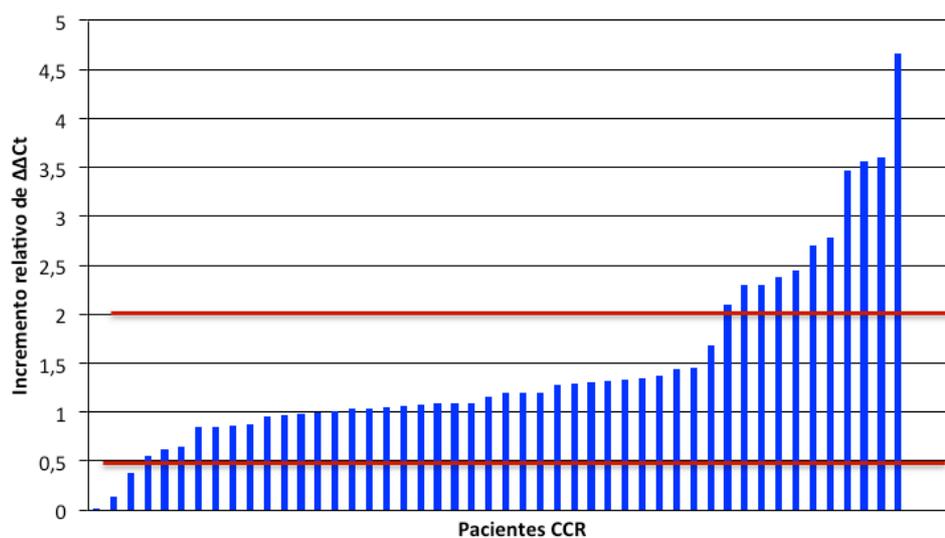
La expresión relativa ( $\Delta\Delta Ct$ ) de EZH2 mostró un valor promedio de 2.66 y una desviación estándar de 5.22, con una mediana de 1.19 y un rango intercuartílico entre 0.98 y 2.30.



**Figura 16.-** Dot plot de los valores de  $\Delta Ct$  en la muestras de tejido tumoral y tejido no tumoral. En rombos rojos la mediana para cada uno de los tejidos.

La figura 17 muestra el  $\Delta\Delta Ct$  de los 51 casos en estudio. Se consideró inhibición de la expresión de EZH2 en las muestras cuyo valor de FC o IR fue inferior a 0.5 y sobreexpresión en aquellas en las que dicho valor fue superior a 2, considerándose que de los 51 casos estudiados se detectó sobreexpresión de EZH2 en el 27.45%(14), expresión normal en el 64.70%(33) e inhibición de la expresión en el 7.84%(4). En la figura 17 se representan los niveles de expresión de los casos analizados.

### Expresión génica de EZH2



**Figura 17.-** Niveles de expresión relativa de EZH2 en la cohorte de casos analizados.

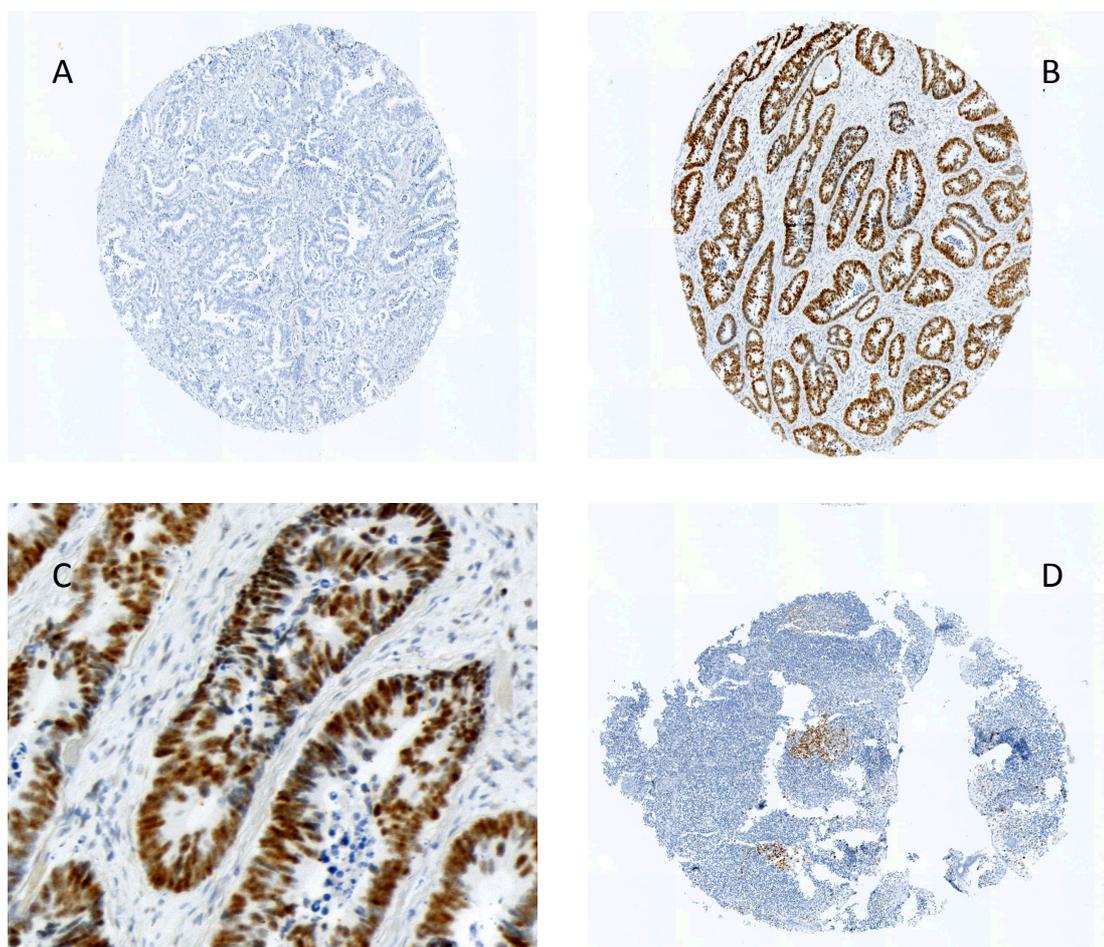
### **Análisis de la expresión proteica de EZH2**

Se analizaron 74 muestras de los 94 pacientes, de las que había disponibilidad de tejido tumoral incluido en parafina. La figura 18 muestra una imagen parcial de la matriz de tejidos (tissue array) empleada para analizar la expresión proteica de EZH2 en las muestras de CCR. Las áreas donde se detecta expresión inmunohistoquímica aparecen teñidas de color marrón.



**Figura 18.-** Imagen de un fragmento de la matriz de tejido "tissue array".

En la figura 19A, con un bajo porcentaje de células positivas para la expresión de EZH2, se observan las criptas y el epitelio glandular de color azul. Sin embargo, en la figura 19B, que muestra un porcentaje de células positivas muy elevado, se pueden observar los núcleos intensamente coloreados para la expresión de la proteína, lo que se observa con más detalle en la figura 19C. La figura 19D muestra la tinción de un fragmento de ganglio linfático, que se empleó como control positivo. Los criterios para evaluar la expresión fueron la intensidad de la inmunotinción y el porcentaje de células positivas, considerando el 20% como umbral. En función de estos dos criterios, las muestras se separaron en dos grupos, según se detectase o no la presencia de la proteína, de este modo se consideró que existía presencia de expresión proteica en el 66.22%(49) de las muestras y ausencia de expresión de la proteína EZH2 en el 33.78%(25).



**Figura 19.-** Imágenes ampliadas de la matriz de tejido. A). Imagen (5x) de una muestra negativa para la expresión de EZH2. B) Imagen (5x) de una muestra positiva para la expresión de EZH2. C) Imagen (20X) de una muestra positiva para la expresión de EZH2. D) Imagen (5x) del el control positivo empleado (ganglio linfático).

### **Análisis de polimorfismos de EZH2**

#### *Búsqueda e identificación de polimorfismos de interés*

La búsqueda e identificación de polimorfismos del gen EZH2 en las bases de Ensemble, y en Snpdatbase del NCBI, en las que se encuentra la información actualizada de la literatura científica y de proyectos de genotipado a gran escala como HapMap o 1000genomes, llevó a la selección tres polimorfismos del gen EZH2 por su posible relevancia desde un punto de vista clínico en base a los estudios publicados: rs6958683, rs2302427 y rs3757441. Las características más relevantes se describen a continuación y se resumen en la tabla 9.

**rs6958683**, se localiza en CHR\_HSCHR7\_2\_CTG6:142559136 (hebra molde), se trata de una variante genética “*upstream*”, cuya consecuencia clínica no está demostrada, por la ausencia de análisis de esta variante en la literatura y en estudios experimentales, pero una de sus posibles consecuencias podría ser la regulación de la transcripción de la proteína. Para este polimorfismo el alelo ancestral sería el C y el de menor frecuencia o MAF (MAF “*minor allele frequency*”) el G, pero no se ha encontrado mucha información acerca del mismo en la literatura.

**rs2302427**, se localiza en Chr.7:148828812 (hebra molde), Se seleccionó este polimorfismo genético porque su posible consecuencia clínica se trata de una variación con cambio de sentido, la presencia del alelo de menor frecuencia, es un tipo de mutación puntual no sinónima “*missense*” en la cual se produce un cambio en un único nucleótido, provocando la aparición de un codón que codifica para un aminoácido diferente. Esto podría provocar un producto proteico incapaz de cumplir su función. Para este polimorfismo el alelo ancestral sería el C y el de menor frecuencia o MAF el G.

**rs3757441**, se localiza en Chr.7:148827660 (hebra molde), se trata de una variante genética intrónica, cuya consecuencia más importante puede ser la modulación de la expresión del producto final proteico, debido al “*splicing*” o procesamiento alternativo del ARNm, se podría ver afectada la transcripción del gen, que tenga lugar el aumento o disminución de la actividad de la proteína. Para este polimorfismo el alelo ancestral sería el T y el de menor frecuencia o MAF el C.

	<b>rs6958683</b>	<b>rs2302427</b>	<b>rs3757441</b>
Localización	CHR_HSCHR7_2_CTG 6:142559136	Chr.7:148828812	Chr.7:148827660
Variante	Variante “ <i>upstream</i> ”	Variante “ <i>missense</i> ”	Variante intrónica
Alelo ancestral	C	C	T
Alelo menor frecuencia (MAF)	G	G	C

**Tabla 9.-** Características de los SNPs, localización, variante, alelo ancestral y de menor frecuencia (MAF).

Una vez realizado el experimento de genotipado se consideraron válidas aquellas muestras que en el análisis mostraron un porcentaje de calidad superior al 98% en las 2 réplicas realizadas. En la tabla 10 se muestra manera resumida la información referente a las

frecuencias alélicas y genotípicas que se describen a continuación para cada uno de los polimorfismos.

#### *Análisis del rs6958683*

En el análisis de este SNP todas las muestras analizadas cumplen con los criterios de inclusión, al presentar porcentajes de calidad superiores al 98%. Las frecuencias alélicas que se han encontrado para este SNP en este estudio fueron las siguientes; para el alelo C, la frecuencia fue del 81% frente al 19% de frecuencia del alelo G. De los 94 pacientes analizados, 61(65%) fueron homocigotos (presentaron el genotipo) CC, 30 fueron heterocigotos CG (32%) y sólo 3 presentaron el genotipo GG (3%). Al realizar la estimación del equilibrio Hardy-Weinberg, se observó que este polimorfismo se encontraba en equilibrio con una  $p=1$ .

#### *Análisis del rs2302427*

En el análisis del polimorfismo rs2302427 se eliminaron de los análisis dos muestras que no cumplían con los criterios de calidad prefijados, por ser su porcentaje de calidad inferior al 97%, por lo que nos quedamos para el análisis con un total de 92.

Las frecuencias alélicas para este SNP fueron del 93% para el alelo C de mayor frecuencia y del 7% para el alelo G de menor frecuencia. Los genotipos de los pacientes fueron el 89% homocigotos para el alelo C, el 8% heterocigotos CG y el 3% homocigotos para el alelo G. Según el análisis del equilibrio Hardy-Weinberg este polimorfismo se encuentra en desequilibrio H-W para esta población ( $p=0.004$ ).

#### *Análisis del rs3757441*

En el análisis del polimorfismo rs3757441 se eliminaron de los análisis dos muestras que no cumplían con los criterios de calidad prefijados por ser su porcentaje de calidad inferior al 97%, por lo que también en este caso, nos quedamos para el análisis un total de 92 muestras.

Las frecuencias alélicas para el rs3757441 fueron del 58% para el alelo T de mayor frecuencia y del 42% para el alelo C de menor frecuencia. El genotipo más frecuente fue el homocigoto para el alelo T con un 41% de los individuos, seguido del 33% heterocigotos TC y en el caso de los homocigotos para el alelo MAF, la frecuencia fue del 26% de los casos. Este polimorfismo se encuentra en desequilibrio H-W en nuestra población estudio con un valor de  $p=0.0014$ .

<b>Equilibrio Hardy-Weinberg</b>						
	<b>Frecuencias genotípicas</b>			<b>Frecuencias alélicas</b>		<b>Valor p</b>
rs6958683	CC	CG	GG	C	G	
n(94)	61(0.65)	30(0.32)	3(0.03)	152(0.81)	36(0.19)	1.000
rs2302427	CC	CG	GG	C	G	
n(94)	82(0.89)	7(0.08)	3(0.03)	171(0.93)	13(0.07)	0.004
rs3757441	TT	TC	CC	T	C	
n(92)	38(0.41)	30(0.33)	24(0.26)	106(0.58)	78(0.42)	0.0014

**Tabla 10.-** Frecuencias genotípicas y alélicas de los SNPs de EZH2 y resultados del análisis del equilibrio Hardy-Weinberg.

### ***Estudios de asociación de los marcadores moleculares con las características basales de los pacientes***

#### ***Asociación de la expresión génica EZH2 con las características sociodemográficas y anatomopatológicas***

Los resultados de los análisis de asociación entre la expresión génica de EZH2 y las características sociodemográficas y anatomopatológicas se resumen en las tablas 11 y 12.

Se observó que, en el caso de la edad, era más frecuente en pacientes mayores de 50 años, de hecho, no se determinó sobreexpresión en ningún paciente menor de 50 años, por este motivo, no se pudo estimar el riesgo expresado en OR ni tampoco los IC al 95%. El 21.43 % de los casos que sobre expresan EZH2 se corresponden con mujeres, el sexo femenino se asocia con un menor riesgo de sobreexpresar el gen EZH2 con un OR=0.56 (0.13-2.42). En ninguna de estas asociaciones las diferencias observadas fueron estadísticamente significativas.

En el análisis de distribución por localización tumoral, que se realizó considerando la distribución en 3 categorías (recto, colon izquierdo y colon derecho) y en 2 categorías (recto y colon), se observó una mayor frecuencia de sobreexpresión del gen en la categoría de recto, y un menor riesgo en el colon con un valor de OR=0.52 (0.07-3.56) tanto en el análisis por dos como por tres categorías pero no de forma estadísticamente significativa.

No se pudo estimar la asociación entre el tipo histológico y la sobreexpresión de EZH2, dado que todos los casos en los que el gen EZH2 aparece sobreexpresado son de tipo adenocarcinoma, esta distribución se encuentra próxima a la significación estadística. Además

los tumores que sobreexpresan EZH2 fueron más frecuentemente de bajo grado, considerando de modo conjunto, bien y moderadamente diferenciados, con un OR= 0.23 (0.02-2.05).

En cuanto al tamaño e invasión del tumor, la mayor parte de las muestras que sobreexpresan EZH2, el 78.57% se correspondieron con una extensión anatómica de pT3, el tumor ha crecido hacia las capas más externas del colon o del recto, pero no ha atravesado estas capas. La sobreexpresión del gen se asoció con una menor afectación de los ganglios linfáticos con un OR=0.62 (0.21-1.75) y con una menor presencia de metástasis a distancia con un OR= 0.39 (0.04-3.63) , pero no de forma estadísticamente significativa.

Esta misma tendencia se observó en el estadio TNM, a mayor estadio menor sobreexpresión de EZH2 aunque no de forma significativa, para el estadio IV el valor de OR= 0.25 (0.01-3.99) y  $p= 0.35$ . Sin embargo, si recodificamos en dos categorías este efecto se suavizó, mostrando una asociación entre una menor sobreexpresión con estadios avanzados III y IV con un OR=0.89 (0.26-3.07). La sobreexpresión génica de EZH2 se asoció con la aparición de segundos tumores primarios con un OR=2.71 (0.73-10.14).

### ***Asociación de la expresión de la proteína EZH2 con las características sociodemográficas y anatomopatológicas***

Los resultados obtenidos en los análisis de asociación entre la presencia de la proteína EZH2 y las características sociodemográficas y anatomopatológicas se describen en las tablas 13 y 14.

En función de las características demográficas, en las mujeres es menor la presencia de la proteína con un OR=0.66 (0.24-1.80), y en el análisis por edad, la presencia de expresión fue más frecuente en el grupo de mayor edad OR=4.47 (0.38-52.13) no existiendo significación estadística en ninguno de los 2 casos.

Tal y como se muestra en las tablas 12 y 13, ninguna de las asociaciones analizadas mostró significación estadística, si bien es interesante describir, tal y como se hace a continuación, algunas de las asociaciones detectadas entre la presencia de proteína EZH2 en los cortes histológicos del tissue array y las características anatomopatológicas de los pacientes estudiados.

	EZH2-		EZH2+		p	OR	IC95%	p
	ARNm	ARNm	ARNm	ARNm	(X <sup>2</sup> /Fisher)			(Wald)
<b>Edad</b>	<50 años	2 (5.71)	0 (0.00)	0 (0.00)	0.363	1	1	-----
	>50 años	33 (94.29)	14 (100.00)	14 (100.00)		---	---	
<b>Género</b>	Hombres	25 (67.57)	11 (78.57)	11 (78.57)	0.513*	1	1	0.445
	Mujeres	12 (32.43)	3 (21.43)	3 (21.43)		0.56 (0.13-2.42)		
<b>Tipo histológico</b>	Adenocarcinoma	28 (77.78)	14 (100.00)	14 (100.00)	0.087*	1	1	-----
	Otros tipos	8 (22.22)	0 (0.00)	0 (0.00)		---	---	
<b>Localización tumoral(x3)</b>	Recto	3 (8.11)	2 (14.29)	2 (14.29)		1	1	
	Colon izquierdo	17 (45.95)	6 (42.86)	6 (42.86)	0.809*	0.81 (0.32-2.05)		0.651
	Colon derecho	17 (45.95)	6 (42.86)	6 (42.86)		0.50 (0.12-1.98)		0.328
	Recto Colon	3 (8.11) 34 (91.89)	2 (14.29) 12 (85.71)	2 (14.29) 12 (85.71)	0.508*	0.52 (0.07-3.56)		0.513
<b>Grado de diferenciación(x3)</b>	Bien diferenciado	12 (37.5)	5 (35.71)	5 (35.71)		1	1	
	Moderadamente diferenciado	12 (37.5)	8 (57.14)	8 (57.14)	0.321*	1.6 (0.4-6.32)		0.495
	Pobresmente diferenciado	8 (25.00)	1 (30.43)	1 (30.43)		0.3 (0.02-3.07)		0.214
<b>Grado diferenciación(x2)</b>	Bajo grado	24 (75.00)	13 (92.86)	13 (92.86)	0.240*	1	1	0.189
	Alto grado	8 (25.00)	1 (7.14)	1 (7.14)		0.23 (0.03-2.05)		

Tabla 11.- Análisis de asociación de la expresión génica con las características sociodemográficas y anatomopatológicas.

	EZH2- ARNm	EZH2+ ARNm	p (X <sup>2</sup> /Fisher)	OR	IC95%	p (Wald)
pT						
pT1	1 (2.78)	0 (0.00)		1		
pT2	3 (8.33)	2 (14.29)	0.388*	6.67 (0.46-101.73)		0.172
pT3	22 (61.11)	11 (78.57)		5.00 (0.57-44.20)		0.148
pT4	10 (27.78)	1 (7.14)		1 (-)		
pN						
pN0	18 (50)	8 (57.14)		1		
pN1	14 (38.89)	6 (42.86)	0.647*	0.96 (0.27-3.42)		0.955
pN2	4 (11.11)	0 (0.00)		1 (-)		
Metástasis a distancia						
M0	31(70.45)	13 (92.86)		1		0.414
M1	6(16.22)	1 (7.14)	0.657*	0.39 (0.04-3.63)		
Estadio I	3 (8.33)	2 (14.29)		1		
Estadio II	14 (38.89)	5 (35.71)		0.53 (0.06-4.20)		0.553
Estadio III	13 (36.11)	6 (42.86)	0.846*	0.69 (0.09-5.29)		0.723
Estadio IV	6 (16.67)	1 (14.29)		0.25 (0.01-3.99)		0.35
Estadio TNM(2x)						
Estadios iniciales (I y II)	17 (47.22)	7 (50.00)		1		0.860
Estadios avanzados (III y IV)	19 (52.78)	7 (50.00)	0.860	0.89 (0.26-3.07)		
Segundos tumores primarios						
No	29 (78.38)	8 (57.14)		1		0.136
Sí	8 (21.62)	6 (27.45)	0.129	2.71 (0.73-10.14)		

Tabla 12.- Análisis de asociación de la expresión génica con las características sociodemográficas y anatomopatológicas

La expresión de la proteína se asoció con localizaciones más frecuentes en colon izquierdo OR=1.39 (0.32-5.97) y en colon derecho OR=1.33 (0.30-5.83), en el análisis por tres categorías. Cuando este análisis se realizó por dos categorías la expresión se asociaba directamente con localización en colon con un valor de OR=1.36 (0.34-5.36).

La distribución de la expresión de la proteína con las diferentes características clínico-patológicas fue la siguiente, por un lado, al estudiar su asociación con el tipo histológico se observó asociación sin alcanzar la significación estadística con tumores con diagnóstico de adenocarcinoma, la asociación entre la presencia de la proteína y otro tipo histológico diferente de adenocarcinoma fue OR=0.63 (0.17-2.25). Por otro lado en cuanto a la relación con el grado de diferenciación, la expresión proteica EZH2 se asoció con tumores de bajo grado y bien diferenciados en el análisis por dos categorías, con un valor de OR= 0.71 (0.23-1.20) y cuando se realizó el análisis por tres categorías se observó que la proteína se expresa menos en los tumores pobremente diferenciados que en los tumores bien diferenciados y moderadamente diferenciados con un OR=0.35 (0.08-1.53).

En cuanto al grado de extensión e invasión del tumor, se observó que la expresión de la proteína se asociaba de forma más frecuente con las categorías pT2 y pT3. Al analizar la afectación de los ganglios la expresión de la proteína se asociaba a una mayor afectación de los ganglios pN2, y con un mayor número de ganglios afectados, aunque no de forma significativa con un OR=1.23 (0.56-2.64). Sin embargo en el caso de la presencia de metástasis a distancia, la sobreexpresión de la proteína se asocia con una menor frecuencia de metástasis a distancia con un OR= 0.36 (0.09-1.32).

Cuando se realizó el análisis por estadio en cuatro categorías se observó que la expresión de la proteína se asociaba con estadios más bajos e intermedios que con estadios avanzados, los casos que en el momento de diagnóstico presentaban estadio IV de la enfermedad expresaban menos la proteína que los casos pertenecientes a los otros estadios, con un valor de OR=0.16 (0.01-1.93). En el análisis realizado de la distribución de la expresión en dos categorías se observó que la sobreexpresión de EZH2 se asociaba a estadios iniciales, mientras que los estadios avanzados mostraban una asociación con una menor expresión de la proteína OR=0.51 (0.19-1.38).

La presencia de expresión de la proteína se asocia inversamente con la aparición de tumores primarios, los pacientes en los que se ha detectado la presencia de proteína tienen un menor riesgo de aparición de segundos tumores primarios con un OR=0.83 (0.28-2.47).

	Proteína EZ2-	Proteína EZ2+	p (x2/Fisher)	OR IC95%	p (Wald)
<b>Tipo histológico</b>					
Adenocarcinoma	19 (79.17)	42 (85.71)	0.478	1	0.481
Otros tipos	5 (20.83)	7 (14.29)		0.63 (0.17-2.25)	
<b>Edad</b>					
<50 años	2 (8.70)	1 (2.08)	0.195	1	0.231
>50 años	21 (91.30)	47 (97.92)		4.47 (0.38-52.13)	
<b>Sexo</b>					
Hombre	15 (60.00)	34 (69.39)	0.419	1	0.421
Mujer	10 (40.00)	15 (30.61)		0.66 (0.24-1.80)	
<b>Localización tumoral (x3)</b>					
Recto	4 (16.00)	6 (12.24)	0.902	1	0.655
Colon izquierdo	11 (44.00)	23 (46.94)		1.39 (0.32-5.97)	0.655
Colon derecho	10 (40.00)	20 (40.82)	0.655	1.33 (0.30-5.83)	0.702
<b>Localización tumoral (x2)</b>					
Recto	4 (16.00)	6 (12.24)	0.655	1	0.656
Colon	21 (84.00)	43 (87.76)		1.36 (0.34-5.36)	
<b>Grado de diferenciación (x3)</b>					
Bien diferenciado	4 (16.67)	16 (36.36)	0.234	1	0.112
Moderadamente diferenciado	13 (54.17)	18 (40.91)		0.34 (0.09-1.27)	0.112
Pobremamente diferenciado	7 (29.17)	10 (22.73)	0.558	0.35 (0.08-1.53)	0.167
<b>Grado de diferenciación (x2)</b>					
Bajo grado	34 (70.83)	51 (77.27)	0.558	1	0.559
Alto grado	7 (29.17)	10 (22.23)		0.71 (0.23-1.20)	

Tabla 13.- Asociación de la expresión proteica con las características sociodemográficas y anatómicas.

	Proteína EZH2-	Proteína EZH2+	p (x2/Fisher)	OR IC95%	p(Wald)
<b>pT</b>					
pT1	0 (0.00)	1 (2.08)		1	
pT2	2 (8.00)	6 (12.50)	0.871	2.00(0.29-13.43)	0.476
pT3	17 (68.00)	32 (66.67)		1.25 (0.38-4.11)	0.408
pT4	6 (24.00)	9 (18.75)		1	-
<b>pN</b>					
pN0	13 (52.00)	27 (56.25)		1	
pN1	12 (48.00)	15 (31.25)	0.116	0.60(0.22-1.65)	0.323
pN2	0 (0.00)	6 (12.5)		1 (-)	
<b>Metástasis a distancia</b>					
M0	19 (76.00)	44 (89.80)		1	0.124
M1	6 (24.00)	5 (10.20)	0.115	0.36 (0.09-1.32)	
<b>Estadio TNM</b>					
Estadio I	1 (4.00)	5 (10.42)		1	
Estadio II	9 (36.00)	22 (45.83)	0.4121	0.48(0.04-4.79)	0.539
Estadio III	9 (36.00)	16 (33.33)		0.35(0.03-3.539)	0.378
Estadio IV	6 (24.00)	5 (10.42)		0.16 (0.01-1.93)	0.152
<b>Estadio TNM(2x)</b>					
Estadíos I y II	10 (40.00)	27 (56.25)		1	0.190
Estadíos III y IV	15 (60.00)	21 (43.25)	0.188	0.51(0.19-1.38)	
<b>Segundos tumores primarios</b>					
No	18 (72.00)	37 (75.51)		1	
Sí	7 (28.00)	12 (24.49)	0.744	0.83 (0.28-2.47)	0.744

Tabla 14.- Análisis de asociación de la expresión génica con las características sociodemográficas y anatomopatológicas.

## **Asociación de los polimorfismos de EZH2 con las variables anatomopatológicas**

### *Distribución de rs6958683 en función de las características anatomopatológicas*

Los resultados de la asociación del rs6958683 y las variables sociodemográficas y anatomopatológicas se muestran en la tabla 15.

Entre el polimorfismo rs6958683 y la edad, recodificada en dos grupos; menores de cincuenta años y mayores de cincuenta, se observó una pequeña asociación OR=1.23 (0.14-11.13) entre la presencia de al menos un alelo G y el grupo de mayor edad aunque dicha asociación no fue estadísticamente significativa en ninguno de los modelos estudiados.

No se encontraron diferencias en este polimorfismo en cuanto a la distribución por género, si se observó una pequeña tendencia entre la presencia de al menos un alelo G con el sexo femenino con un valor de OR=1.19 (0.55-2.59) pero no de forma significativa.

Se observó asociación entre el genotipo GG estudiado desde un modelo de herencia recesivo y la presencia de tipos histológicos en los tumores diferentes a adenocarcinoma (adenocarcinoma mucinoso y carcinoma de células en anillo de sello) con significación estadística con un valor de OR=11.54 (0.97-136.64) y un valor de  $p=0.047$ . Se confirma esta asociación al emplear el modelo *log-additive* se confirma la asociación aunque no de forma significativa, entre portar al menos un alelo G con diagnóstico histológico diferente de adenocarcinoma con un OR=1.72 (0.67-4.38) y un valor de  $p=0.27$

Se recodificó la localización tumoral en dos categorías; colon y recto, y se analizó la distribución de este SNP tomando el valor que proporciona el modelo *log-additive* que la presencia de al menos un alelo G se asociaba con un menor riesgo de localización en colon con un OR=0.83 (0.31-2.27) y un valor de  $p=0.72$ . Además el genotipo GG de este polimorfismo mostró asociación con localización en recto o asociación inversa con localización en colon con OR=0.35 (0.03-4.31),  $p=0.72$ .

En cuanto a la distribución de este polimorfismo según el grado de diferenciación tampoco se observó ninguna asociación significativa pero la presencia de al menos un alelo G se asoció a tumores de bajo grado de diferenciación con un OR= 0.87 (0.33-2.27),  $p=0.78$ .

Los portadores de al menos un alelo G mostraron una mayor presencia de metástasis a distancia en el momento del diagnóstico, la asociación hallada para este genotipo y la presencia de metástasis fue OR=1.52 (0.55-4.26),  $p=0.25$ .

El genotipo GG se asocia a estadios TNM más iniciales con OR=0.50 (0.04-5.81) y  $p=0.85$ . Ninguno de los portadores de este genotipo GG presentaban sobreexpresión génica de EZH2 y

la presencia de al menos un alelo G se asocia con una menor expresión del gen OR=0.74 (0.21-2.61) y una  $p=0.64$ .

En cuanto a la expresión de la proteína, se observó que el alelo G en homocigosis, tanto en el modelo codominante como en el recesivo presenta la mitad de riesgo de sobreexpresión de la proteína con un valor de OR= 0.53 (0.03-9.11) para el modelo dominante y un OR=0.52 (0.03-8.71) para el modelo recesivo, sin significación estadística. Este efecto no se observa en el modelo *log-additive* en el que la presencia de al menos un alelo G no se asocia con un menor riesgo de sobreexpresión de la proteína, valor de OR muy próximo a 1, OR= 0.94 (0.39-2.30).

Se calculó la asociación de este genotipo GG y los eventos de muerte y recidiva y se observó, que ninguno de los portadores de este genotipo habían tenido eventos de recidiva y mostraron asociación con eventos de *exitus* con un OR=3.31(0.27-40.88) y la presencia de al menos un alelo G se asoció a un mayor riesgo de presentar un evento de *exitus* OR=1.73(0.68-4.40).

#### *Distribución de rs2302427 en función de las características anatomopatológicas*

Los resultados de la asociación del rs2302427 y las variables sociodemográficas y anatomopatológicas se muestran en la tabla 16.

En el análisis de la distribución del polimorfismo rs2302427 se observaron diferencias en la distribución por sexo, estadísticamente significativas, la presencia de al menos un alelo G se asocia con el género masculino, las mujeres tienen un riesgo cuatro veces menor de presentar al menos un alelo G con un valor de OR=0.23 (0.03-1.58) y un valor de  $p=0.048$ .

Cuando analizamos este SNP en función de la edad, recodificada en dos grupos, se observó que los genotipos portadores de un alelo G se asocian inversamente con el grupo de edad de más de 50 años con OR=0.53(0.09-3.22), aunque no de forma significativa, lo que implica que en el grupo de menos de 50 años la frecuencia de portar al menos un alelo G fue mayor que en el grupo de más de 50, se asocia a individuos jóvenes.

En cuanto a la asociación del polimorfismo con el tipo histológico, no se pudo hallar el cálculo de OR, ya que no existe representación del alelo de menor frecuencia que es el alelo G, ya que no existe ningún representante que porte este alelo y presente un tipo histológico diferente a adenocarcinoma. No se pudieron obtener resultados de asociación entre la localización tumoral en tres categorías y este polimorfismo debido a que no existe representación de ningún alelo G con localización en recto. Al recodificar en dos categorías la variable localización tumoral, y debido a que no existe representación del alelo G en

localización tumoral en recto, dado que todos los casos tanto homocigotos para G como heterocigotos se localizaron en el colon.

Por otro lado, en cuanto a la distribución de este polimorfismo en función del grado de diferenciación se observó, que el alelo G en homocigosis, se asocia a tumores de alto grado de diferenciación o pobremente diferenciados pero no de forma significativa, con un valor de OR=1.47 (0.13-17.15) y una  $p=0,76$ .

La presencia de al menos un alelo G se asoció con un menor riesgo de metástasis a órganos distantes en el momento del diagnóstico, con un valor de OR=0.63 (0.1-3.84) y una  $p=0.58$ . El genotipo GG de este polimorfismo se asocia a una menor presencia de metástasis a distancia OR=0.51(0.04-5.88). De la misma forma se observó cuando analizamos la distribución de este polimorfismo en función del estadio, recodificado en dos categorías estadios iniciales (Estadio I y II) y estadios avanzados (III) que la presencia de al menos un alelo G se asocia con un menor riesgo de estadios o enfermedad avanzada con un valor de OR=0.74 (0.28-1.96) y un valor de  $p=0.53$ , este riesgo disminuye, cuando el alelo G aparece en homocigosis con un valor de OR=0.52 (0.05-5.98) y un valor de  $p=0.82$ .

No se pudo analizar la distribución de este polimorfismo en función de la expresión de EZH2 detectada a partir de ARNm, dado que no existe ningún individuo homocigoto para G o heterocigoto que sobreexpresen el gen. Todos, el 100% de los individuos que sobreexpresan EZH2 presentan fenotipo CC.

Al analizar la expresión proteica, se observó que la presencia de al menos un alelo G se asociaba a una menor sobreexpresión de la proteína EZH2 con un valor de OR=0.72 (0.23-2.28) y una  $p=0.58$ . Este riesgo disminuye si el alelo G se encuentra en homocigosis, el valor de OR=0.51 (0.03-8.53) y un  $p=0.64$ , para el genotipo GG(modelo recesivo).

Al analizar este polimorfismo en función de los eventos de recidiva y muerte, se observó que ningún individuo con genotipo GG presentó eventos de recidiva o muerte.

#### *Distribución de rs3757441 en función de las características anatomopatológicas*

En el análisis del polimorfismo rs3757441 en función de las diferentes características anatomopatológicas se han observado las siguientes distribuciones cuyos resultados se muestran completos en la tabla 17.

La presencia de al menos un alelo C se asoció con el género femenino de forma significativa, este alelo aparece de forma más frecuente en mujeres que en hombres, con un OR=1.65 e IC95% (0.95-2.86) con un valor de  $p=0.07$  para el modelo *Log-additive*. Cuando este

alelo aparece en homocigosis el efecto de dicha asociación es más evidente con OR=2.68 (0.87-8.25).

Al analizar la distribución de este polimorfismo en grupo de edad <50 años y >50 años, observamos que la presencia de al menos un alelo C se asocia al grupo de mayor edad, con un OR=1.36 (0.30-6.21) pero no de forma estadísticamente significativa,  $p=0.68$ .

No se observan diferencias significativas en cuanto al tipo histológico, si se ha observado una tendencia, el alelo de menor frecuencia el alelo C, en homocigosis parece estar asociado a tipos histológicos diferentes de Adenocarcinoma. En cuanto a la localización tumoral, no se observaron diferencias en cuanto a la localización anatómica y ninguno de los genotipos, los valores de OR para todos los modelos fueron muy próximos a 1.

En la distribución de este SNP y la presencia de metástasis a distancia se observó un riesgo del doble para los individuos con genotipo heterocigoto CT frente a los individuos homocigotos para C o T, con valores de OR = 2.25 (0.66-7.69),  $p=0.2$ . Los individuos que tenían el genotipo homocigoto C/C, presentaron la mitad de riesgo para la presencia de metástasis a distancia OR=0.54 (0.11-2.68) y un valor de  $p=0.43$ .

Cuando se quiso comprobar la asociación de este polimorfismo en función de estadio, en dos categorías, enfermedad inicial y avanzada, se observó que la presencia de al menos un alelo GC proporcionaba un efecto protector frente a la enfermedad a distancia aunque no de forma significativa con un valor de OR= 0.62 (0.31-1.23) y una  $p= 0.16$ , en los individuos que fueron homocigotos para este alelo C, CC, el riesgo disminuye y casi alcanza la significación estadística OR=0.38 (0.14-1.03) y un valor de  $p=0.051$ .

En cuanto a la sobreexpresión a partir de ARN, observamos que el genotipo CC de rs3757441 se asociaba a una menor expresión del gen con OR= 0.86 (0.15-4.87) y una  $p=0.86$ .

La misma tendencia se observó en el análisis de la proteína ya que la presencia de al menos un alelo C se asocia a una menor presencia de la proteína OR=0.59 e IC95 %(0.32-1.10) y un valor de  $p=0.093$  y en los individuos con genotipo CC el valor disminuía a OR= 0.35 (0.10-1.18).

Este genotipo GG se asoció además con una menor presencia de eventos de recidiva OR=0.49 (0.14-1.77) pero no se observó asociación con los evento de *exitus* OR=1.07(0.27-4.25).

Modelo de herencia rs6958683 C (Ancestral)/ G(MAF)		Codominante			Dominante			Recesivo			Sobredominante			Log-additive
		C/C	C/G	G/G	C/C	C/G-G/G	C/C- C/G	G/G	C/C- G/G	C/C	C/G	C/C- G/G		
Edad (<50 años y > 50 años)	OR (IC95%)	1	1.00 (0.09-11.53)	NA (0.00-NA)	1	1.11 (0.10-12.77)	1	NA (0.00-NA)	1	0.95 (0.08-10.91)	1	0.95 (0.08-10.91)	1.23 (0.14-11.13)	
	p-valor		0.900			0.930		0.640		0.970		0.970	0.850	
Sexo	OR (IC95%)	1	1.28 (0.51-3.21)	1.11 (0.09-12.95)	1	1.26 (0.52-3.08)	1	1.02 (0.09-11.66)	1	1.27 (0.51-3.17)	1	1.27 (0.51-3.17)	1.19 (0.55-2.59)	
	p-valor		0.870			0.610		0.990		0.600		0.600	0.650	
Tipo histológico	OR (IC95%)	1	0.89 (0.25-3.17)	11.11 (0.91-135.80)	1	1.28 (0.41-3.99)	1	11.54 (0.97-136.64)	1	0.74 (0.21-2.56)	1	0.74 (0.21-2.56)	1.72 (0.67-4.38)	
	p-valor		0.140			0.670		0.047		0.630		0.630	0.270	
Localización tumoral	OR (IC95%)	1	1.10 (0.31-3.93)	0.35 (0.03-4.31)	1	0.95 (0.29-3.13)	1	0.34 (0.03-4.05)	1	1.18 (0.34-4.13)	1	1.18 (0.34-4.13)	0.83 (0.31-2.27)	
	p-valor		0.720			0.940		0.430		0.790		0.790	0.720	
Grado de diferenciación	OR (IC95%)	1	1.08 (0.37-3.11)	0.00 (0.00-NA)	1	0.98 (0.34-2.79)	1	0.00 (0.00-NA)	1	1.13 (0.39-3.25)	1	1.13 (0.39-3.25)	0.87 (0.33-2.27)	
	p-valor		0.550			0.960		0.280		0.820		0.820	0.780	
Metástasis a distancia	OR (IC95%)	1	2.35 (0.68-8.05)	0.00 (0.00-NA)	1	2.08 (0.61-7.07)	1	0.00 (0.00-NA)	1	2.48 (0.72-8.49)	1	2.48 (0.72-8.49)	1.52 (0.55-4.26)	
	p-valor		0.260			0.240		0.360		0.150		0.150	0.430	
Estadio TNM	OR (IC95%)	1	0.93 (0.38-2.27)	0.50 (0.04-5.81)	1	0.88 (0.37-2.08)	1	0.51 (0.04-5.84)	1	0.96 (0.40-2.32)	1	0.96 (0.40-2.32)	0.85 (0.40-1.80)	
	p-valor		0.850			0.780		0.580		0.930		0.930	0.670	
Expresión ARNm	OR (IC95%)	1	0.87 (0.22-3.41)	0.00 (0.00-NA)	1	0.80 (0.21-3.09)	1	0.00 (0.00-NA)	1	0.91 (0.23-3.54)	1	0.91 (0.23-3.54)	0.74 (0.21-2.61)	
	p-valor		0.700			0.740		0.410		0.890		0.890	0.640	
Expresión proteica	OR (IC95%)	1	1.07 (0.38-3.03)	0.53 (0.03-9.11)	1	1.01 (0.37-2.77)	1	0.52 (0.03-8.71)	1	1.10 (0.39-3.09)	1	1.10 (0.39-3.09)	0.94 (0.39-2.30)	
	p-valor		0.900			0.990		0.650		0.860		0.860	0.900	
Recidiva	OR (IC95%)	1	2.48 (0.86-7.10)	0.00 (0.00-NA)	1	2.17 (0.76-6.15)	1	0.00 (0.00-NA)	1	2.62 (0.91-7.50)	1	2.62 (0.91-7.50)	1.57 (0.65-3.81)	
	p-valor		0.130			0.150		0.250		0.074		0.074	0.320	
Exitus	OR (IC95%)	1	1.66 (0.52-5.30)	3.31 (0.27-40.88)	1	1.78 (0.58-5.45)	1	2.75 (0.23-32.42)	1	1.53 (0.49-4.77)	1	1.53 (0.49-4.77)	1.73 (0.68-4.40)	
	p-valor		0.530			0.310		0.450		0.470		0.470	0.260	

Tabla 15.- Distribución de rs6958683 en función de las características anatomopatológicas.

Modelo de herencia		Codominante			Dominante			Recesivo			Sobredominante			Log-additive
rs2302427		C/C	C/G	G/G	C/C	C/G-G/G	C/C- C/G	G/G	C/C- G/G	C/C- G/G	C/G	C/G	---	
C (Ancestral)/ G(MAF)														
Edad (<50 años y > 50 años)	OR (IC95%)	1	0.16 (0.01-2.06)	NA (0.00-NA)	1	0.24 (0.02-2.96)	1	NA (0.00-NA)	1	0.16 (0.01-1.98)	1	0.16 (0.01-1.98)	0.53 (0.09-3.22)	
	p-valor		0.410			0.310		0.640		0.200		0.200	0.530	
Sexo	OR (IC95%)	1	0.29 (0.03-2.52)	NA (0.00-NA)	1	0.19 (0.02-1.60)	1	0.00 (0.00-NA)	1	0.31 (0.04-2.66)	1	0.31 (0.04-2.66)	0.23 (0.03-1.58)	
	p-valor		0.130			0.066		0.110		0.230		0.230	0.048	
Tipo histológico	OR (IC95%)	1	0.00 (0.00-NA)	0.00 (0.00-NA)	1	0.00 (0.00-NA)	1	0.00 (0.00-NA)	1	0.00 (0.00-NA)	1	0.00 (0.00-NA)	0.00 (0.00-NA)	
	p-valor		0.140			0.049		0.290		0.100		0.100	0.049	
Localización tumoral	OR (IC95%)	1	NA (0.00-NA)	NA (0.00-NA)	1	NA (0.00-NA)	1	NA (0.00-NA)	1	NA (0.00-NA)	1	NA (0.00-NA)	NA (0.00-NA)	
	p-valor		0.170			0.059		0.310		0.120		0.120	0.059	
Grado de diferenciación	OR (IC95%)	1	0.57 (0.06-5.18)	1.42 (0.12-16.58)	1	0.81 (0.15-4.25)	1	1.47 (0.13-17.15)	1	0.56 (0.06-5.09)	1	0.56 (0.06-5.09)	0.98 (0.32-3.00)	
	p-valor		0.830			0.800		0.760		0.590		0.590	0.970	
Metástasis a distancia	OR (IC95%)	1	1.06 (0.12-9.67)	0.00 (0.00-NA)	1	0.71 (0.08-6.14)	1	0.00 (0.00-NA)	1	1.11 (0.12-10.08)	1	1.11 (0.12-10.08)	0.63 (0.10-3.84)	
	p-valor		0.650			0.740		0.350		0.930		0.930	0.58	
Estadio TNM	OR (IC95%)	1	0.77 (0.16-3.65)	0.51 (0.04-5.88)	1	0.68 (0.18-2.60)	1	0.52 (0.05-5.98)	1	0.79 (0.17-3.73)	1	0.79 (0.17-3.73)	0.74 (0.28-1.96)	
	p-valor		0.820			0.570		0.590		0.760		0.760	0.740	
Expresión ARNm	OR (IC95%)	---	---	---	---	---	---	---	---	---	---	---	---	
	p-valor		---			---		---		---		---	---	
Expresión proteica	OR (IC95%)	1	0.75 (0.12-4.82)	0.50 (0.03-8.38)	1	0.67 (0.14-3.24)	1	0.51 (0.03-8.53)	1	0.77 (0.12-4.92)	1	0.77 (0.12-4.92)	0.72 (0.23-2.28)	
	p-valor		0.860			0.620		0.640		0.780		0.780	0.580	
Recidiva	OR (IC95%)	1	3.64 (0.73-18.11)	0.00 (0.00-NA)	1	2.08 (0.48-9.05)	1	0.00 (0.00-NA)	1	3.80 (0.77-18.89)	1	3.80 (0.77-18.89)	1.24 (0.40-3.80)	
	p-valor		0.170			0.350		0.260		0.120		0.120	0.720	
Exitus	OR (IC95%)	1	2.12 (0.37-12.14)	0.00 (0.00-NA)	1	1.33 (0.25-6.97)	1	0.00 (0.00-NA)	1	2.22 (0.39-12.66)	1	2.22 (0.39-12.66)	0.95 (0.26-3.53)	
	p-valor		0.420			0.740		0.300		0.390		0.390	0.940	

Tabla 16.- Distribución de rs2302427 en función de las características anatomopatológicas.

Modelo de herencia		Codominante			Dominante			Recesivo			Sobredominante			Log-additive
rs3757441		T/T	T/C	C/C	T/T	T/C-C/C	T/T-T/C	C/C	T/T-C/C	T/T-T/C	T/C	T/C	---	
T (Ancestral)/ C(MAF)														
Edad (<50 años y > 50 años)	OR (IC95%)	1	0.40 (0.03-4.62)	NA (0.00-NA)	1	0.72 (0.06-8.27)	1	NA (0.00-NA)	1	0.24 (0.02-2.78)	1	0.24 (0.02-2.78)	1.36 (0.30-6.21)	
	p-valor		0.3			0.790		0.180		0.240		0.240	0.680	
Sexo	OR (IC95%)	1	2.50 (0.86-7.28)	2.68 (0.87-8.25)	1	2.58 (1.00-6.67)	1	1.71 (0.65-4.50)	1	1.63 (0.65-4.06)	1	1.63 (0.65-4.06)	1.65 (0.95-2.86)	
	p-valor		0.130			0.044		0.280		0.300		0.300	0.070	
Tipo histológico	OR (IC95%)	1	0.60 (0.14-2.62)	1.44 (0.38-5.38)	1	0.94 (0.30-2.98)	1	1.76 (0.52-5.93)	1	0.51 (0.13-2.01)	1	0.51 (0.13-2.01)	1.17 (0.58-2.36)	
	p-valor		0.530			0.920		0.370		0.320		0.320	0.670	
Localización tumoral	OR (IC95%)	1	0.63 (0.17-2.29)	1.04 (0.22-4.84)	1	0.76 (0.23-2.50)	1	1.31 (0.33-5.18)	1	0.62 (0.19-1.97)	1	0.62 (0.19-1.97)	0.98 (0.48-1.99)	
	p-valor		0.720			0.650		0.700		0.420		0.420	0.950	
Grado de diferenciación	OR (IC95%)	---	---	---	---	---	---	---	---	---	---	---	---	
	p-valor		---			---		---		---		---	---	
Metástasis a distancia	OR (IC95%)	1	2.06 (0.52-8.12)	0.79 (0.13-4.68)	1	1.47 (0.41-5.28)	1	0.54 (0.11-2.68)	1	2.25 (0.66-7.69)	1	2.25 (0.66-7.69)	0.98 (0.46-2.10)	
	p-valor		0.420			0.550		0.430		0.20		0.20	0.960	
Estadio TNM	OR (IC95%)	1	1.24 (0.47-3.26)	0.41 (0.14-1.24)	1	0.78 (0.34-1.82)	1	0.38 (0.14-1.03)	1	1.71 (0.71-4.14)	1	1.71 (0.71-4.14)	0.69 (0.41-1.17)	
	p-valor		0.140			0.570		0.051		0.230		0.230	0.160	
Expresión ARNm	OR (IC95%)	1	1.38 (0.36-5.28)	1.00 (0.16-6.35)	1	1.26 (0.37-4.36)	1	0.86 (0.15-4.87)	1	1.38 (0.39-4.86)	1	1.38 (0.39-4.86)	1.08 (0.46-2.49)	
	p-valor		0.880			0.710		0.860		0.610		0.610	0.870	
Expresión proteica	OR (IC95%)	1	0.70 (0.22-2.24)	0.35 (0.10-1.18)	1	0.51 (0.19-1.41)	1	0.41 (0.14-1.22)	1	1.06 (0.38-2.98)	1	1.06 (0.38-2.98)	0.59 (0.32-1.10)	
	p-valor		0.230			0.190		0.110		0.910		0.910	0.093	
Recidiva	OR (IC95%)	1	0.27 (0.07-1.09)	0.49 (0.14-1.77)	1	0.37 (0.13-1.05)	1	0.77 (0.23-2.62)	1	0.35 (0.09-1.31)	1	0.35 (0.09-1.31)	0.62 (0.31-1.23)	
	p-valor		0.130			0.059		0.670		0.092		0.092	0.160	
Exitus	OR (IC95%)	1	0.82 (0.21-3.22)	1.07 (0.27-4.25)	1	0.93 (0.29-2.93)	1	1.16 (0.33-4.11)	1	0.80 (0.23-2.80)	1	0.80 (0.23-2.80)	1.02 (0.50-2.06)	
	p-valor		0.940			0.900		0.820		0.720		0.720	0.960	

Tabla 17.- Distribución de rs3757441 en función de las características anatomopatológicas.

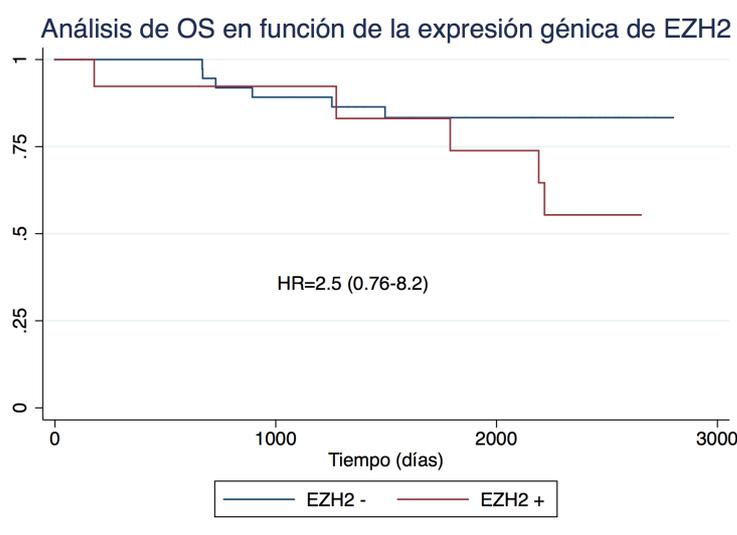
### **Análisis de supervivencia en función de los marcadores moleculares**

A continuación se muestran los análisis de supervivencia teniendo en cuenta los dos modelos propuestos, por un lado, la supervivencia global u *overall survival* (OS), en la que el evento final considerado es la muerte del paciente independientemente de la causa, y por otro la supervivencia libre de enfermedad o *disease-free survival* en la que el evento final considerado es la aparición de la enfermedad, sea una recidiva y/o la muerte del paciente.

### **Análisis de la supervivencia global en función de los marcadores moleculares**

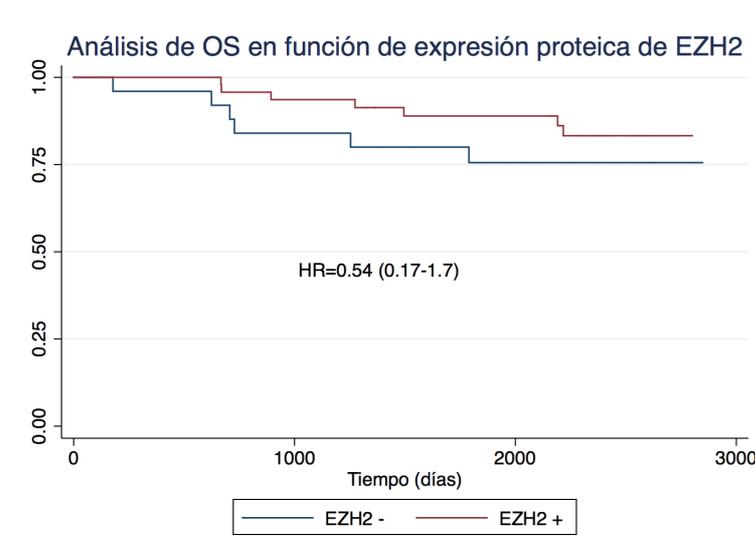
En el periodo de seguimiento de los casos desde la fecha de diagnóstico de CCR hasta Enero de 2017 el tiempo máximo de seguimiento fue de 88 meses, y ocurrieron un total de 15 eventos de muerte (15.62%), y se registraron como datos censurados 9 casos cuyo seguimiento se perdió antes del punto final de seguimiento, y de los que no se ha obtenido información sobre su posible fallecimiento (Tabla 18).

En el análisis crudo, la sobreexpresión génica de EZH2 se asocia con una menor supervivencia global con HR=2.50 (0.76-8.2) y una p=0.13, no siendo significativa esta asociación. Cuando se ajustó por edad y estadio en el análisis multivariante, el efecto de esta asociación aumenta con un valor de HR=2.66 (0.81-8.76). Dicha asociación no alcanza en ningún caso la significación estadística. En las curvas Kaplan-Meier que se muestran en la figura 20 se observa que la supervivencia OS es menor en el grupo que sobreexpresó el gen.



**Figura 20.-** Curvas Kaplan Meier para el estudio de la OS en función de la sobreexpresión génica de EZH2.

Cuando se valoró la supervivencia en función de la expresión de la proteína de EZH2, se obtuvo una asociación de dicha expresión con un mejor pronóstico y una mayor supervivencia global (OS) tanto en el análisis crudo con un valor de HR=0.55 (0.18-1.64) y una p=0.284 como en el análisis multivariante ajustado por edad y estadio con un HR=0.54 (0.17-1.70) y un valor de p=0.294. En las curvas Kaplan-Meier que se muestran en la figura a continuación se observó la supervivencia OS de los pacientes en los que se analizó la expresión proteica de EZH2 (N=74). El grupo en los que se detectó la proteína (N=49) mostraron una mayor supervivencia OS (Figura 21).



**Figura 21.-** Curvas Kaplan-Meier para el análisis de la OS en función de la expresión proteica de EZH2.

Cuando se analizó el papel de los distintos SNPs estudiados en la supervivencia global considerando un modelo codominante de herencia, en el análisis crudo del rs6958683, el genotipo GG se asoció a una menor supervivencia global con un valor de HR= 2.82 (0.35-22.69) con una p= 0.328, y el genotipo heterocigoto CG se asoció también a una menor supervivencia global con un HR=1.52 (0.53-4.38) y una p=0.437. En el análisis multivariante ajustado por edad y estadio se observó la misma tendencia, los homocigotos para G, se asociaron a una menor OS con un valor de OR=2.59 (0.31-21.29) Los heterocigotos CG, presentan un riesgo menor, este genotipo aparece asociado también a una menor supervivencia OS con un HR=1.53 (0.53-4.43) y una p=0.427, por lo que la presencia de al menos una alelo G parece estar asociado a una menor supervivencia y a un peor pronóstico (Figura 22.A).

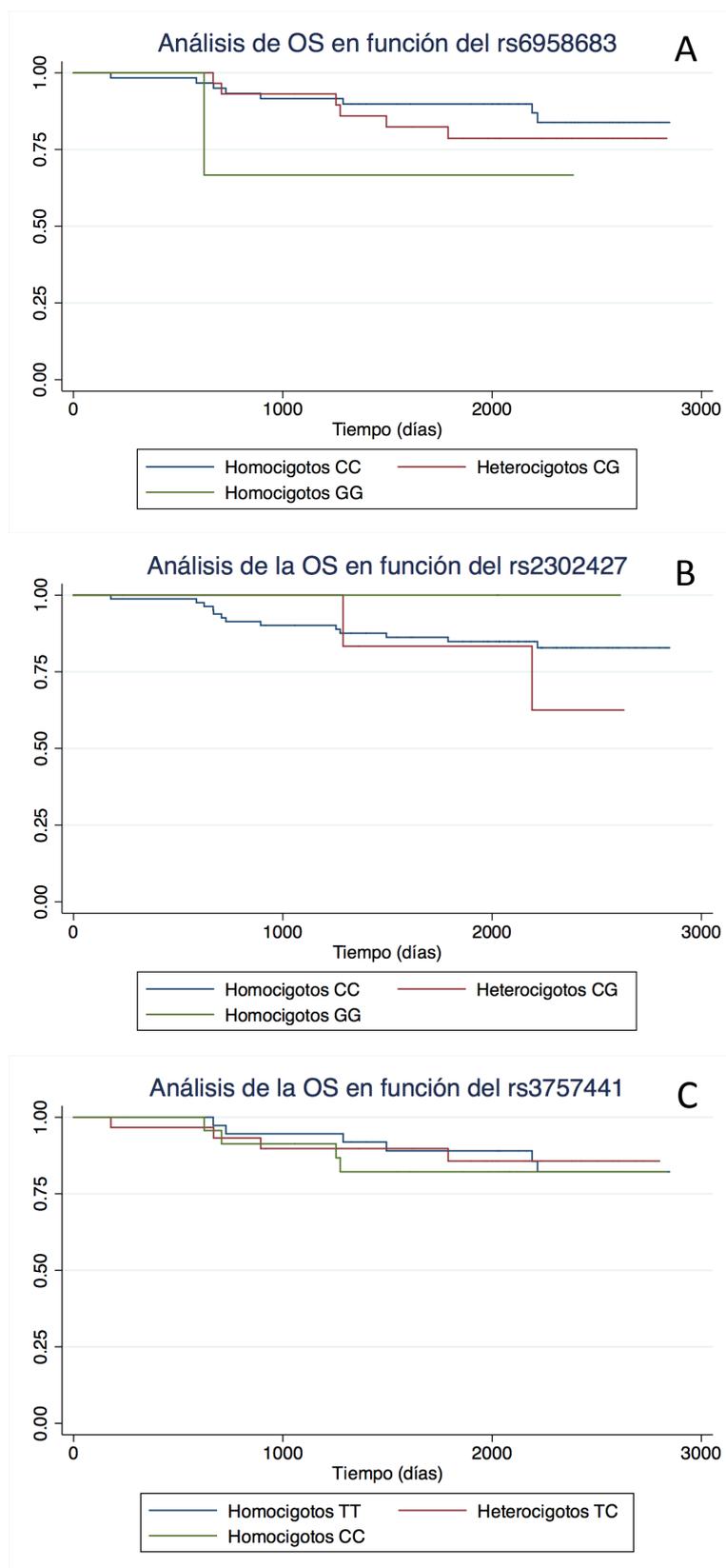
En el análisis crudo del polimorfismo rs2302427, el genotipo GG se asocia con una mayor OS,  $HR=1.66 \cdot 10^{-15}$  pero no se ha podido realizar el análisis de probabilidad ni estimar el intervalo de confianza debido a la ausencia de individuos que presenten dicho genotipo y

tengan evento de muerte, es decir la totalidad de los individuos con este genotipo se encuentran vivos al terminar de realizarse el seguimiento. Por el contrario, los individuos heterocigotos CG presentaron una menor OS con un valor de HR=1.95 (0.44-8.69) y una p=0.376. En el análisis ajustado por edad y estadio se observaron las mismas tendencias que en el análisis crudo, para el genotipo GG no existe ningún caso de *exitus* por lo que el HR que se ha estimado es HR=  $1.72 \times 10^{-15}$ , pero al no existir ningún individuo no se pudo estimar ni los intervalos de confianza ni el valor de p. Los individuos heterocigoto CG presentaron una menor supervivencia global con un HR=1.93 (0.44-8.98) y una p=0.373 (Figura 22.B).

Finalmente, en el análisis crudo del SNPs rs3757441 el genotipo CC se asoció a una menor supervivencia con un riesgo de HR=1.24 (0.34-4.41) y una p=0.734, para los heterocigotos CG el riesgo fue menor con valores HR=0.96 (0.27-3.42) y con una p=0.953. En el análisis ajustado se observó la misma asociación, el alelo C en homocigosis se relacionó con menor supervivencia global con un valor de HR=1.33 (0.36-4.82) y una p=0.661. Sin embargo los heterocigotos TC se asocian a una mayor supervivencia OS con un HR=0.85 (0.36-4.82) y una p=0.661 (Figura 22.C).

		OSc	IC95%	p	OSa	IC95%	p
<b>Sobreexpresión ARN</b>		2.5	0.76-8.2	0.13	2.66	0.81-8.76	0.107
<b>Expresión Proteica</b>		0.55	0.18-1.64	0.284	0.54	0.17-1.70	0.294
<b>rs6958683</b>	<b>CC</b>	1			1		
	<b>CG</b>	1.52	0.53-4.38	0.437	1.53	0.53-4.43	0.427
	<b>GG</b>	2.82	0.35-22.69	0.328	2.59	0.31-21.29	0.375
<b>rs2302427</b>	<b>CC</b>	1			1		
	<b>CG</b>	1.95	0.44-22.69	0.376	1.93	0.44-8.98	0.373
	<b>GG</b>	$1.66 \times 10^{-15}$	0	1	$1.72 \times 10^{-15}$	0	1
<b>rs3757441</b>	<b>TT</b>	1			1		
	<b>TC</b>	0.96	0.27-3.42	0.953	0.85	0.23-3.04	0.807
	<b>CC</b>	1.24	0.34-4.41	0.739	1.33	0.36-4.82	0.661

**Tabla 18.-** Análisis crudo y ajustado por edad y estadio de la supervivencia global OS en función de los marcadores moleculares de EZH2. OSc (modelo crudo) OSa (modelo ajustado por edad y estadio TNM).

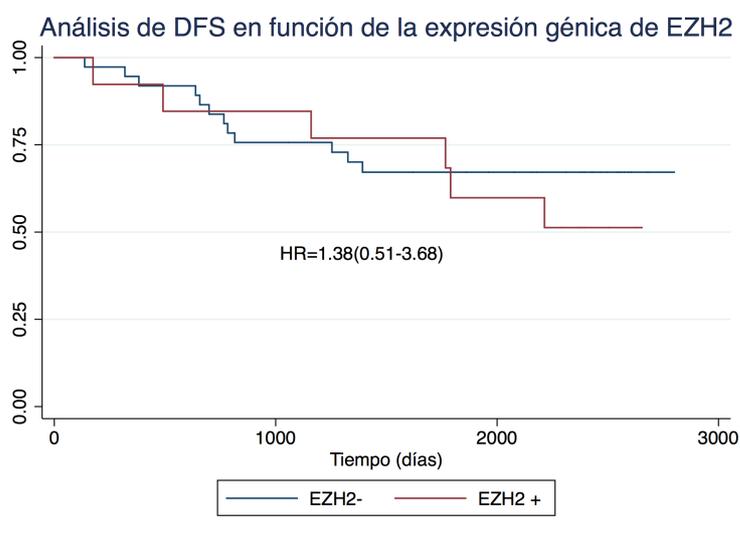


**Figura 22.-** Curvas Kaplan-Meier de la supervivencia global OS en función de los diferentes genotipos de los polimorfismos analizados: rs6958683 (A), rs2302427 (B) y rs3757441 (C).

### **Análisis de la supervivencia libre de enfermedad en función de los marcadores moleculares**

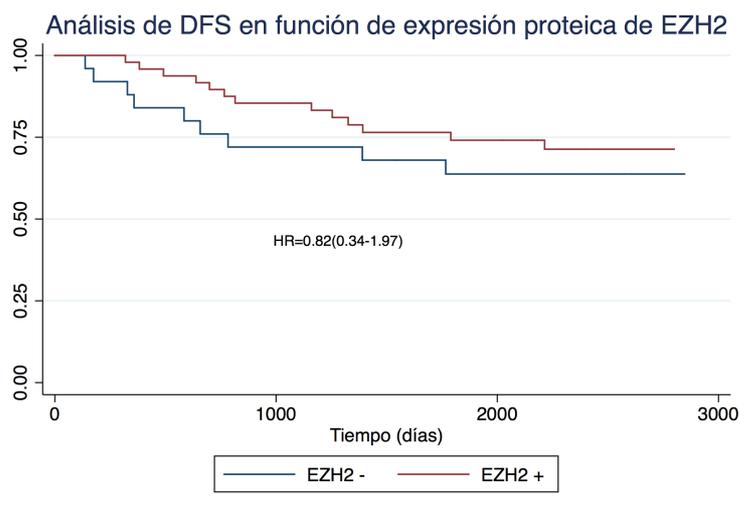
Asumiendo el mismo tiempo de seguimiento que para la supervivencia global, desde la fecha de diagnóstico hasta Enero de 2017, de los 94 pacientes con CCR, 18 (19.15%) sufren una recidiva de la enfermedad, y otros 9 fallecen durante este periodo de tiempo, por lo que en total los eventos a considerar en el modelo de análisis de supervivencia libre de enfermedad fueron 27, lo que supone el 28.72% de los pacientes a estudio (Tabla 19).

En el análisis crudo de la expresión génica se observó que los casos que sobreexpresaban EZH2 presentaron una menor DFS HR= 1.38 (0.51-3.68) y una p=0.519. En el análisis ajustado por edad y estadio, la sobreexpresión se asocia a una menor RFS, presentando un efecto similar al hallado en el análisis crudo HR=1.33 (0.49-3.56) con un valor de p=0.567 (Figura 23).



**Figura 23.-** Curvas Kaplan-Meier para la representación de la supervivencia libre de enfermedad DFS en función de la sobreexpresión del gen EZH2.

En el análisis de la DFS en función de la expresión de la proteína se observó que la presencia de la proteína se asociaba a una mayor supervivencia DFS con un HR=0.67 (0.28-1.57) y una p=0.363. Cuando se realizó este análisis ajustado por edad y estadio recodificado en dos categorías se observó la misma tendencia pero con un efecto más moderado HR=0.82 (0.34-1.97) y un valor de p=0.655. Ninguna de las asociaciones halladas para la expresión génica o proteica y la DFS fueron estadísticamente significativas (Figura 24).



**Figura 24.-** Curvas Kaplan-Meier para la representación de la supervivencia libre de enfermedad DFS en función de la expresión proteica de EZH2.

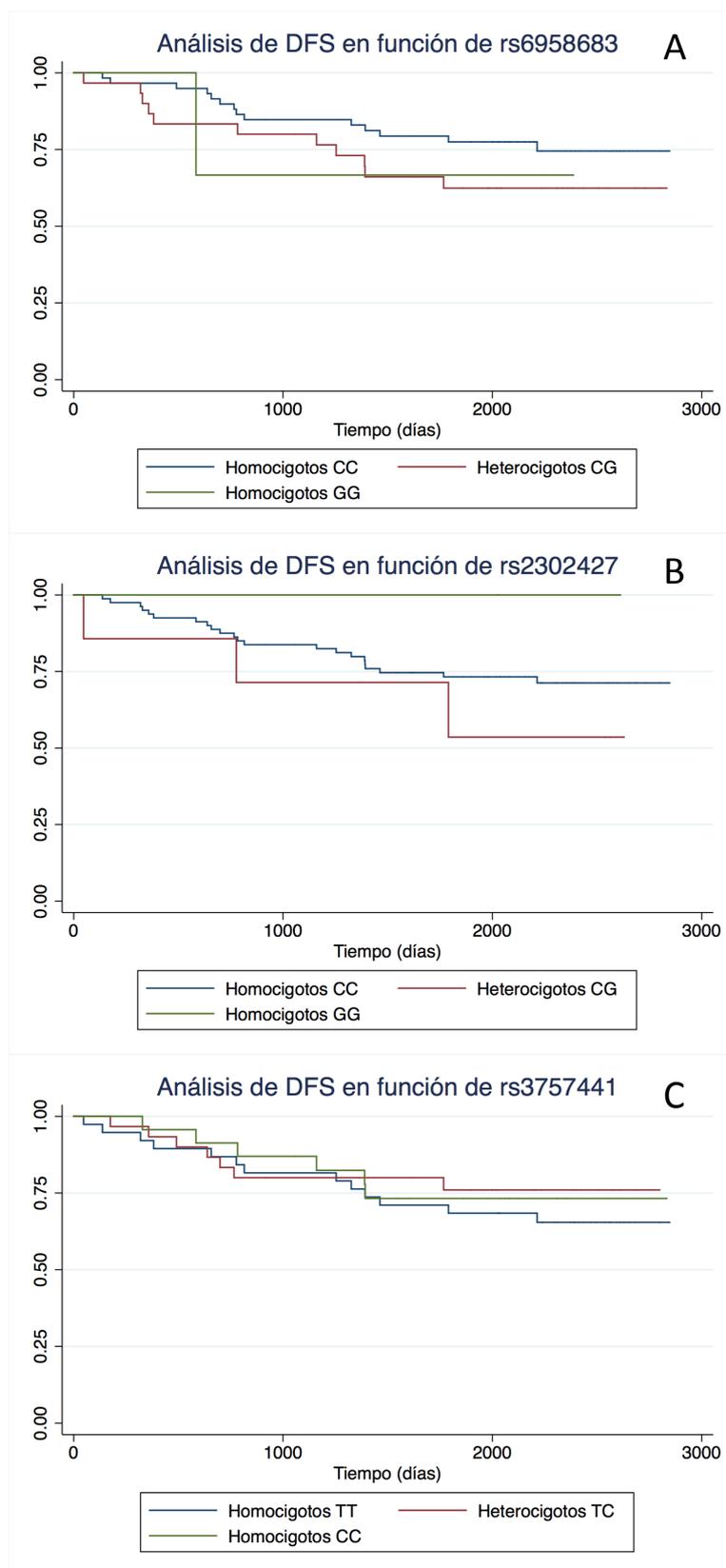
Respecto al papel en la DFS de los diferentes polimorfismos del gen EZH2, en primer lugar señalar que para el polimorfismo rs6958683, en el análisis crudo el genotipo homocigoto para el alelo G se asocia a una menor DFS con un valor  $HR=1.54$  (0.20-11.77) y una  $p=0.674$ . Se observó un riesgo mayor para los heterocigotos con genotipo CG, con un valor de  $HR=1.71$  (0.77-3.76) y una  $p=0.183$ . En el análisis multivariante, ajustado por edad y estadio, se observaron tendencias similares al análisis crudo, el genotipo homocigoto para el alelo G, se asocia a una menor DFS con un  $HR=1.34$  (0.17-10.47) y un valor de  $p=0.777$ . En los heterocigotos esta tendencia es mayor, de tal forma que se asocian a una menor supervivencia DFS con un valor de  $HR=1.79$  (0.82-3.96) y una  $p=0.194$  (Figura 25.A).

En el caso del polimorfismo rs2302427, en el análisis crudo, se observó el mismo efecto que en el caso de la OS, para el genotipo GG no se ha podido estimar los parámetros de IC95% ni el valor de  $p$ , ya que no existe ningún caso que haya sufrido una recaída en la enfermedad con ese genotipo, por tanto el riesgo que se estimó fue de  $HR=1.65 \times 10^{-15}$ . El genotipo heterocigoto al contrario, se asocia a una menor DFS, con un  $HR=1.76$  (0.52-5.88) y un valor de  $p=0.358$ . En el análisis ajustado, las tendencias observadas son similares pero el efecto del riesgo se ve aumentado y sucede lo mismo que en el análisis univariante, no se puede estimar el valor de  $p$  ni el intervalo de confianza al 95% para el genotipo GG, debido a la ausencia de individuos con este genotipo que hayan sufrido una recidiva, el HR estimado es de  $7.26 \times 10^{-16}$ . En el caso de los individuos heterocigotos presentaron una menor DFS con  $HR=2.24$  (0.66-7.58) y un valor de  $p=0.194$  (Figura 25.B).

Por último se realizó el análisis de DFS para el polimorfismo rs3757441. En el análisis crudo tanto el genotipo homocigoto para el alelo C, como el genotipo heterocigoto muestran una asociación con una mayor supervivencia DFS, con HR=0.77 (0.29-2.03) y una p=0.598 para los homocigotos CC, y un HR=0.71 (0.28-1.79) y un valor de p=0.478 para los heterocigotos TC, por lo que la presencia de al menos un alelo C se asociaría a un mejor pronóstico y una mayor supervivencia DFS. Cuando se realiza el análisis multivariante ajustado por edad y estadio, este efecto desaparece en los individuos homocigotos para el alelo C que no se observó asociación con la DFS, con valores de HR muy próximos a 1, HR=1.02 (0.38-2.72) y un valor de p=0.964, para el genotipo TC se sigue manteniendo el efecto protector observado, ya que se asoció a una mayor DFS con HR=0.60 (0.24-1.53) y un valor de p=0.291 (Figura 25.C).

		DFSc	IC95%	p	DFSa	IC95%	p
Sobreexpresión ARN		1.38	0.51-3.68	0.519	1.33	0.49-3.56	0.567
Expresión Proteica		0.67	0.28-1.57	0.363	0.82	0.34-1.97	0.665
rs6958683	CC	1			1		
	CG	1.71	0.77-3.76	0.183	1.79	0.82-3.96	0.145
	GG	1.54	0.2-11.77	0.674	1.34	0.17-10.47	0.777
rs2302427	CC	1			1		
	CG	1.76	0.52-5.88	0.358	2.24	0.66-7.58	0.194
	GG	1.65*10 <sup>-15</sup>	0	1	7.26*10 <sup>-15</sup>	0	1
rs3757441	TT	1			1		
	TC	0.71	0.28-1.79	0.475	0.60	0.24-1.53	0.291
	CC	0.77	0.29-2.03	0.598	1.02	0.38-2.72	0.964

**Tabla 19.-** Análisis crudo y ajustado por edad y estadio de la supervivencia libre de enfermedad (DFS) en función de los marcadores moleculares de EZH2. DFSc (modelo crudo) DFSa (modelo ajustado por edad y estadio TNM).



**Figura 25.-** Curvas Kaplan-Meier de la supervivencia libre de enfermedad DFS en función de los diferentes genotipos de los polimorfismos analizados: rs6958683 (A), rs2302427 (B) y rs3757441 (C).



## ***DISCUSIÓN***



El objetivo principal del presente estudio fue analizar el papel que desempeña EZH2 en el CCR. Para su desarrollo, el trabajo fue realizado en dos fases que se discuten a continuación por separado. En la primera fase, se llevó a cabo una revisión sistemática de la literatura científica orientada a analizar la asociación entre la expresión de EZH2 y la supervivencia, y se realizó un meta-análisis a partir de los datos de supervivencia global (OS) del CCR encontrados en la revisión. En la segunda fase, se investigó la expresión génica y proteica de EZH2, incluyendo una caracterización de tres variantes genéticas de tipo SNP y se analizó su asociación con características anatomopatológicas y con la supervivencia global (OS) y libre de enfermedad (DFS), en una muestra de pacientes del estudio MCC-Spain.

### **Revisión sistemática**

La implicación del EZH2 como componente de la tumorigénesis del CCR se está debatiendo actualmente, y aunque no existen muchos trabajos que hayan analizado su papel en la aparición, progresión y supervivencia de la enfermedad, en los últimos años se han reportado nuevos estudios que no habían sido incluidos en revisiones y meta-análisis previos. La ausencia de meta-análisis publicados que evalúen la participación de EZH2 en el CCR en profundidad incluyendo los artículos más recientes, y el papel contradictorio que describen los estudios observacionales similares al realizado en este trabajo, son las premisas que justifican una nueva revisión sistemática de la literatura<sup>[136,140,145,169,170,171]</sup>.

La revisión realizada permitió identificar diez artículos que evalúan la supervivencia del CCR en función de la expresión de EZH2 en tejido tumoral, empleando 8 de ellos la OS como medida de supervivencia. Estos 8 estudios fueron agrupados con el fin de realizar un meta-análisis, que incluyó un total de 1059 pacientes de CCR, para estimar la asociación entre la expresión de EZH2 y la supervivencia.

La sobreexpresión de EZH2 en el CCR se asoció de manera estadísticamente significativa a una mayor OS, con un valor de HR global=0.61 (0.38-0.84). Estos resultados fueron consistentes con los valores obtenidos por Wang *et al*<sup>[171]</sup> y Jiang *et al*<sup>[170]</sup>, en dos meta-análisis publicados que, si bien son recientes, no contienen todos los artículos incluidos en la presente revisión. Los valores estimados por dichos autores para la relación de sobreexpresión de EZH2 y la supervivencia del CCR fueron: HR global =0.91 (0.63-1.19)<sup>[171]</sup> y HR=0.75 (0.28-1.22)<sup>[170]</sup>

respectivamente. Sin embargo, un meta-análisis previo <sup>[145]</sup>, que incluía otros tumores malignos y sólo dos estudios sobre el CCR, con un total de 291 pacientes y un resultado de HR=1.13 (0.16-8.21), ofrece resultados discordantes con los obtenidos en el presente meta-análisis.

El hecho de haber encontrado datos que apuntan a un posible papel protector de la sobreexpresión de EZH2 para el CCR es destacable, dado que en otros tumores sólidos, como próstata, mama o pulmón <sup>[172-174]</sup> los resultados publicados apoyan el efecto contrario, presentándose en estos tumores la sobreexpresión de EZH2 como un marcador de mal pronóstico. Además, EZH2 aparece sobreexpresado en tejido tumoral respecto al tejido <sup>[135,136]</sup> lo que sugiere que posee un papel oncogénico relacionado con el funcionamiento de PRC2 a través de la metilación de la histona, que induciría la condensación de la cromatina reprimiendo la expresión de diferentes genes supresores tumorales <sup>[173-175]</sup>. Este proceso se asociaría, por tanto, con un peor pronóstico y menores tasas de supervivencia <sup>[135-136]</sup>. Sin embargo, estudios recientes sugieren que EZH2 podría tener un papel dual en el cáncer, actuando como un represor transcripcional, pero también como un activador <sup>[114,115,175]</sup>.

Otras posibles razones por las que los resultados de expresión de EZH2 pueden presentar estos resultados aparentemente contradictorios en la supervivencia, serían las modificaciones post-traduccionales <sup>[176]</sup>, las variaciones en la asociación con otras subunidades del complejo PRC2 como SUZ12 y EED <sup>[119,177]</sup>, la actividad de EZH2 independiente del PRC2 <sup>[119,120,177,178]</sup> o el papel de los SNPs en la expresión o en la funcionalidad <sup>[127,179,180]</sup>. Además, otros estudios han demostrado que la pérdida del estado de trimetilación de H3K27 se relaciona con menores tasas de supervivencia en cánceres como los de mama, ovario, páncreas, y recto <sup>[181,182]</sup>, lo que sería consistente con los resultados encontrados en éste meta-análisis.

Recientemente, Wassef y Margueron <sup>[114,115]</sup> han planteado que, dadas las conocidas funciones de EZH2 como regulador biológico de la expresión génica, una disminución de la actividad de PRC2 podría poseer un efecto oncogénico, puesto que se activarían numerosos genes que permanecían silentes.

Todos estos estudios apoyan la hipótesis de que existe un patrón de metilación entre las histonas, los diferentes genes (*loci* específicos) y el resto del genoma no codificante, por lo que la carcinogénesis podría darse por hipermetilación de genes específicos e hipometilación de las histonas y del genoma no codificante <sup>[72,183,186]</sup>.

Cabe destacar, como fortalezas del presente estudio, que se trata del meta-análisis más completo realizado hasta la fecha sobre el papel de EZH2 en la supervivencia de CCR. Se empleó una estrategia de búsqueda exhaustiva incluyéndose un mayor número de estudios que en otros meta-análisis publicados. Además, de un modo acorde con la guía de publicación REMARKS se realizó un análisis pormenorizado de la calidad metodológica de todos los estudios, para su selección e inclusión en la revisión sistemática.

Adicionalmente, se analizó la heterogeneidad de los estudios y los posibles sesgos de publicación. En este sentido, podemos indicar que no se encontraron sesgos de publicación empleando los métodos de Egger y de Begg, aunque la interpretación de los resultados del análisis de sesgos se ve restringida por una de las limitaciones de este trabajo, que fue el número de artículos disponibles en la literatura científica<sup>[185]</sup>. Tampoco se debe olvidar que solo se incluyeron trabajos escritos en inglés, y que la tendencia en el mundo académico es la de publicar estudios con resultados estadísticamente significativos y descubrimientos de alto impacto en revistas internacionales, mientras que los resultados negativos o menos llamativos suelen tener una difusión más limitada.

En relación a la heterogeneidad, aunque el valor de  $I^2$  fue de 29.8% y no resultó estadísticamente significativo, podemos decir que los estudios incluidos presentaron diferencias en cuanto a las características basales de los pacientes, incluyendo la edad, el estadio de la enfermedad o la duración del seguimiento. También mostraron diferencias en cuanto al uso de diferentes técnicas para evaluar la expresión de EZH2 y a la ausencia de un punto de corte común para considerar sobreexpresión en cada una de las técnicas. Todas esas fuentes de variabilidad implican limitaciones en la comparación, tal y como ha sido descrito en otros meta-análisis<sup>[170,171]</sup>.

Cabe destacar, como fuente potencial de sesgo, el método empleado para extrapolar los HR y sus respectivos intervalos de confianza. Si bien en algunos trabajos el dato se obtuvo directamente a partir de la información publicada, en otros el valor de HR tuvo que extrapolarse de las curvas de supervivencia, lo que conlleva necesariamente a hacer suposiciones sobre los datos censurados.

En conclusión, el meta-análisis realizado sugiere un papel protector de la sobreexpresión de EZH2 en la supervivencia del CCR. Sin embargo, el escaso número de estudios disponibles limita su aplicación como biomarcador de pronóstico. Son necesarios más estudios, tanto prospectivos como retrospectivos, que indaguen en el papel de la expresión de EZH2 como

marcador de supervivencia y permitan clarificar los mecanismos moleculares en los que participa para contribuir a la patogenia del CCR. En este contexto, se plantea la segunda fase de esta Tesis Doctoral cuyos resultados se discuten a continuación.

## ***Estudio de la cohorte de casos del proyecto MCC-Spain***

### ***Análisis descriptivo de la muestra***

En la segunda fase del presente trabajo se abordó el análisis de la expresión génica, la expresión proteica y el genotipado de tres polimorfismos de tipo SNP del gen *EZH2*, en una serie de casos integrada por pacientes de CCR del Área de Salud de León (ASL), que participan en el proyecto MCC-Spain. Este estudio observacional ha sido de tipo prospectivo, con un seguimiento de los casos de hasta 88 meses como máximo desde la fecha de diagnóstico. A partir de los datos obtenidos, se realizó un estudio de asociación entre los parámetros moleculares del gen *EZH2* y las características anatomopatológicas, así como un análisis de supervivencia de los pacientes del estudio.

Al describir la muestra se observó que la distribución en cuanto al sexo seguía un ratio 2:1. Este dato concuerda con la bibliografía publicada relativa al ASL en esa fecha, que coincidía con los datos sobre la incidencia de la enfermedad a nivel nacional<sup>[23-25]</sup>.

En cuanto a la edad de los casos, la media fue de 71.73 años, con un rango de 46 a 85 años, lo que coincide con otros estudios de este mismo tipo, en los que la media se sitúa en torno a los 70 años, si bien el rango varía en función de los diferentes criterios de inclusión empleados en la selección de la población estudiada<sup>[15,45,46]</sup>. La media de edad es, sin embargo, diferente entre hombres y mujeres (72.66 años en hombres con un rango de 50.53 a 85.34 años frente a 69.79 años en mujeres con un rango de 46 a 84.13 años). Esto concuerda con los datos publicados por Galcerán *et al*, que describen cómo la incidencia del CCR se asocia al género y a la edad. Así, a partir de los 50 años, el riesgo de presentar cualquier tipo de cáncer sería superior en hombres; en el caso del CCR, observaríamos estas diferencias en la incidencia a partir de los 70 años, incrementándose con la edad y observándose cómo dicha diferencia aumenta entre ambos sexos de forma progresiva<sup>[22]</sup>.

En lo que se refiere a la localización anatómica del tumor, en nuestra muestra se observó que el mayor porcentaje de casos de CCR, 85.11%, aparece en el colon, frente a un 14.89%,

que se desarrollan en el recto. Estos datos no son representativos de la población de estudio, integrada, como se ha comentado, por pacientes del ASL que aceptaron colaborar en el proyecto MCC-Spain: en una publicación previa de carácter descriptivo, que incluyó a 388 pacientes<sup>[187]</sup> la distribución por localización anatómica fue la siguiente: 28.5% en colon derecho, 32.1% en colon izquierdo, y 39,1% en recto, siendo ésta la localización más frecuente. Esta discordancia es debida fundamentalmente a que, en el momento en el que se desarrolló la fase de obtención de muestras biológicas para el proyecto MCC-Spain, ya se estaba implantando en el Hospital Universitario de León una nueva técnica quirúrgica para cáncer de recto, que implicaba la exéresis del mesorrecto, y cuyo protocolo de fijación completa en formol limitaba la obtención de tejido congelado a aquellas piezas que no se operaban por este procedimiento. Por ello, la mayor parte de los casos de los que se pudo obtener muestra de tejido congelado correspondieron a cáncer de colon.

En cuanto a la distribución por estadio, tenemos una sobrerrepresentación de los estadios II y III, y una infrarrepresentación de los estadios I y IV. Podemos atribuir esta variación, por un lado, a que la participación voluntaria en el estudio puede introducir un sesgo de selección diferencial, que provocaría que los estadios más avanzados estén infrarrepresentados, ya que el desfavorable estado de salud del paciente mermaría su predisposición a participar en el proyecto. Por otro lado, el estadio I suele coincidir con tumores de tamaño pequeño, por lo que no siempre es posible recoger muestra excedente para investigación. Finalmente, en el estadio IV la gravedad de la enfermedad y el riesgo y comorbilidad asociado a la intervención influye sobre la decisión de someter al paciente a tratamiento quirúrgico y, consecuentemente, hace inviable la disponibilidad de muestras de tejido de estos pacientes.

### ***Análisis de la expresión génica de EZH2 y su asociación con variables anatomopatológicas.***

La expresión génica fue analizada en muestras pareadas tumoral y no tumoral de 51 casos de CCR, detectándose sobreexpresión de *EZH2* en el 27.45% de los casos, e inhibición de la expresión en el 7.84%, mientras que en un 64.70% de casos no se evidenció expresión diferencial. La frecuencia de sobreexpresión detectada en nuestra serie de casos es inferior a la mostrada por otros autores que han analizado este mismo gen en muestras pareadas, como Liu *et al*<sup>[136]</sup> o Mimori *et al*<sup>[137]</sup>, que estiman respectivamente un porcentaje de sobreexpresión de 41.4% y de 52.45%.

En el presente trabajo se ha empleado la metodología utilizada por Liu *et al*, incluyendo los mismos *primers* y condiciones de PCR descritas para amplificar *EZH2*, el mismo método de

cuantificación y el mismo umbral para considerar sobreexpresión del gen <sup>[159]</sup> con la intención de reproducir su diseño experimental en nuestra población. Las diferencias encontradas con Liu *et al*, pueden deberse a que se trata de dos poblaciones distintas, aunque similares en cuanto a la distribución en estadios de la enfermedad, y a que se desconocen el resto de características de los pacientes que participan en el estudio, como la edad, el sexo, el resto de características anatomopatológicas o si los pacientes han recibido tratamiento de quimioterapia previo a la obtención de la muestra<sup>[136]</sup>.

En cuanto al estudio de Mimori *et al*, el contraste con nuestros datos puede atribuirse a la diferencia metodológica al definir el umbral de sobreexpresión. Mientras que estos autores lo han definido como la mediana calculada a partir del ratio de expresión tumor/normal ( $2\Delta\Delta Ct$ ) y el valor de la mediana se corresponde con un IR de 1.45, en nuestro trabajo se han empleado criterios más restrictivos a la hora de considerar sobreexpresión del gen ( $IR > 2$ )<sup>[137]</sup>.

No se ha encontrado ninguna asociación estadísticamente significativa entre la expresión génica de *EZH2* y las características anatomopatológicas. Así, la sobreexpresión de *EZH2* no se ha asociado a una mayor invasión linfática, ni a una mayor presencia de metástasis a distancia, ni a estadios TNM más avanzados. En el trabajo de Mimori *et al*<sup>[137]</sup> tampoco hallan diferencias significativas en la asociación de la expresión génica de *EZH2* y las características anatomopatológicas descritas. Por el contrario, el trabajo de Liu *et al*<sup>[136]</sup> sí encuentra asociación significativa tanto con la afectación de los ganglios, como con la presencia de metástasis a distancia y con el estadio TNM avanzado.

Liu *et al*<sup>[136]</sup> analizaron, junto con *EZH2*, *SUZ12* y *EED*, también componentes del complejo PRC2, y observaron que su sobreexpresión conjunta se asociaba con características anatomopatológicas de mal pronóstico. Estos resultados sugieren que *EZH2* está interactuando con otros componentes del complejo PRC2, favoreciendo su actividad pro-oncogénica. En nuestra población se desconoce la expresión del resto de integrantes de este complejo. No obstante, hay que tener en cuenta que, tal y como se ha descrito recientemente, *EZH2* puede presentar actividad independiente del complejo PRC2, tanto represora como activadora, por lo que se está investigando su posible papel como gen supresor tumoral<sup>[114,115,119,120]</sup>.

### ***Análisis de la expresión proteica de EZH2 y su asociación con variables anatomopatológicas***

El análisis de la expresión proteica incluyó 74 tejidos tumorales de la serie de casos de los que se dispuso de muestra en parafina. La proteína se detectó en el 66.22% de las muestras, un valor inferior a los obtenidos por otros autores como Takawa *et al*<sup>[166]</sup> o Fornaro *et al*<sup>[138]</sup>

que encontraron expresión de EZH2 en el 91.4% y en el 86% de los casos estudiados, respectivamente. En otros trabajos, como el de Kodach *et al*<sup>[165]</sup> y el de Kurihara *et al*<sup>[168]</sup> se alcanzó el 87.5% y el 85%, respectivamente, y con unos elevados niveles de expresión en ambos casos.

En cuanto al análisis de asociación de la proteína con las características anatomopatológicas, se han observado diferentes tendencias, si bien los datos no alcanzaron significación estadística. La expresión proteica se asoció con afectación de los ganglios, pero no con la presencia de metástasis a distancia, ni a la enfermedad estadios avanzados (III y IV). Estos resultados están en línea con los publicados por Takawa *et al*<sup>[166]</sup>, que encuentran asociación significativa entre la presencia de expresión y la ausencia de metástasis a distancia. Otros autores<sup>[140,165,168]</sup> no hallaron asociaciones estadísticamente significativas entre las características anatomopatológicas y la expresión de la proteína. En ninguno de estos estudios se ha realizado el cálculo de OR y de sus intervalos de confianza como medida de asociación, ni se han recodificado las variables anatomopatológicas siguiendo los mismos criterios que en nuestro trabajo. Esto dificulta la comparación entre estudios.

En el presente trabajo sólo se valoró la presencia o ausencia de expresión de la proteína conforme a lo realizado por Takawa *et al*<sup>[166]</sup>; no obstante, otros autores establecieron diferentes *scores* para valorar y medir la expresión proteica. Esta puede ser una de las razones por las que obtenemos resultados discrepantes con dichos autores, si bien al no existir un consenso sobre que *score* es el más adecuado, no podemos considerar como inadecuado el método empleado en este trabajo<sup>[137]</sup>.

### ***Relación entre la expresión génica y proteica***

La coincidencia entre la expresión génica y proteica se evidenció en un 64.70% de los casos estudiados. Las diferencias halladas entre ambos tipos de expresión pueden deberse tanto a cuestiones metodológicas, como a mecanismos biológicos. En primer lugar, el resultado está influenciado por la técnica empleada y el tipo de valoración que se realiza. En nuestro caso, la expresión relativa de *EZH2* se cuantificó empleando PCR en tiempo real, y evaluando la expresión del gen en la muestra tumoral con respecto a su pareada no tumoral, con un valor de corte definido en el diseño experimental. En cuanto a la expresión proteica, realizada mediante técnica inmunohistoquímica sobre un *tissue array*, se valoró la presencia o ausencia de la proteína en el tumor, sin comparación con lo que ocurre en el tejido no tumoral, y sin una cuantificación del grado de expresión de la proteína.

Por otro lado, existen mecanismos biológicos que originan diferencias entre la expresión génica y proteica. En el análisis de la expresión génica estamos valorando un amplicón, generado de un transcrito que puede sufrir modificaciones post-transcripcionales o post-traduccionales que modifiquen la proteína, de modo que no sea detectada por el anticuerpo<sup>[176]</sup>. Además, la presencia de una variante génica concreta puede modular tanto la expresión génica como la proteica, e incluso en el caso de un polimorfismo que se encuentre en una región codificante, impedir que la proteína se traduzca y sea detectable<sup>[189,191]</sup>.

No se ha hallado asociación entre la expresión génica o proteica de EZH2 y las características anatomopatológicas de mal pronóstico, como estadios avanzados o metástasis a distancia en el momento del diagnóstico. No obstante, la tendencia encontrada fue que los tumores que muestran expresión inhibida o no detectada, tienen estadios TNM más elevados y mayor presencia de metástasis a distancia, lo que concuerda con lo descrito por otros autores<sup>[166]</sup>. También existen resultados discordantes que relacionan la sobreexpresión de EZH2 con una mayor invasión linfática y venosa<sup>[136]</sup>.

### ***Análisis de polimorfismos de EZH2***

Se ha descrito en publicaciones recientes basadas en estudios de GWAS (*Genome-wide association study*) cómo la variación genética influye en la predisposición o susceptibilidad al cáncer y en la respuesta a la carcinogénesis<sup>[195]</sup>. El impacto que presentan estas variaciones a nivel génico en función de la región del gen en la que se localicen (exón, intrón, 5'UTR, 3'UTR) puede influenciar la actividad del promotor que regula la expresión del gen, la conformación o estabilidad del ARNm y su localización subcelular, y la regulación de la traducción y conformación del producto proteico. Cualquiera de estos procesos podría llevar a la aparición o evolución del CCR.

La relación entre los polimorfismos del gen *EZH2* y la aparición y desarrollo del CCR es un tema que se encuentra poco estudiado. En nuestro trabajo hemos analizado tres polimorfismos de tipo SNP que fueron seleccionados en base a su localización y posible consecuencia funcional, con la intención de investigar el papel que desempeñan estas variantes y los diferentes genotipos de cada una de ellas en el CCR.

El polimorfismo rs6958683 consiste en una variante localizada en la 5'UTR (*untranslated region*), localizada aguas arriba del marco de lectura del gen, implicando como consecuencia funcional la regulación de la expresión génica. Es un polimorfismo poco estudiado del que aún no se encuentran disponibles datos sobre su distribución poblacional en el mundo. Por el

momento se encuentra en fase 3 del análisis de *1000genomes*, pero no existen estudios publicados al respecto <sup>[152-154]</sup>. En este trabajo, rs6958683 se encuentra en equilibrio H-W con una frecuencia de 0.19 para el alelo de menor frecuencia (G). Cuando analizamos la relación de este polimorfismo con las características anatomopatológicas, se observó que la presencia del alelo G se asociaba con la existencia de metástasis a distancia, y con adenocarcinomas mucinosos. En cuanto a la expresión génica y proteica, mostró una asociación inversa, sobre todo en homocigosis GG, sugiriendo que la presencia del alelo G modifica la expresión de la proteína o el producto final proteico, pero no hay, hasta el momento, estudios similares que permitan discutir en profundidad estos resultados.

rs2302427 es una variante *missense* o “mutación con cambio de sentido” que se localiza en el exón 6 de *EZH2*. La presencia del alelo G origina la incorporación de un ácido aspártico en lugar de una histidina a la proteína final, lo que podría implicar que se generase un producto proteico con función alterada. Según las bases de datos internacionales este alelo se asocia a un buen pronóstico <sup>[152-157]</sup>. Existen estudios de casos y controles recientes que han analizado esta variante en otros tumores, mostrando que los homocigotos GG presentaban un menor riesgo de carcinoma escamoso oral <sup>[188]</sup> o de carcinoma de células uroteliales <sup>[190]</sup> y sugiriendo que presenta un papel protector en esos tumores.

En nuestro estudio, aunque las frecuencias alélicas coinciden con las descritas para la población mundial <sup>[152-157]</sup>, existe un desequilibrio H-W en la muestra, y la representación de individuos con genotipo GG fue muy pequeña. Cabe destacar que ninguno de los pacientes portadores de este genotipo presentó metástasis a distancia en el momento del diagnóstico ni estadios avanzados de la enfermedad. Estos resultados son acordes a lo publicado respecto a este genotipo, sugiriendo que se asocia a un mejor pronóstico. Además, la presencia del genotipo GG se asocia con una menor expresión génica y proteica. En nuestro estudio, ningún tumor con este genotipo sobreexpresó el gen *EZH2*. Este papel protector para la tumorigénesis podría explicar la situación de desequilibrio H-W, ya que al tratarse de una muestra de pacientes de CCR, el genotipo protector sería menos frecuente que en la población general. Por otro lado, otra explicación para la baja representación de este genotipo, es que aparezca como respuesta al cáncer, estimulando mecanismos pro-apoptóticos en el tumor. Así lo sugieren Gao *et al* <sup>[189]</sup> en un estudio en el que observaron que el genotipo GG aparecía en muestras de carcinoma hepatocelular, mientras que en las mucosas adyacentes al tumor aparecían otros genotipos (CG y CC). Tras verificar estos resultados por secuenciación, concluyeron que podría tratarse de una mutación somática y sugirieron que su incorporación

es un mecanismo de defensa de la célula tumoral, que podría jugar un papel fundamental en la transición de una hiperplasia benigna al cáncer, al iniciar mecanismos protectores contra la tumorigénesis, como la apoptosis, la autofagia y la senescencia.

El polimorfismo rs3757441, es una variante localizada en el intrón 6 del gen. Una de las consecuencias más importantes de este SNP puede ser la modulación de la expresión del producto final proteico, por dos razones bien verse afectado por el “*splicing*” o procesamiento alternativo del ARNm o bien dar lugar a cambios en los sitios de unión de factores de transcripción <sup>[135]</sup>. De una u otra forma, la transcripción del gen podría verse afectada, lo que conllevaría el aumento o la disminución de la actividad de la proteína. El alelo de menor frecuencia de este SNP (C) presenta en la población mundial una frecuencia de 0.19, alcanzando un valor de 0.34 en determinadas poblaciones y de 0.25 en nuestro país. Sin embargo, en nuestra población aparece con una frecuencia de 0.42.

En nuestro estudio, el polimorfismo rs3757441 se encuentra en desequilibrio H-W, con una mayor frecuencia de individuos heterocigotos. El alelo C en homocigosis se asocia a una menor presencia de metástasis a distancia, a un menor estadio TNM y a una menor expresión proteica. Esto contrasta con lo descrito por Fornaro *et al*, <sup>[137]</sup> que asocia el genotipo CC con la sobreexpresión de la proteína y con un mal pronóstico. No obstante, en su estudio no realizaron análisis de asociación con variantes anatomopatológicas y no describieron las características de los pacientes que participan en el estudio, por lo que otros factores que se desconocen pueden estar influenciando el pronóstico. Por otro lado, los resultados de esta Tesis Doctoral sí son coincidentes con los publicados en el ya citado estudio de casos y controles en carcinoma hepatocelular de <sup>[189]</sup>, en el que los homocigotos CC presentaron un menor riesgo de desarrollar este tumor. Por lo tanto, en estos momentos resulta controvertido el papel como marcador de mal pronóstico que se ha atribuido al alelo C.

### ***Análisis de la supervivencia en función de la expresión de EZH2***

Como ya se ha expuesto en la revisión sistemática, el papel de EZH2 en la supervivencia es controvertido. Se ha señalado en diferentes puntos de la discusión que EZH2 podría presentar un papel dual en la tumorigénesis del CCR: en función del contexto molecular podría activar y/o reprimir la transcripción de numerosos genes y tanto su sobreexpresión o aumento de actividad como su inhibición podrían resultar patológicas <sup>[114,115,119,120]</sup>. También se ha comentado que existen estudios publicados que evidencian el papel del gen como marcador de mal pronóstico y menor supervivencia en patologías como mama y próstata <sup>[172,173]</sup>. Sin embargo los únicos trabajos que analizan la relación de EZH2 en la supervivencia del CCR son

los analizados en la primera fase de esta Tesis Doctoral y dos estudios de reciente aparición con resultados de signo contradictorio, puesto que en uno de ellos la sobreexpresión de EZH2 aparece como marcador de mal pronóstico y en el otro como marcador de buen pronóstico<sup>[192,193]</sup>.

En nuestra serie de casos, cuando analizamos la supervivencia en función de la expresión génica de *EZH2*, se obtuvo que su sobreexpresión se asociaba a una menor OS, con un valor de HR=2.50 (0.76-8.2) aunque no de forma significativa. Esto concuerda con lo descrito por Mimori *et al*<sup>[137]</sup> que obtuvo valores similares, HR=2.17 (0.93-5.07), único trabajo publicado hasta la fecha que analiza el papel de la expresión génica de EZH2 con la supervivencia global del CCR. En el análisis ajustado por edad y estadio confirmamos esta tendencia con un valor de HR=2.66 (0.81-8.76).

En cuanto al análisis de la DFS, se observó que la expresión génica era un factor de mal pronóstico para el CCR, aunque el efecto fue menos acusado que el observado para la OS. El análisis crudo ofreció un valor de HR=1.38 (0.51-3.68) y el ajustado por edad y estadio TNM, un valor de HR= 1.33 (0.49-3.56). Estos resultados mostraron una tendencia similar a los publicados por Liu *et al*<sup>[136]</sup>, si bien estos autores obtuvieron valores mucho más elevados tanto en el análisis crudo, con un HR= 4.18 (2.08-8.36) como en el análisis ajustado con un HR=2.52 (1.10 -5.73).

Liu *et al*<sup>[136]</sup> también observaron que la sobreexpresión conjunta de *SUZ12* y *EED* se asociaba a una menor DFS, con valores de HR=2.82 (1.48-5.37) y HR=2.69 (1.46-4.96), respectivamente. De nuevo, estos resultados apoyan la hipótesis de que EZH2 interactúa con otros componentes del complejo PRC2 favoreciendo su actividad pro-oncogénica.

La expresión proteica de EZH2 mostró asociación con una mayor OS, con un HR=0.55 (0.18-1.64). En el análisis ajustado por edad y estadio se obtuvo un valor de HR= 0.54 (0.17-1.7). Estos datos fueron similares a los encontrados por Benard *et al*<sup>[140]</sup>, que reportaron un valor de HR=0.74 (0.54-1.03)<sup>[140]</sup>, y acordes a los resultados globales del meta-análisis, con un HR=0.61 (0.38-0.84).

Al comparar los resultados de expresión proteica con los publicados en los trabajos que integran el meta-análisis, observamos concordancia con aquellos que en la revisión sistemática obtuvieron un *score* metodológico más elevado<sup>[140,168]</sup>. Más recientemente, Yamamoto *et al* encontraron que el grupo de pacientes que mostraba inhibición en la expresión proteica de EZH2 se asociaba a menor OS y DFS<sup>[191]</sup>. Otros autores han mostrado, sin embargo, resultados

de sentido contrario al de nuestro trabajo, como Kodach *et al*<sup>[165]</sup> y Wang *et al*<sup>[141]</sup>. Es necesario subrayar que la puntuación de estos trabajos en el *score* metodológico no fue muy elevada, 63% y 40% respectivamente y que, además, en cada uno de los trabajos se emplean diferentes umbrales para considerar sobreexpresión proteica lo que, a efectos comparativos, dificulta la interpretación de los resultados.

Nuestros datos muestran asociación entre la presencia de expresión proteica de EZH2 y una mayor DFS, con valores de HR=0.67 (0.28-1.57) en el análisis crudo y HR=0.82 (0.34-1.97) en el análisis ajustado, que son un poco más elevados aunque similares a los descritos por Kurihara *et al*<sup>[168]</sup>, con valores de HR =0.51 (0.29-0.77) y HR=0.46 (0.27-0.76), respectivamente. En ambos trabajos se realizó el análisis multivariante ajustando por edad y estadio TNM. Estos resultados son contrarios, sin embargo, a los descritos por Wang *et al*<sup>[141]</sup>, en los que la sobreexpresión de la proteína se asociaba a una menor DFS, con HR=3.21(1.06-9.72). La inclusión en el trabajo de Wang *et al*, de una mayor proporción de pacientes en estadio IV, junto con el uso de un *score* que valora la sobreexpresión de EZH2 en tejido tumoral respecto al no tumoral, podrían explicar las diferencias en los resultados. Asimismo, la media de edad de los pacientes de este estudio es relativamente baja (62 años), lo que hace pensar que, probablemente, se ha incluido una proporción mayor de pacientes con síndromes hereditarios, que sufren la enfermedad más tempranamente y con un peor pronóstico, aunque esto no es referido por los autores<sup>[197]</sup>.

Los resultados obtenidos en el análisis de OS y DFS en función de la expresión génica y proteica de EZH2 variaron en función de la técnica empleada, acorde a lo descrito en la revisión sistemática realizada en la primera fase de este trabajo. Una posible explicación biológica a este hecho, es que para cada gen existe un conjunto particular de eventos que controlan la expresión, modificaciones postranscripcionales y/o modificaciones post-traduccionales que afecten a la detección de la proteína; además, la expresión del ARNm dependerá del tejido en el que se encuentre y de la concentración de ARNm disponible para su traducción a proteína, que es variable en función del momento biológico, si se está sintetizando o degradando o si es estable y funcional en el citoplasma.<sup>[119,120]</sup> Desde un punto de vista metodológico, también pueden existir diferencias a la hora de valorar el papel como marcador de pronóstico del CCR y emplear como medida de supervivencia la OS.

La OS contabiliza los fallecimientos independientemente de la causa, mientras que la DFS contabiliza los fallecimientos y las recaídas durante la enfermedad. Cuando se analizó la OS en función de la expresión génica y proteica se encontraron diferencias notables entre los

resultados de ambos tipos de expresión. Sin embargo, cuando se analizó la DFS, estas diferencias se redujeron. Estos datos, junto con que no se observaron diferencias entre ambos tipos de expresión en función de las características anatomopatológicas relacionadas con el pronóstico, como el estadio TNM o la presencia de metástasis a distancia, implicaría que quizá la OS, no sea una buena medida de supervivencia para valorar el efecto de un marcador molecular como la expresión de EZH2 en el CCR.

### ***Análisis de supervivencia en función de los polimorfismos de EZH2***

El análisis del polimorfismo rs6958683 muestra que, en nuestra población, tanto el genotipo homocigoto (GG) como heterocigoto (CG) están asociados a una menor OS y a una menor DFS. Por tanto, la presencia del alelo G parece asociarse a un mal pronóstico en el CCR. Esta variante genética sólo ha sido analizada en un estudio previo en CCR<sup>[179]</sup>, pero los autores no encontraron asociación con la supervivencia. Por otra parte, las bases de datos de proyectos de genotipado no registran información sobre sus posibles consecuencias clínicas.

En lo que se refiere al polimorfismo rs2302427, como ya se ha comentado, ninguno de los pacientes portadores del alelo G en homocigosis falleció a lo largo del seguimiento, lo que sugiere su relación con un papel protector frente al CCR, en concordancia con la información recogida en la base de datos<sup>[152-157]</sup> sobre la consecuencia clínica de este alelo. En carcinoma hepatocelular se ha encontrado relación entre el genotipo GG de este SNP y una mayor OS y DFS.<sup>[189]</sup> Un mecanismo que se ha sugerido para explicar el posible efecto protector del genotipo GG es que puede jugar un papel antagonista en el desarrollo del cáncer, manteniendo e induciendo en las células tumorales los procesos celulares de apoptosis, autofagia y senescencia<sup>[189]</sup>.

En el caso del último polimorfismo analizado rs3757441, el genotipo CC se asoció en el análisis crudo con una mayor OS HR=0.77 (0.29-2.03) y DFS HR=0.77 (0.51-3.68). Esta tendencia varía cuando se ajusta por edad y estadio, pasando a obtenerse valores de HR = 1.33 (0.36-4.82) en la supervivencia OS y de HR=1.02 (0.38-2.72) en la DFS. Estos resultados discrepan con los mostrados por Crea *et al*, que encontraron que el genotipo CC se asociaba tanto a una menor OS con HR=2.85 (1.02-7.93) como a una menor DFS con HR=2.21 (1.05-4.62)<sup>[179]</sup>. Estas diferencias podrían ser debidas a que la muestra estudiada por estos autores estaba integrada por pacientes en un estadio avanzado de la enfermedad, con CCR metastásico. En contraste, en nuestra serie de casos hay representación de todos los estadios TNM. Debe recordarse que un estadio avanzado de la enfermedad es un predictor independiente de mal pronóstico y menor supervivencia<sup>[152-157]</sup>.

En definitiva, en el análisis de la supervivencia en función de la expresión de EZH2 y de sus polimorfismos en CCR realizado en este trabajo, se han encontrado diferentes tendencias. Por un lado, la sobreexpresión génica de EZH2 se ha asociado a una menor supervivencia, mientras que la presencia de la proteína se ha asociado a un buen pronóstico y una mayor supervivencia. Por otro lado, en el análisis de polimorfismos, tanto rs2302427 como rs6958683 parecen tener un papel relevante en el pronóstico de la enfermedad. Así, el genotipo GG de rs2302427 desempeñaría un papel protector, predictor de mayor supervivencia y el genotipo GG del rs6958683 se presentaría como un genotipo de riesgo, asociándose a una menor supervivencia de la enfermedad. No se ha encontrado asociación del SNP rs3757441 con la supervivencia del CCR.

Se ha señalado que el papel de la expresión de EZH2 es contradictorio en función de la técnica empleada y eso se ha visto reflejado en la literatura científica. La expresión del gen puede verse afectada por modificaciones post-transcripcionales o post-traduccionales como la fosforilación, ubiquitinación y los procesos degradativos de proteínas o proteólisis, que pueden modificar la detección de la proteína<sup>[176]</sup>. Es necesario ahondar en la relación entre el fenotipo de la proteína y la expresión génica, con técnicas que implementen este estudio, permitiendo cuantificar la proteína aunque el estudio del proteoma sea más complejo que el estudio del genoma debido a todas las modificaciones existentes, como el proceso de editado del ARN, los sitios alternativos de iniciación y *splicing* además de las modificaciones postraduccionales. Por este motivo, se precisan más avances con la incorporación de las nuevas tecnologías que nos permitan cuantificar las proteínas y además e identificar su localización celular, las modificaciones y sus funciones<sup>[196]</sup>. Técnicas como la espectrometría de masas y la cromatografía nos permitirán concretar los mecanismos que se producen y esclarecer la relación y las discrepancias entre las diferencias entre expresión génica de EZH2 y expresión proteica.<sup>[196,198,199]</sup>

Se han observado notables diferencias en los análisis en función de la medida de supervivencia utilizada. La OS sólo refleja si la persona ha fallecido o si se ha perdido su seguimiento, no se especifica la causa de la muerte y no se incluyen eventos de recaída en la enfermedad. Sin embargo, en el análisis de la DFS se incluyen los eventos de recidiva. En nuestro trabajo, al emplear la DFS se ha observado cómo los resultados y tendencias encontradas en el análisis de OS se moderan, por lo que quizá la OS sea una medida de supervivencia insuficiente para establecer el efecto predictivo de un marcador molecular.

Por último, se ha observado en el análisis de polimorfismos del gen *EZH2*, que los genotipos homocigoto GG del rs6958386 y homocigoto GG del rs2302427 tienen un efecto predictor en la supervivencia, y una relación de asociación inversa con la expresión del gen y de la proteína. Sería interesante reproducir estos resultados en una muestra de características similares para validar este efecto en el pronóstico, y comprobar la relación de estos polimorfismos con la expresión génica y proteica de *EZH2* en otra población similar a la que ha participado en este estudio.

### ***Limitaciones y fortalezas***

Al tratarse de un estudio basado en una serie de casos, las asociaciones detectadas se encuentran sometidas a las limitaciones propias de este tipo de estudios, a la hora de establecer la causalidad en dichas asociaciones. Además, la recogida de información a partir de las historias clínicas ha supuesto pérdidas que pueden introducir sesgos que limiten la validez de los resultados obtenidos. Otra de las principales limitaciones fue el tamaño de la muestra, careciendo del poder suficiente para examinar, en algunos casos, el efecto de la estratificación.

Existen varios factores que han podido introducir sesgos de selección en el estudio, como la disponibilidad de muestra biológica excedente del proceso diagnóstico: el procesamiento asistencial rutinario de aquellas piezas de recto en las que se realizó resección total, no permitió la obtención inmediata de tejido para su congelación, y de ahí la escasa representación de casos con esta localización; tampoco se pudo analizar el ARNm ni la proteína de todos los individuos (N=94), sino sólo de 51 muestras, en el caso del ARNm y 74 en el caso de la proteína, y aunque la selección se realizase al azar, se ha podido introducir un sesgo en el estudio.

Otras limitaciones, de tipo metodológico, se relacionan con el proceso de recabar información a partir de las historias clínicas. La causa de la muerte no siempre ha quedado reflejada, y en el caso particular de tres pacientes, se consideró incluirlos en el modelo en lugar de censurarlos, porque se tuvo constancia de su ingreso en la unidad de cuidados paliativos. Este hecho ha podido introducir sesgos en el análisis de DFS.

A la hora de realizar el diseño experimental para valorar la expresión génica diferencial de *EZH2*, además de haber empleado un tejido no tumoral pareado con la muestra tumoral, la comparación con el tejido de una persona sana habría supuesto una aproximación más cercana a la realidad. El tejido adyacente al área tumoral, puede contener alteraciones moleculares, así como niveles de expresión de *EZH2* alterados, lo que podría implicar la

presencia de un sesgo en la cuantificación de la expresión. En un estudio de Sanz-Pamplona *et al*, se describe como se ha detectado en muestras de mucosa adyacente al tumor, expresión aberrante de numerosos genes que no estaban presentes en mucosas sanas de personas no enfermas. Esos patrones alterados pueden ser inducidos por el propio tumor, el microambiente tumoral o como una respuesta contra el mismo<sup>[200]</sup>.

Finalmente, en el estudio de la expresión de la proteína, se ha discriminado entre ausencia y presencia de expresión. Otros autores emplean diferentes *scores*, estableciendo categorías según el patrón de expresión de la proteína. Por eso, en nuestro trabajo no hemos hablado de sobreexpresión proteica sino simplemente de presencia o ausencia de expresión de la proteína.

La primera fortaleza que señalamos en este trabajo ha sido la realización, en una primera fase del diseño experimental, de una revisión sistemática de la literatura científica y un meta-análisis, para evaluar el papel del gen *EZH2* en el pronóstico del CCR, incluyendo un análisis de sesgos de publicación. Los resultados obtenidos sientan unas premisas que se han confirmado en la segunda fase, tras la realización de la parte experimental de esta Tesis Doctoral.

Como fortaleza de esta segunda fase, podemos señalar la disponibilidad de una muestra poblacional con una buena caracterización clínica y epidemiológica, con todos los datos anatomopatológicos necesarios para el diagnóstico y la estadificación, similar a los realizados por otros estudios enfocados en el estudio de *EZH2* como biomarcador en CCR.

El presente trabajo incluye el análisis de tres variantes polimórficas del gen, así como de su expresión génica y proteica. Es el único estudio que realiza los tres análisis, y aborda la relación entre el genotipo y el fenotipo de *EZH2*. Serían necesarios estudios adicionales de la expresión de *EZH2* y de los polimorfismos en muestras independientes para confirmar las tendencias encontradas. Además, el estudio de otros polimorfismos, así como de la expresión de las otras subunidades del PRC2, ayudarían a comprender mejor el papel de *EZH2* en la progresión y supervivencia del CCR.

En este caso, la importante colección de muestras biológicas obtenidas y toda la información clínica recogida durante este periodo, suponen un punto de partida para futuros experimentos que permitan un mejor entendimiento del papel de *EZH2* en el CCR.

## ***CONCLUSIONES***



1. El efecto de la sobreexpresión de EZH2 en la supervivencia global OS del CCR resulta protector, estimado a partir de la realización de un meta-análisis realizado de la literatura científica con un HR global ajustado de HR de 0.61 (0.38-0.84).
2. La sobreexpresión génica de EZH2 se asocia a estadios menos avanzados y con ausencia de metástasis a distancia en la serie de casos analizada, se asocia a una menor supervivencia OS con HR=2.50 (0.76-8.2) en el análisis crudo y HR=2.66 (0.81-8.76) en el análisis ajustado y a una menor supervivencia DFS con HR=1.38 (0.51-3.68) en el análisis crudo y HR= 1.33 (0.49-3.56) en el análisis ajustado.
3. La expresión de la proteína se asocia a estadios menos avanzados y menor presencia de metástasis a distancia, a una mayor supervivencia OS HR=0.55 (0.18-1.64) en el análisis crudo y en el ajustado HR= 0.54 (0.17-1.7) y a una mayor supervivencia DFS con HR=0.67 (0.28-1.57) en el análisis crudo y HR=0.82 (0.34-1.97) en el análisis ajustado.
4. El alelo G del polimorfismo rs6958683 se asocia con mayor presencia de metástasis a distancia OR=1.52(0.55-2.26) y con los adenocarcinomas mucinosos OR=11.11(0.91-135.8). Los portadores del genotipo GG presentaron menos expresión proteica y no sobreexpresan el gen y se asociaron a una menor OS=2.82 (0.35-22.69) y a una menor DFS=1.54(0.2-11.77)
5. En el análisis del polimorfismo rs2302427,ninguno de los portadores del genotipo GG muestran metástasis a distancia ni estadios avanzados en el momento del diagnóstico. Esta variante asimismo parece asociarse a una mayor supervivencia global (OS) y a una mayor supervivencia libre de enfermedad (DFS) siendo un posible marcador de buen pronóstico.
6. En el análisis del polimorfismo rs3757441, el genotipo CC parece ser protector frente a la OS, pero apenas si muestra efecto sobre la DFS. Este alelo C no se asocia a una mayor sobreexpresión del gen *EZH2* ni de la proteína.



## ***BIBLIOGRAFÍA***



1. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100(1):57-70.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144(5):646-74.
3. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF et al. SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. Bethesda, MD, [https://seer.cancer.gov/csr/1975\\_2013/](https://seer.cancer.gov/csr/1975_2013/), basado en datos enviados a SEER en Noviembre de 2015, publicados en la web de SEER en Abril de 2016.
4. American Cancer Society. Cancer Facts & Figures 2017. Atlanta, Ga: American Cancer Society; 2017.
5. Colina F, Ibarrola C. Protocolo e información sistematizada para los estudios histopatológicos relacionados con el carcinoma colorrectal. *Rev Esp Patol* 2004;37(1):73-90.
6. Fernández W, Gaete F, Nilsen E, Olave E, Peñaloza P, Pisano R; Zapata L. Protocolo para el procesamiento de biopsias en cáncer colorrectal. *Manual de Protocolos para Anatomía Patológica 2010-2011*.81:95.
7. Oncoguía del Cáncer colorrectal de la Comunidad Valenciana.2007
8. Bosman FT; World Health Organization;International Agency for Research on Cancer. WHO classification of tumours of the digestive system Lyon : IARC Press, 2010
9. Matthew Fleming, Sreelakshmi Ravula, Sergei F. Tatishchev, Hanlin L. Wang Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol* 2012; 3(3): 153–173
10. Robbins y Cotran. Atlas de Anatomía patológica. Capítulo 7 - Tubo digestivo, Pages 153-193 Edited by:Edward C. Klatt, MD (ISBN: 978-84-8086-275-2) 2007
11. Pritchard DJ, Korf BR. *Genética Médica: Lo esencial de un vistazo*.3ª edición 2015.
12. Tariq K, Ghias K. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol Med*. 2016 Mar; 13(1): 120–135
13. Raskov H, Pommergaard HC, Burcharth J, Rosenberg J. Colorectal carcinogenesis-update and perspectives. *World J Gastroenterol* 2014; 20(48): 18151–18164.
14. Varios autores. Libre pathology. Disponible en <https://librepathology.org/> (Acceso: 10/07/2017).
15. Hemminki K, Santi I, Weires M, Thomsen H, Sundquist J, Bermejo JL. Tumor location and patient characteristics of colon and rectal adenocarcinomas in relation to survival and TNM classes. *BMC Cancer* 2010,10:688.
16. Sobin LH. TNM: evolution and relation to other prognostic factors. *Semin Surg Oncol* 2003;21(1):3-7

17. Greene FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG et al. American Joint Committee on Cancer: AJCC Cancer Staging Manual. 6th ed, New York, NY, USA: Springer, 2002:157-164
18. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, AJCC cancer staging manual, 7th ed.2010:143-164.
19. Obrocea FL, Sajin M, Marinescu EC, Stoica D. Colorectal cancer and the 7th revision of the TNM staging system: review of changes and suggestions for uniform pathologic reporting. Rom J Morphol Embryol 2011;52(2):537-544.
20. Compton C, Fenoglio-Preiser CM, Pettigrew N, Fielding LP. American Joint Committee on Cancer Prognostic Factors Consensus Conference: Colorectal Working Group. Cancer 2000,88(7):1739-1757.
21. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F.GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet].Lyon, France: International Agency for Research on Cancer; 2013. <http://globocan.iarc.fr>
22. Galceran J, Ameijide A, Carulla M, Mateos A, Quirós J, Rojas D et al . REDECAN Working Group . Cancer incidence in Spain, 2015. Clin Transl Oncol. 2017;19(7):799-825.
23. Lopez-Abente G, Ramis R, Pollán M, Aragonés N, Pérez-Gómez B, Gómez-Barroso D et al. Atlas municipal de mortalidad por cáncer en España 1989-1998. Instituto de Salud Carlos III. Madrid, 2006.
24. Martín Sánchez V, Muínelo Voces MI, Jorquera Plaza F, Molina De La Torre AJ, De Abajo Olea S, Tamames Gómez S et al. Tendencia y distribución municipal de la incidencia de cáncer colorrectal en el área de salud de León (1994-2008). Gastroenterol Hepatol 2012; 35(5):299-308.
- 25 López-Abente G, Aragonés N, Pérez-Gómez B, Pollán M, García-Pérez J, Ramis R et al. Time trends in municipal distribution patterns of cancer mortality in Spain. BMC Cancer. 2014;24;14:535.
26. Pérez Cueto FJ. Colorectal cancer prevention with diet and physical activity: recent update. Nutr Hosp 2011;26(6):1491
27. Wu K, Feskanich D, Fuchs CS, Chan AT, Willett WC, Hollis BW et al. Interactions between plasma levels of 25-hydroxyvitamin D, insulin-like growth factor (IGF)-1 and C-peptide with risk of colorectal cancer. PLoS One. 2011;6(12):e28520.
28. Lee DH, Keum N, Giovannucci EL. Colorectal Cancer Epidemiology in the Nurses' Health Study. Am J Public Health. 2016;106(9):1599-607.

29. Namasivayam V, Lim S. Recent advances in the link between physical activity, sedentary behavior, physical fitness, and colorectal cancer. *F1000Res* 2017 Mar;1;6:199.
30. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ* 2016;9;354:i3857.
31. Halle M, Schoenberg MH. Physical activity in the prevention and treatment of colorectal carcinoma. *Dtsch Arztebl Int* 2009;106(44):722-7.
32. Schoenberg MH. Physical Activity and Nutrition in Primary and Tertiary Prevention of Colorectal Cancer. *Visc Med.* 2016;32(3):199-204.
33. 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;1;125(1):171-80.
34. Asano T., McLeod R.S. Dietary fibre for the prevention of colorectal adenomas and carcinomas. *Cochrane database of systematic reviews.* 2002;2.
35. Hague A., Manning A.M., Hanlon K.A., Huschtscha L.I., Hart D., Paraskeva C. Sodium butyrate induces apoptosis in human colonic tumour cell lines in a p53-independent pathway: implications for the possible role of dietary fibre in the prevention of large-bowel cancer. *Int J Cancer.* 1993;55:498–505
36. Jarosz M., Sekuła W., Rychlik E. Trends in dietary patterns, alcohol intake, tobacco smoking, and colorectal cancer in Polish population in 1960-2008. *BioMed Research International* 2013:183204
37. Moskal A, Freisling H, Byrnes G, Assi N, Fahey MT, Jenab M et al. Main nutrient patterns and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition study. *Br J Cancer* 2016;22;115(11):1430-1440.
38. Shin S, Saito E, Sawada N, Ishihara J, Takachi R, Nanri A et al; JPHC Study Group. Dietary patterns and colorectal cancer risk in middle-aged adults: A large population-based prospective cohort study. *Clin Nutr* 2017 May 5. pii: S0261-5614(17)30154-1.
39. Mehta RS, Song M, Nishihara R, Drew DA, Wu K, Qian ZR et al. Dietary Patterns and Risk of Colorectal Cancer: Analysis by Tumor Location and Molecular Subtypes. *Gastroenterology* 2017;152(8):1944-1953.
40. Hagggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg.* 2009;22(4):191-7.
41. Jasperson KW, Tuohy TM, Neklason DW, Burt RW Hereditary and familial colon cancer. *Gastroenterology* 2010 ;138(6):2044-58.

42. Hamilton SR, Liu B, Parsons RE, Papado-poulos N, Jen J, Powell SM et al. The molecular basis of Turcot's syndrome. *N Engl J Med* 1995;332:839-47.12
43. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 42, D1001–1006.
44. Stigliano V, Sanchez-Mete L., Martayan A, Anti M. Early-onset colorectal cancer: a sporadic or inherited disease? *World J Gastroenterol* 2014; 20(35): 12420–12430.
45. Westra JL, Plukker JT, Buys CH, Hofstra RM. Genetic alterations in locally advanced stage II/III colon cancer: a search for prognostic markers. *Clin Col Cancer* 2004;4 (4):252-259.
46. Tepjar S, Bertagnolli M, Bosman F, Lenz HJ, Garraway L, Waldman F et al. Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery. *The Oncologist* 2010; 15:390-404
47. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci* 1971; 68: 820–823.
48. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61(5):759-67.
49. Jasperson K, Burt RW. The Genetics of Colorectal Cancer. *Surg Oncol Clin N Am* 2015;24(4):683-703.
50. Vogelstein B, Fearon ER, Kern SE, Hamilton SR, Preisinger AC, Nakamura Y et al. Allelotype of colorectal carcinomas. *Science*;1989;244: 207-211.
51. Kaemmerer E, Klaus C, Kyung Jeon M, Gassler N. Molecular classification of colorectal carcinomas: The genotype-to-phenotype relation. *World J Gastroenterol* 2013;19:8163-7.
52. Lynch HT, Lynch J. Lynch syndrome: genetics, natural history, genetic counseling, and prevention. *J Clin Oncol* 2000;1;18(21 Suppl):19S-31S.
53. Maehama T, Dixon JE. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J Biol Chem* 1998;29;273(22):13375-8.
54. Riegert-Johnson DL, Gleeson FC, Roberts M, Tholen K, Youngborg L, Bullock M et al. Cancer and Lhermitte-Duclos disease are common in Cowden syndrome patients. *Hered Cancer Clin Pract* 2010;17;8(1):6
55. You YN, Wolff BG, Boardman LA, Riegert-Johnson DL, Qin R Peutz-Jeghers syndrome: a study of long-term surgical morbidity and causes of mortality. *Fam Cancer*;2010;9(4):609-16.
56. Hemminki A. The molecular basis and clinical aspects of Peutz-Jeghers syndrome. *Cell Mol Life Sci* 1999;55(5):735-50.
57. Stoffel EM Heritable Gastrointestinal Cancer Syndromes. *Gastroenterol Clin North Am* 2016;45(3):509-27.

58. Chun P, Waingberg ZA. Adjuvant chemotherapy for stage II colon cancer: The role of molecular markers in choosing therapy. *Gastrointest Cancer Res* 2009;3:191-196.
59. Wang W, Wang GQ, Sun XW, Chen G, Li YF, Zhang LY et al. Prognostic values of chromosome 18q microsatellite alterations in stage II colonic carcinoma. *World J Gastroenterol* 2010;16(47):6026-6034.
60. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. *Gastroenterology* 2010 Jun;138(6):2059-72.
61. Lengauer C, Kinzler KW, Vogelstein B. DNA methylation and genetic instability in colorectal cancer cells *Proc Natl Acad Sci* 1997;18;94(6):2545-50.
62. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat Genet* 2011;43(10):964-8.
63. Castillo A, Morse HC, Godfrey VL, Naeem R, Justice MJ. Overexpression of Eg5 causes genomic instability and tumor formation in mice. *Cancer Res* 2007;67(21):10138-47.
64. Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* 2009;460(7252):278-82
65. Gertler R, Rosenberg R, Stricker D, Friederichs J, Hoos A, Werner M et al. Telomere length and human telomerase reverse transcriptase expression as markers for progression and prognosis of colorectal carcinoma. *J Clin Oncol* 2004;22(10):1807-14.
66. Baker SJ, Preisinger AC, Jessup JM, Paraskeva C, Markowitz S, Willson JK et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* 1990;50(23):7717-22
67. Carmona FJ, Esteller M. Epigenomics of human colon cancer. *Mutation Research* 2010;693:53-60
68. Laiho P, Launonen V, Lahermo P, Esteller M, Guo M, Herman JG et al. Low level microsatellite instability in most colorectal carcinomas. *Cancer Res* 2002;62:1166-1170.
69. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;21(11):1350-6
70. Müller MF, Ibrahim AE, Arends MJ Molecular pathological classification of colorectal cancer. *Virchows Arch* 2016 ;469(2):125-34
71. Waddington CH. The epigenotype 1942. *Int J Epidemiol* 2012;41(1):10-3.
72. Esteller M. Cancer epigenetics: DNA methylation and chromatin alterations in human cancer. *Adv Exp Med Biol* 2003;532:39-49
73. Novak K. Epigenetics changes in cancer cells. *Med Gen Med.* 2004;20;6(4):17.

74. Rodenhiser DI. Epigenetic contributions to cancer metastasis. *Clin Exp Metastasis* 2009;26(1):5-18.
75. Jovanovic J, Rønneberg JA, Tost J, Kristensen V. The epigenetics of breast cancer. *Mol Oncol* 2010;4(3):242-54
76. Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Mol Biotechnol* 2010;44(1):71-81.
77. Schnekenburger M, Diederich M. Epigenetics Offer New Horizons for Colorectal Cancer Prevention. *Curr Colorectal Cancer Rep* 2012;8(1):66-81.
78. Das PM, Singal R. DNA methylation and cancer. *J Clin Oncol* 2004;22(22):4632-42.
79. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* 2011;18;8(12):686-700.
80. Suzuki K, Suzuki I, Leodolter A, Alonso S, Horiuchi S, Yamashita K et al. Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. *Cancer Cell* 2006;9(3):199-207.
81. Matsubara N Epigenetic regulation and colorectal cancer. *Dis Colon Rectum* 2012;55(1):96-104.
82. Roy S, Majumdar A. Cancer Stem Cells in Colorectal Cancer: Genetic and Epigenetic Changes. *J Stem Cell Res Ther* 2012;17;7(6) pii: 1034.
83. Choong MK, Tsafnat G. Genetic and epigenetic biomarkers of colorectal cancer *Clin Gastroenterol Hepatol.* 2012;10(1):9-15.
84. van Engeland M, Derks S, Smits KM, Meijer GA, Herman JG. Colorectal cancer epigenetics: complex simplicity. *J Clin Oncol* 2011;1;29(10):1382-91.
85. Lopes EC, Valls E, Figueroa ME, Mazur A, Meng FG, Chiosis G. A Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. *Cancer Res* 2008;15;68(18):7258.
86. Jones PA. DNA methylation and cancer. *Oncogene* 2002;12;21(35):5358-60.
87. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* 2012; 6;150(1):12-27.
88. Spange S, Wagner T, Heinzl T, Krämer OH. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int J Biochem Cell Biol* 2009;41(1):185-98.
89. Chevillard-Briet M, Quaranta M, Grézy A, Mattera L, Courilleau C, Philippe M et al. Interplay between chromatin-modifying enzymes controls colon cancer progression through Wnt signaling. *Hum Mol Genet* 2014;15;23(8):2120-31.
90. Barneda-Zahonero B, Parra M. Histone deacetylases and cancer. *Mol Oncol* 2012;6(6):579-89.

91. Rönsch K, Jäger M, Schöpflin A, Danciu M, Lassmann S, Hecht A. Class I and III HDACs and loss of active chromatin features contribute to epigenetic silencing of CDX1 and EPHB tumor suppressor genes in colorectal cancer. *Epigenetics* 2011;6(5):610-22.
92. Wang Y, Jia S. Degrees make all the difference: the multifunctionality of histone H4 lysine 20 methylation. *Epigenetics* 2009;4(5):273-6.
93. Jie D, Zhongmin Z, Guoqing L, Sheng L, Yi Z, Jing W et al. Positive expression of LSD1 and negative expression of E-cadherin correlate with metastasis and poor prognosis of colon cancer. *Dig Dis Sci* 2013;58(6):1581-9.
94. Nakazawa T, Kondo T, Ma D, Niu D, Mochizuki K, Kawasaki T et al. Global histone modification of histone H3 in colorectal cancer and its precursor lesions. *Hum Pathol* 2012;43(6):834-42.
95. Urasaki Y, Heath L, Xu CW Coupling of glucose deprivation with impaired histone H2B monoubiquitination in tumors. *PLoS One* 2012;7(5):e36775.
96. Vaiopoulos AG, Kostakis ID, Athanasoula KCh, Papavassiliou AG .Targeting transcription factor corepressors in tumor cells. *Cell Mol Life Sci* 2012;69(11):1745-53.
97. Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 2013;154(3):490-503.
98. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;6;489(7414):57-74.
99. Costa FF. Non-coding RNAs: new players in eukaryotic biology. *Gene* 2005;12;357(2):83-94.
100. Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 2011;12(2):136-49.
101. Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. *Science* 2007;318(5858):1931-4
102. Slaby O, Svoboda M, Michalek J, Vyzula R. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. *Mol Cancer* 2009;8:102.
103. Zhai H, Fesler A, Schee K, Fodstad O, Flatmark K, Ju J. Clinical significance of long intergenic noncoding RNA-p21 in colorectal cancer. *Clin Colorectal Cancer* 2013;12(4):261-266.
104. Zeng J, Kirk BD, Gou Y, Wang Q, Ma J. Genome-wide polycomb target gene prediction in *Drosophila melanogaster*. *Nucleic Acids Res* 2012;40(13):5848-63.
105. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature* 1978;7;276(5688):565-70.

106. van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H, Berns A. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* 1991;31;65(5):737-52.
107. Haupt Y, Alexander WS, Barri G, Klinken SP, Adams JM. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell* 1991 May 31;65(5):753-63.
108. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 2006;441(7091):349-53.
109. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 2006;125(2):301-13.
110. Schuettengruber B, Cavalli G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* 2009;136(21):3531-42.
111. Klose RJ, Cooper S, Farcas AM, Blackledge NP, Brockdorff N. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins *PLoS Genet* 2013;9(8):e1003717.
112. Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, Jones RS. Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 2004;14(5):637-46
113. Juan AH, Derfoul A, Feng X, Ryall JG, Dell'Orso S, Pasut A et al. Polycomb EZH2 controls self-renewal and safeguards the transcriptional identity of skeletal muscle stem cells. *Genes Dev* 2011;25(8):789-94.
114. Wassef M, Michaud A, Margueron R. Association between EZH2 expression, silencing of tumor suppressors and disease outcome in solid tumors. *Cell Cycle* 2016;15(17):2256-62.
115. Wassef M, Margueron R. The Multiple Facets of PRC2 Alterations in Cancers. *J Mol Biol* 2017;429(13):1978-1993.
116. Cao R, Zhang Y The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* 2004;14(2):155-64.
117. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011;469(7330):343-9.
118. Holoch D, Margueron R. Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends Biochem Sci.* 2017;42(7):531-542.
119. Deb G, Thakur VS, Gupta S. Multifaceted role of EZH2 in breast and prostate tumorigenesis: epigenetics and beyond. *Epigenetics* 2013;8(5):464-76.

120. Deb G, Singh AK, Gupta S. EZH2: not EZHY (easy) to deal. *Mol Cancer Res* 2014;12(5):639-53.
121. Gall Trošelj K, Novak Kujundzic R, Ugarkovic D. Polycomb repressive complex's evolutionary conserved function: the role of EZH2 status and cellular background. *Clin Epigenetics* 2016;8:55.
122. Yan KS, Lin CY, Liao TW, Peng CM, Lee SC, Liu YJ, Chan WP, Chou RH. EZH2 in Cancer Progression and Potential Application in Cancer Therapy: A Friend or Foe? *Int J Mol Sci*. 2017 May 31;18(6).
123. He A, Shen X, Ma Q, Cao J, von Gise A, Zhou P et al. Prc2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev* 2012; 26:37–42.
124. Lee JM, Lee JS, Kim H, Kim K, Park H, Kim JY et al. Ezh2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. *Mol. Cell* 2012;48,:572–586.
125. Shi B, Liang J, Yang X, Wang Y, Zhao Y, Wu H et al. Integration of estrogen and wnt signaling circuits by the polycomb group protein EZH2 in breast cancer cells. *Mol Cell Biol* 2007;27:5105–5119.
126. Jung HY, Jun S, Lee M, Kim HC, Wang X, Ji H et al. Paf and EZH2 induce wnt/-catenin signaling hyperactivation. *Mol Cell* 2013; 52:193–205.
127. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT et al. Ezh2 oncogenic activity in castration-resistant prostate cancer cells is polycomb-independent. *Science* 2012;338:1465–1469.
128. Lee ST, Li Z, Wu Z., Aau M, Guan P, Karuturi RK et al. Context-specific regulation of NF- $\kappa$ B target gene expression by EZH2 in breast cancers. *Mol Cell* 2011;43:798–810.
129. Kim E, Kim M, Woo DH, Shin Y, Shin J, Chang, N et al. Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* 2013;23:839–852.
130. Tatton-Brown K, Murray A, Hanks S, Douglas J, Armstrong R, Banka S et al; Childhood Overgrowth Consortium, Rahman N. Weaver syndrome and EZH2 mutations: Clarifying the clinical phenotype. *Am J Med Genet A*. 2013;161A(12):2972-80.
131. Tam EK, Nguyen TM, Lim CZ, Lee PL, Li Z, Jiang X et al. 3-Deazaneplanocin A and neplanocin A analogues and their effects on apoptotic cell death. *ChemMedChem* 2015;10(1):173-82.
132. Bachmann IM, Halvorsen OJ, Collett K, Stefansson IM, Straume O, Haukaas SA et al. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in

cutaneous melanoma and cancers of the endometrium, prostate, and breast. *J Clin Oncol* 2006;24:268–73.

133. Wang X, Hu B, Shen H, Zhou H, Xue X, Chen Y et al. et al. Clinical and prognostic relevance of EZH2 in breast cancer: A meta-analysis. *Biomed Pharmacother.* 2015;75: 218–25.

134. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002;419:624–629.

135. Crea F, Hurt EM. & Farrar, W. L. Clinical significance of Polycomb gene expression in brain tumors. *Mol Cancer* 2010;9:265.

136. Liu YL, Gao X, Jiang Y, Zhang G, Sun ZC, Cui BB et al. Expression and clinicopathological significance of EED, SUZ12 and EZH2 mRNA in colorectal cancer. *J Cancer Res. Clin. Oncol.* 2015; 141:661–9.

137. Mimori K, Ogawa K, Okamoto M, Sudo T, Inoue H, Mori M. Clinical significance of enhancer of zeste homolog 2 expression in colorectal cancer cases. *Eur J Surg Oncol* 2005;31: 376–380.

138. Fornaro L, Faviana P, De Gregorio V, Vivaldi C, Paolicchi E, Masi G et al. Molecular and pathological characterization of the EZH2 rs3757441 single nucleotide polymorphism in colorectal cancer. *BMC Cancer* 2015;15:874.

139. Fluge Ø, Gravdal K, Carlsen E, Vonen B, Kjellevoid K, Refsum S et al. Expression of EZH2 and Ki-67 in colorectal cancer and associations with treatment response and prognosis. *Br J Cancer* 2009;101:1282–9.

140. Benard A, Goossens-Beumer IJ, van Hoesel AQ, Horati H, Putter H, Zeestraten EC et al. Prognostic value of polycomb proteins EZH2, BMI1 and SUZ12 and histone modification H3K27me3 in colorectal cancer. *PLoS One* 2014;9(9):e108265.

141. Wang CG, Ye YJ, Yuan J, Liu FF, Zhang H, Wang S. EZH2 and STAT6 expression profiles are correlated with colorectal cancer stage and prognosis. *World J Gastroenterol* 2010;6:2421–2427.

142. Ibáñez-Sanz G, Díez-Villanueva A, Alonso MH, Rodríguez-Moranta F, Pérez-Gómez B, Bustamante M et al. Corrigendum: Risk Model for Colorectal Cancer in Spanish Population Using Environmental and Genetic Factors: Results from the MCC-Spain study. *Sci Rep* 2017;7:46807.

143. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): Explanation and Elaboration. *PLoS Med.* 2012;9: e1001216.

144. Steels E, Paesmans M, Berghmans T, Branle F, Lemaitre F, Mascaux C et al. Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis. *Eur Respir J* 2011;18:705–19.
145. Chen S, Huang L, Sun K, Wu D, Li M, Li M et al. Enhancer of zeste homolog 2 as an independent prognostic marker for cancer: a meta-analysis. *PLoS One* 2015;10:e0125480.
146. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16.
147. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
148. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.
149. Begg, C. B. & Mazumdar, M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:088–101.
150. CIE-O, 3ª edición, Clasificación Internacional de Enfermedades para Oncología
151. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E et al. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res.*2005;33(6):e56.
152. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S et al. The Ensembl gene annotation system Database 2016, baw093.
153. McLaren W, Gil L, Hunt S, Riat H, Ritchie G, Thormann A et al. The Ensembl Variant Effect Predictor *Genome Biology* 2016;7:122.
154. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-11.
155. The 1000 Genomes Project Consortium. A global reference for human genetic variation, *Nature* 2015;526:68-74.
156. International HapMap Consortium. The International HapMap Project *Nature.* 2003;426(6968):789-96.
157. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. dbSNP accession: {ss1 or ss1 – ss100}, (dbSNP Build ID: {build ID}). Disponible en: <http://www.ncbi.nlm.nih.gov/SNP/> (Acceso: 10/07/2017).
158. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* 2001;25(4):402-8.

159. Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4(7):844-7.

160. Solé X, Guinó E, Valls J, Iñesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics* 2006;22(15):928-1929.

161. Jinushi T, Shibayama Y, Kinoshita I, Oizumi S, Jinushi M, Aota T et al. Low expression levels of microRNA-124-5p correlated with poor prognosis in colorectal cancer via targeting of SMC4. *Cancer Med* 2014;3:1544–52.

162. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 2011;71:6320–6.

163. Ishikawa S, Hayashi H, Kinoshita K, Abe M, Kuroki H, Tokunaga R et al. Statins inhibit tumor progression via an enhancer of zeste homolog 2-mediated epigenetic alteration in colorectal cancer. *Int J Cancer* 2014;135:2528–36.

164. Tamagawa H, Oshima T, Numata M, Yamamoto N, Shiozawa M, Morinaga S et al. Global histone modification of H3K27 correlates with the outcomes in patients with metachronous liver metastasis of colorectal cancer. *Eur J Surg Oncol* 2013;39:655–61.

165. Kodach LL, Jacobs RJ, Heijmans J, van Noesel CJ, Langers AM, Verspaget HW et al. The role of EZH2 and DNA methylation in the silencing of the tumour suppressor RUNX3 in colorectal cancer. *Carcinogenesis* 2010;31(9):1567-75.

166. Takawa M, Masuda K, Kunizaki M, Daigo Y, Takagi K, Iwai Y et al. Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. *Cancer Sci* 2011;102(7):1298-305.

167. Lin YW, Ren LL, Xiong H, Du W, Yu YN, Sun TT et al. Role of STAT3 and vitamin D receptor in EZH2-mediated invasion of human colorectal cancer. *J Pathol* 2013;230:277–290.

168. Kurihara H, Maruyama R, Ishiguro K, Kanno S, Yamamoto I, Ishigami K et al. The relationship between EZH2 expression and microRNA-31 in colorectal cancer and the role in evolution of the serrated pathway. *Oncotarget* 2016;7:12704–17.

169. Wang W, Wang F, Zong G, Liu R, Zhang Y, Luan Y et al. Prognostic significance of EZH2 expression in patients with digestive cancers: a meta-analysis. *Int J Clin Exp Med* 2015;8:16043–9.

170. Jiang T, Wang Y, Zhou F, Gao G, Ren S, Zhou C. Prognostic value of high EZH2 expression in patients with different types of cancer: a systematic review with meta-analysis. *Oncotarget* 2016;7:4584–97.

171. Wang X, Zhao H, Lv L, Bao L, Wang X, Han S. Prognostic Significance of EZH2 Expression in Non-Small Cell Lung Cancer: A Meta-analysis. *Sci Rep* 2016;6:19239.
172. van Leenders GJ, Dukers D, Hessels D, van den Kieboom SW, Hulsbergen CA, Witjes JA et al. Polycomb-Group Oncogenes EZH2, BMI1, and RING1 Are Overexpressed in Prostate Cancer With Adverse Pathologic and Clinical Features. *Eur Urol* 2007;52:455–463.
173. Klee CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci* 2003;100:11606–11.
174. Sato T, Kaneda A, Tsuji S, Isagawa T, Yamamoto S, Fujita T et al. PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. *Sci Rep* 2013;3:1911.
175. Sashida G, Iwama A. Multifaceted role of the polycomb-group gene EZH2 in hematological malignancies. *Int J Hematol* 2017;105:23-30.
176. Lu H, Li G, Zhou C, Jin W, Qian X, Wang Z et al. Regulation and role of post-translational modifications of enhancer of zeste homologue 2 in cancer development. *Am J Cancer Res* 2016;6:2737-2754.
177. Sauvageau M, Sauvageau G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell*. 2010;7:299-313.
178. Wen Y, Cai J, Hou Y, Huang Z, Wang Z. Role of EZH2 in cancer stem cells: from biological insight to a therapeutic target. *Oncotarget* 2017;8(23):37974-37990.
179. Crea F, Fornaro L, Paolicchi E, Masi G, Frumento P, Loupakis F et al. An EZH2 polymorphism is associated with clinical outcome in metastatic colorectal cancer patients. *Ann Oncol* 2012;23:1207–1213.
180. Zhou Y, Du WD, Wu Q, Liu Y, Chen G, Ruan J, Xu S et al. EZH2 genetic variants affect risk of gastric cancer in the Chinese Han population. *Mol Carcinog* 2014;53:589–9.
181. Wei Y, Xia W, Zhang Z, Liu J, Wang H, Adsay NV et al. Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. *Mol Carcinog* 2008;47:701–6.
182. Bae WK, Yoo KH, Lee JS, Kim Y, Chung JJ, Park MH et al. The methyltransferase EZH2 is not required for mammary cancer development, although high EZH2 and low H3K27me3 correlate with poor prognosis of ER-positive breast cancers. *Mol Carcinog* 2015;54:1172–80.
183. Benard A, van de Velde CJ, Lessard L, Putter H, Takeshima L, Kuppen PJ et al. Epigenetic status of LINE-1 predicts clinical outcome in early-stage rectal cancer. *Br J Cancer* 2013;109: 3073–83.

184. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet.* 2007;8:286–98.
185. Palma Pérez S, Delgado Rodríguez M. Practical considerations on detection of publication bias. *Gac Sanit* 2006;20;3:10–6.
186. Simon J, Kingston RE. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 2009;10:697–708.
187. Vilorio-Marqués L, Molina AJ, Díez-Tascón C, Álvarez-Cuenllas B, Álvarez-Cañas C, Hernando-Martín M et al. Características clínicas, anatomopatológicas y moleculares en casos de cáncer colorrectal según localización tumoral y grado de diferenciación. *Rev Colomb Cancerol* 2015;19:193-203
188. Su KJ, Lin CW, Chen MK, Yang SF, Yu YL. Effects of EZH2 promoter polymorphisms and methylation status on oral squamous cell carcinoma susceptibility and pathology. *Am J Cancer Res* 2015;5(11):3475-84.
189. Gao SB, Sun SL, Zheng QL, Zhang L, Zhu Y, Jin GH et al. Genetic alteration and misexpression of Polycomb group genes in hepatocellular carcinoma. *Am J Cancer Res* 2015;5(10):2969-79.
190. Yu YL, Su KJ, Hsieh MJ, Wang SS, Wang PH, Weng WC et al. Impact of EZH2 polymorphisms on urothelial cell carcinoma susceptibility and clinicopathologic features. *PLoS One* 2014;9(4):e93635.
191. Yu YL, Su KJ, Hsieh YH, Lee HL, Chen TY, Hsiao PC et al. Effects of EZH2 polymorphisms on susceptibility to and pathological development of hepatocellular carcinoma. *PLoS One* 2013;8(9):e74870.
192. Chen JF, Luo X, Xiang LS, Li HT, Zha L, Li N et al. EZH2 promotes colorectal cancer stem-like cell expansion by activating p21cip1-Wnt/ $\beta$ -catenin signaling. *Oncotarget* 2016;7(27):41540-41558.
193. Yamamoto I, Nosho K, Kanno S, Igarashi H, Kurihara H, Ishigami K et al. EZH2 expression is a prognostic biomarker in patients with colorectal cancer treated with anti-EGFR therapeutics. *Oncotarget* 2017;8(11):17810-17818.
194. Lu H, Li G, Zhou C, Jin W, Qian X, Wang Z et al. Regulation and role of post-translational modifications of enhancer of zeste homologue 2 in cancer development. *Am J Cancer Res* 2016;6(12):2737-2754.
195. Shastri BS. Genetic susceptibility to primary angle closure glaucoma (PACG) *Discov Med* 2013;15(80):17-22

196. Mori K, Toiyama Y, Otake K, Fujikawa H, Saigusa S, Hiro J et al. Proteomics analysis of differential protein expression identifies heat shock protein 47 as a predictive marker for lymph node metastasis in patients with colorectal cancer. *Int J Cancer*. 2017;140(6):1425-1435.
197. Liu S, Zheng R, Zhang M, Zhang S, Sun X, Chen W. Incidence and mortality of colorectal cancer in China, 2011. *Chinese J Cancer Res* 2015;27(1):22-28.
198. Alvarez-Chaver P, Otero-Estévez O, Páez de la Cadena M, Rodríguez-Berrocal FJ, Martínez-Zorzano VS. Proteomics for discovery of candidate colorectal cancer biomarkers. *World J Gastroenterol*. 2014;20(14):3804-24
199. Yamaguchi H, Hung MC. Regulation and Role of EZH2 in Cancer. *Cancer Res Treat* 2014;46(3):209-22.
200. Sanz-Pamplona R, Berenguer A, Cordero D, Molleví DG, Crous-Bou M, Sole X et al. Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. *Mol Cancer* 2014;13:46.



***ANEXOS***



## ANEXO 1

### Protocolo de procesamiento de biopsias en cáncer colorrectal.

El diagnóstico anatomopatológico de las muestras de CCR de los casos estudiados, se realizó en el Servicio de Anatomía Patológica del CAULE, de acuerdo con un protocolo general estandarizado que resumimos a continuación. Todos los especímenes obtenidos tras cualquier procedimiento quirúrgico, endoscópico se remiten al Servicio de Anatomía Patológica, acompañados de una hoja de petición de estudio en la que consta la información demográfica y clínica siguiente: identificación del paciente (nombre y apellidos, edad, fecha de nacimiento, sexo y número de historia clínica), fecha del procedimiento, antecedentes relevantes, diagnóstico clínico, procedimiento (diagnóstico, curativo o paliativo) y técnica de obtención de la muestra. En función del tipo de espécimen recibido, se aplica el protocolo que corresponda:

Si el espécimen procede de **biopsia endoscópica**, la muestra irá acompañada de una petición de informe anatomopatológico en el que se harán constar los siguientes datos, tamaño de la lesión, localización anatómica, tipo (sésil o pediculado), e incidencias. En el informe anatomopatológico se hará constar el tipo histológico, el grado de displasia (adenoma) y el grado de diferenciación.

Si la pieza o el espécimen procede de **polipectomía endoscópica**, irá acompañada de una petición de informe anatomopatológico en el que se harán constar los siguientes datos, tamaño del pólipo, localización anatómica, tipo sésil o pediculado, resección completa o incompleta e incidencias.

Los datos descritos en el informe anatomopatológico serán los siguientes en caso de tratarse de un adenoma: el tipo histológico y el grado de displasia. En caso de tratarse de un adenocarcinoma se describe el tipo histológico, el nivel de invasión, el grado de diferenciación histológica y la presencia de invasión vascular o linfática además se valora y describe la extirpación total o no, del tumor, del borde quirúrgico.

Si la pieza procede de **resección quirúrgica** irá acompañada de un informe en el que constarán los siguientes datos; el tipo de intervención practicada y el objetivo del proceder quirúrgico (curativo o paliativo).

El anatomopatólogo procederá al estudio macroscópico (localización, tipo macroscópico, tamaño, descripción, distancia al borde de resección quirúrgica más próximo) y al estudio histopatológico, determinando los siguientes puntos

- Tipo de cirugía.
- Descripción macroscópica
- Características del tumor

Existen otros protocolos específicos para piezas (biopsias y resecciones quirúrgicas) en los que está indicada la quimioterapia y radioterapia preoperatoria y para especímenes con sospecha de poliposis adenomatosa familiar (PAF) y carcinoma colorrectal hereditario no polipósico (CCHNP o síndrome de Lynch) así como protocolos específicos para metástasis hepáticas de adenocarcinoma colorrectal o especímenes de pacientes en seguimiento tras el cáncer.

La clasificación anatomopatológica, emplea los datos recogidos tras el examen de muestras obtenidas del tumor y nos proporciona datos precisos que se utilizan para valorar el pronóstico y los resultados finales. Tanto para el diagnóstico de extensión clínico como anatomo-patológico se emplea la clasificación TNM. En cada caso, la valoración clínica y la evaluación del diagnóstico de extensión deben ser revisadas por un grupo multidisciplinario, para determinar la opción terapéutica más adecuada en cada caso.

## ANEXO 2

 <b>SERVICIO ANATOMÍA PATOLÓGICA</b>	<b>PROTOCOLO DE COLECTOMIA</b>	Código: P-08 Edición: 02 Fecha: 01-01-10
--	--------------------------------	--

## PROTOCOLO MACRO A LA BIOPSIA:

**TIPO DE ESPECIMEN:**

- I01 – Hemicolectomía derecha
- I02 – Colectomía transversa
- I03 – Hemicolectomía izquierda
- I04 – Sigmoidectomía
- I05 – Colon rectosigmoide/rectal (resección anterior baja)
- I06 – Colectomía abdominal total
- I07 – Resección abdominoperineal
- I08 – Exciisión local transanal
- I09 – Otros:
- I10 – No especificado

**LONGITUD DEL ESPECIMEN:**

- I11 – Colon: cm.
- I12 – Ileon: cm.
- Válvula ileocecal: cm.
- Apéndice cecal: cm.
- I15 – Mucosa y piel perianal: cm.

**LOCALIZACION DEL TUMOR:**

- I16 – Válvula ileocecal
- I17 – Ciego
- I18 – Colon ascendente (derecho)
- I19 – Angulo hepático
- I20 – Colon transverso
- I21 – Angulo esplénico
- I22 – Colon descendente (izquierdo)
- I23 – Sigma
- I24 – Rectosigma
- I25 – Recto
- I26 – No especificado
- I27 – No se puede determinar

**CONFIGURACION TUMORAL:**

- I28 – Polipoide
- I29 – Sesil
- I30 – Ulcerado
- I31 – No valorable

**TUMOR:**

- I32 – Tamaño tumoral: × cm.
- I33 – Espesor: cm.
- I34 – Distancia tumor-márgenes:
  - Proximal: cm.
  - Distal: cm.
- I37 – -Radial anterior: cm.
- I38 – -Radial posterior: cm.
- I39 – -Radial lateral derecha: cm.
- I40 – -Radial lateral izquierda: cm.

**CIRCUNFERENCIA AFECTADA:**

- I41 – Menos del 25%
- I42 – 25-50%
- I43 – 50-75%
- I44 – 75-100%
- I45 – Anular

**PERFORACION TUMORAL:**

- I46 – Presente
- I47 – No identificada
- I48 – No se puede determinar

**MESORRECTO:**

- I49 – No aplicable
- I50 – Completo
- I51 – Casi completo
- I52 – Incompleto
- I53 – No se puede determinar

**POLIPOS:**

- I54 – No
- I55 – Sí
  - Número:
  - Tipo de pólipo:
  - Distancia a margen distal: cm.

**RODETES:**

- I59 – No aplicable
- I60– Sí
- I61– No

**OTROS:**



## ANEXO 3

 <b>SERVICIO ANATOMÍA PATOLÓGICA</b>	<b>PROTOCOLO DE COLECTOMIA</b>	Código: P-06 Edición: 04 Fecha: 06-03-12
--	--------------------------------	--

## PROTOCOLO MICRO A LA BIOPSIA:

**TIPO HISTOLOGICO (OMS):**

- I63 – Adenocarcinoma
- I64 – Adenocarcinoma mucinoso (más del 50% mucinoso)
- I65 – Carcinoma de cél. en anillo de sello (más del 50% de cél en anillo de sello)
- I66 – Carcinoma de célula pequeña
- I67 – Carcinoma epidermoide
- I68 – Carcinoma adenoescamoso
- I69 – Carcinoma medular
- I70 – Carcinoma indiferenciado
- I71 – Otros:

**GRADO HISTOLOGICO:**

- I72 – No aplicable
- I73 – No se puede valorar
- I74 – Bajo grado (Bien-moderadamente diferenciado)
- I75 – Alto grado (Pobremente diferenciado-indiferenciado)

**CARACTERISTICAS HISTOLOGICAS SUGESTIVAS DE****INESTABILIDAD DE MICROSATELITES:**

- I76 – Respuesta linfocitaria intratumoral:
- I77 – -ninguna
- I78 – -ligera-moderada (0-2 por C.G.A.)
- I79 – -marcada (3 o más por C.G.A.)
- I80 – Respuesta linfocitaria peritumoral (Crohn-like):
- I81 – -ninguna
- I82 – -ligera-moderada
- I83 – -marcada
- I84 – Subtipo tumoral y diferenciación:
- I85 – -componente mucinoso (porcentaje:        )
- I86 – -componente medular
- I87 – -alto grado histológico (pobremente diferenciado)

**MARGENES QUIRURGICOS:**

- I88 – Libres de tumor
- I89 – El tumor afecta a los siguientes márgenes quirúrgicos:  
proximal, distal, radial

**GRADOS DE REGRESIÓN TUMORAL (modificado de Ryan y col):**

- I90 – Grado 0 (respuesta completa)
- I91 – Grado 1 (respuesta moderada)
- I92 – Grado 2 (respuesta mínima)
- I93 – Grado 3 (respuesta pobre o ninguna)

**INVASION LINFATICA (vasos pequeños):**

- I94 – No identificada
- I95 – Presente
- I96 – Indeterminada

**INVASION VENOSA (vasos grandes):**

- I97 – No identificada
- I98 – Presente:
- I99 – -Intramural
- I100 – -Extramural
- I101 – Indeterminada

**INVASION PERINEURAL:**

- I102 – No se puede valorar
- I103 – Ausente
- I104 – Presente

**DEPÓSITOS TUMORALES (extensión extramural discontinua):**

- I105 – No identificados
- I106 – Presentes
- I107 – Indeterminados

**TIPO DE PÓLIPO SOBRE EL QUE ASIENTA EL CARCINOMA:**

- I108 – No identificado
- I109 – Adenoma tubular
- I110 – Adenoma vellosa
- I111 – Adenoma tubulo-vellosa
- I112 – Adenoma serrado tradicional
- I113 – Adenoma serrado sesil
- I114 – Pólipo hamartomatoso
- I115 – Indeterminado

**ESTADIO PATOLÓGICO (pTNM):**

- I116 – Descriptor:
- I117 – m (tumores primarios múltiples)
- I118 – r (recidiva tumoral)
- I119 – y (post-tratamiento radio-quimioterápico)
- I120 – Tumor primario:
- I121 – pTX: El tumor primario no se puede valorar
- I122 – pT0: No hay evidencia de tumor primario
- I123 – pTis: El tumor no es invasivo (intraepitelial)
- I124 – pTis: El tumor invade la lámina propia
- I125 – pT1: El tumor invade la submucosa
- I126 – pT2: El tumor invade la muscular propia
- I127 – pT3: El tumor invade la subserosa o tejidos pericólicos o perirectales no peritonealizados
- I128 – pT4a: El tumor penetra el peritoneo visceral
- I129 – pT4b: El tumor directamente invade o está adherido a otros órganos o estructuras
- I130 – Ganglios linfáticos:
- I131 – Número examinado:
- I132 – Número afectado:
- I133 – pNX: No se puede valorar
- I134 – pN0: No metástasis en ganglios regionales
- I135 – pN1a: Metástasis en 1 ganglio linfático
- I136 – pN1b: Metástasis en 2 o 3 ganglios
- I137 – pN1c: Depósitos tumorales en el tejido adiposo sin metástasis en ganglios linfáticos regionales.
- I138 – pN2a: Metástasis en 4 - 6 ganglios linfáticos
- I139 – pN2b: Metástasis en 7 o más ganglios linfáticos
- I140 – Metástasis a distancia:
- I141 – No aplicable
- I142 – pM1a: Metástasis a un solo órgano o lugar. Especificar:
- I143 – pM1b: Metástasis a más de un órgano/lugar o al peritoneo. Especificar:

 <p>SERVICIO ANATOMÍA PATOLÓGICA</p>	<p><b>PROTOCOLO DE COLECTOMIA</b></p>	<p>Código: P-06 Edición: 04 Fecha: 06-03-12</p>
---	---------------------------------------	---

**INTESTINO NO TUMORAL:**

- I144 – No muestra anomalías
- I145 – Adenoma
- I146 – Pólipo hiperplásico
- I147 – Divertículos
- I148 – Colitis ulcerosa
- I149 – Enfermedad de Crohn
- I150 – Otros:

**INMUNOHISTOQUÍMICA PARA PROTEÍNAS**

**REPARADORAS DE DNA:**

- I151– MLH1:
- I152– -Presencia de expresión
- I153– -Ausencia de expresión
- I154– MSH2:
- I155– -Presencia de expresión
- I156– -Ausencia de expresión
- I157– MSH6:
- I158– -Presencia de expresión
- I159– -Ausencia de expresión
- I160– PMS2:
- I161– -Presencia de expresión
- I162– -Ausencia de expresión
- I163– No valorable
- I164– No realizado

**ANÁLISIS MUTACIONAL DE KRAS:**

- I165– Solicitado. Su resultado se enviará en un informe específico.
- I166– No solicitado.

## ANEXO 4

**Table 1.** The REMARK checklist [1–7].

<b>INTRODUCTION</b>	
1	State the marker examined, the study objectives, and any pre-specified hypotheses.
<b>MATERIALS AND METHODS</b>	
<i>Patients</i>	
2	Describe the characteristics (for example, disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.
3	Describe treatments received and how chosen (for example, randomized or rule-based).
<i>Specimen characteristics</i>	
4	Describe type of biological material used (including control samples) and methods of preservation and storage.
<i>Assay methods</i>	
5	Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.
<i>Study design</i>	
6	State the method of case selection, including whether prospective or retrospective and whether stratification or matching (for example, by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
7	Precisely define all clinical endpoints examined.
8	List all candidate variables initially examined or considered for inclusion in models.
9	Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.
<i>Statistical analysis methods</i>	
10	Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
11	Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.
<b>RESULTS</b>	
<i>Data</i>	
12	Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the number of patients and the number of events.
13	Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.
<i>Analysis and presentation</i>	
14	Show the relation of the marker to standard prognostic variables.
15	Present univariable analyses showing the relation between the marker and outcome, with the estimated effect (for example, hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.
16	For key multivariable analyses, report estimated effects (for example, hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.
17	Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.
18	If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.
<b>DISCUSSION</b>	
19	Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.
20	Discuss implications for future research and clinical value.

Note: we have changed 'univariate' to 'univariable' in item 15 for consistency with 'multivariable'.  
doi:10.1371/journal.pmed.1001216.t001



## ANEXO 5

714

E. STEELS ET AL.

defined. Sometimes, the manipulations might be inadequate: for example, the use of the antibody 1801 on a paraffin-embedded tissue, because of its specific immunoreactivity [104], or the absence of prior reaction of epitope unmasking on fixed tissue [105]. Finally, immunohistochemistry can produce false-negative results because some mutations are associated with the absence of detectable protein [106], or false-positive results because p53 can be quantitatively increased or stabilized in circumstances other than gene mutations [107–109]. Thus, there is no absolute correlation between immunohistochemistry and molecular biology. To try to exclude the technique bias, a subgroups analyses was performed according to the most frequently used methods: immunohistochemistry with antibody 1801, immunohistochemistry with antibody DO-7 and molecular biology (figs. 1–3). In any situation, the results were consistent, with poorer survival in cases where there was an abnormal p53 status, making the techniques improbable as a source of bias.

Some biases can be due to the methodology used to perform the meta-analysis. As already highlighted, the negative studies reported less detailed results, making them unlikely to be evaluated. Of 18 studies excluded for the meta-analysis due to a lack of reported data, 16 were negative, which represents a significantly higher proportion than in the trials evaluable for the meta-analysis. There is thus, a potential bias in favour of positive trials. Another potential source of biases is related to the method of extrapolation of the HR. If they were not reported by the authors, they were calculated from the data available in the article and, if this was not possible, they were extrapolated from the survival curves, hence making assumptions on the censoring process. Moreover, the choice of time intervals for reading mortality rates on the curves is not the object of a consensus. Finally, it should again be emphasized that a global meta-analysis did not appear meaningful because of the heterogeneity of the patient populations. This potential population bias makes it hazardous to generalize the results of the meta-analysis of subgroups of patients that were not included in the data aggregation, such as those with SCLC.

The present results are based on an aggregation of data obtained by univariate survival analysis in a majority of retrospective trials. It should be stressed that an HR of 1.5–2 does not mean a large effect of p53 on survival and, therefore, p53 does not constitute a useful factor for prediction of shorter survival at the level of individual patients. The results need to be confirmed by an adequately designed prospective study and the exact value of an abnormal p53 status needs to be determined by an appropriate multivariate analysis, taking into account the classical well-defined prognostic factors for lung cancer. A meta-analysis based on the individual data of the patients included into the studies [110] would help to define, by multivariate methods, the prognostic role of p53, but it will require the collection of a huge amount of retrospective data, with the potential problem of dealing with a lot of missing data. Moreover, it will

not have the definitive value of a well-designed prospective study [111].

To conclude, a systematic review of the lung cancer literature suggested a poor prognostic value for survival of an abnormal p53 status in patients with nonsmall cell lung cancer, whatever the biological test used. Data are insufficient to determine its role in small cell lung cancer. This observation requires further investigation with an adequate statistical methodology to determine its exact place among the other known survival prognostic factors for lung cancer.

#### **Appendix: European Lung Cancer Working Party quality scale for biological prognostic factors for lung cancer**

Except when specified, the attributed value per item is 2 points if it is clearly defined in the article, 1 point if its description is incomplete or unclear and 0 point if it is not defined or is inadequate.

##### *Scientific design*

1. Study objective definition.
2. Study design: prospective (2 points); retrospective or retrolective (1 point); not defined (0 point).
3. Outcome definition.
4. Statistical considerations: fully reported with a preliminary assessment of the patient/sample number to be included and/or analysed (2 points); patient/sample number to be included and/or analysed justified by the number of studied variables (minimum 10 patients per variable) (1 point); not defined (0 point).
5. Statistical methods and tests description.

##### *Laboratory methodology*

1. Blinding in the biological assays performance: double-blind (2 points); simple-blind (1 point); unblinded or not defined (0 point).
2. Tested factor description: DNA (types of exons analysed), messenger RNA (complete or partial with description of the primers used), protein (nuclear, cytoplasmic or extracted from cellular components), antibodies (type of tissue or liquid sampled).
3. Tissue sample conservation: either fresh tissue or conservation requiring freezing at  $\leq -80^{\circ}\text{C}$  in presence of an anti-RNAase for RNA or freezing at  $\leq -20^{\circ}\text{C}$  for DNA, protein and serum, or fixation in formol, alcohol or paraffin.
4. Description of the revelation test procedure of the biological factor: PCR with mention of primers, polymerase type, general reaction conditions (concentration of the various reagents, cycle number, duration and temperature of the various steps); DNA-sequencing with the

method used, the electrophoresis characteristics (gel composition, duration, temperature and generator voltage) and revelation procedure; reverse transcription (RT) with the transcriptase-reverse type and general conditions of incubation (reagent concentration, temperature and duration); SSCP with the gel composition (acrylamide percentage, glycerol content), other electrophoresis characteristics (duration, temperature and generator voltage), coloration method (the SSCP must be followed by the abnormally migrated fragments sequencing if there is no negative internal control); DGGE with gel composition and gel composition gradient, other electrophoresis characteristics (duration, temperature and generator voltage) and coloration method (the DGGE must be followed by the abnormally migrated fragments sequencing if there is no negative internal control); restricted fragment length polymorphism (RFLP) with the restriction enzyme type, temperature and incubation time, electrophoresis characteristics (gel composition, duration, temperature and generator voltage), coloration method (the RFLP must be followed by the abnormally migrated fragments sequencing if there is no negative internal control); IHC with the first antibody type and clone identification, second antibody type, reaction characteristics (antibodies concentration, duration and temperature of incubation), colouration method (peroxydase, alkaline phosphatase or chromagenic method), epitope unmasking method in case of fixed tissue, endogenous peroxydase activity inhibition method if the colouration method requires peroxydase; ELISA with the type of antibodies/antigens used, general reaction conditions (reagents concentration, temperature, duration), unspecific sites blockage (type of serum used), colouration and reading methods (fluorometry, spectrophotometry, *etc.*); Western blot with types of antibodies/antigens used, gel composition, other electrophoresis characteristics (duration, temperature, generator voltage), colouration method; immunoblot with the type of antibodies/antigens used, the type of membrane, other electrophoresis characteristics (duration, temperature, generator voltage), colouration method.

5. Description of the negative and positive control procedures.
6. Test reproducibility control: between investigator's centres if the study is multicentric, or inside the centre if it is unicentric.
7. Definition of the level of positivity of the test (only evaluated in case of IHC): validated (2 points); arbitrary (1 point); not described (0 point).

*Generalizability*

1. Patient selection criteria, including histological type, disease stage and treatment.

2. Patients' characteristics, including histology type, disease stage and treatment.
3. Initial workup.
4. Treatment description.
5. Source of samples.
6. Number of unassessable samples with exclusion causes.

*Results analysis*

1. Follow-up description, including the number of events.
2. Survival analysis according to the biological marker.
3. Univariate analysis of the prognostic factors for survival: report of the relative risk with the confidence interval (2 points); results without evaluation of the relative risk and its confidence interval (1 point); not reported or inadequate (0 point).
4. Multivariate analysis of the prognostic factors for survival: report of the relative risk with the confidence interval (2 points); results without evaluation of the relative risk and its confidence interval (1 point); not reported or inadequate (0 point).

*Acknowledgements.* The authors would like to thank D. Larsimont, P. Martiat, P. Vermeylen, B. Martin, D. Ganeberg and L. Lespagnard for their assistance in the design of the methodological scale for biological prognostic factors.

**References**

1. Paesmans M, Sculier J-P. Facteurs pronostiques des cancers bronchopulmonaires. *In: Milleron B, Depierre A eds. Cancers Bronchopulmonaires.* France, Arnette, 1998; pp. 239-247.
2. Paesmans M, Sculier JP, Lecomte J, *et al.* for the European Lung Cancer Working Party. Prognostic factors in patients with small cell lung cancer: analysis of a series of 763 patients included in four consecutive prospective trials and with a minimal 5-year follow-up duration. *Cancer* 2000; 89: 523-533.
3. Strauss GM. Prognostic markers in resectable non-small cell lung cancer. *Hematol Oncol Clin North Am* 1997; 11: 409-434.
4. Paesmans M, Sculier JP, Libert P, *et al.* Prognostic factors for survival in advanced non-small-cell lung cancer: univariate and multivariate analyses including recursive partitioning and amalgamation algorithms in 1,052 patients. The European Lung Cancer Working Party. *J Clin Oncol* 1995; 13: 1221-1230.
5. Buccheri G, Ferrigno D. Prognostic factors in lung cancer: tables and comments. *Eur Respir J* 1994; 7: 1350-1364.
6. Kanters SD, Lammers JW, Voest EE. Molecular and biological factors in the prognosis of non-small cell lung cancer. *Eur Respir J* 1995; 8: 1389-1397.

## ANEXO 6

Estudio MCC-Spain - ConsentFormA

### Estudio MCC-Spain

#### Hoja de información

##### Introducción

Solicitamos su participación en un estudio destinado a investigar el efecto que determinados factores ambientales, laborales, alimentarios y tratamientos médicos tienen sobre la salud. El estudio se está llevando a cabo en éste y otros centros sanitarios de España, y está coordinado desde el Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP).

Su participación en el estudio es voluntaria. Puede rechazar participar y puede retirarse en cualquier momento sin que esto afecte su atención médica posterior.

Por favor, lea atentamente este documento y haga tantas preguntas como crea necesarias.

##### Explicación de los procedimientos

Si acepta participar en el estudio, su aportación consistirá en responder a las preguntas de un cuestionario, donar una muestra de sangre/saliva, donar una muestra de pelo y una muestra de uña del pie y dar permiso para recoger información de su historial clínico. El cuestionario consta de dos partes. Una parte consiste en preguntas sobre sus hábitos de vida e historial clínico mediante una entrevista personal de unos 50 minutos. La otra parte es el cuestionario de dieta el cual le será entregado para que usted lo conteste y nos lo devuelva completado. Una enfermera extraerá una muestra de sangre (30 ml). La entrevistadora o la enfermera le proporcionarán un recipiente para recoger una muestra de saliva. En el caso de que sea intervenido quirúrgicamente, le pedimos permiso para obtener muestras de tejido. Mediremos la distancia anogenital.

Al final de la entrevista le haremos algunas medidas antropométricas, como la cintura y la cadera y también la longitud de los dedos índice y anular. También le pedimos una muestra de la uña del dedo gordo del pie y una muestra de pelo.

Las muestras recogidas serán utilizadas exclusivamente con fines científicos relacionados con los objetivos del estudio. Las muestras biológicas serán identificadas con un código para no ser identificables con el participante. Estas muestras se guardarán en congeladores o almacenes adecuados hasta su análisis. En caso de retirada del consentimiento del participante, serán destruidas. Con las muestras de sangre, saliva, uñas, pelo y tejido se realizarán análisis genéticos, bioquímicos y celulares que pueden estar asociados con la salud y la enfermedad. Le entregaremos una hoja informativa adicional para el consentimiento en los análisis genéticos.

Por último, deseáramos poder contactar en el futuro con usted o alguno de sus familiares para recabar información adicional que no se ha recogido en esta entrevista.

##### Beneficios potenciales

Usted no se beneficiará directamente de este estudio, pero su participación es muy importante ya que contribuirá a aumentar el conocimiento de los efectos de factores ambientales y genéticos sobre la salud. En caso de que los resultados de los análisis mostraran algún hallazgo importante para su salud, su médico sería informado para que se lo comunicara.

##### Riesgos

Es poco probable que, al hacer la extracción de sangre, le salga un hematoma ("morado").

##### Confidencialidad

La información obtenida en el estudio será confidencial, y almacenada en un fichero de datos automatizados, de acuerdo con lo que establece la Ley Orgánica 15/99 de Protección de datos de Carácter Personal (LOPD). Nadie, excepto los miembros del estudio y el Comité Ético de Investigación Clínica, tendrá acceso a su información. Los resultados obtenidos sólo se podrán publicar de forma anónima y nunca de forma individual.



**Hoja de información y consentimiento informado – estudio genético**

Utilizaremos una parte de la sangre, saliva y tejido para examinar su ADN y ARN (material genético). Nos gustaría realizar análisis exclusivamente con fines científicos de su ADN y ARN que están relacionados con el objetivo del estudio.

Las muestras biológicas serán identificadas con un código para no ser identificables con el participante. Todo el material que no sea utilizado inmediatamente será conservado durante un máximo de 50 años para realizar, posiblemente, posteriores determinaciones. Las muestras se guardarán en el Centre de Regulació Genòmica de Barcelona, programa Genes y Enfermedad. En caso de retirada del consentimiento del participante, serán destruidas.

En un futuro, usted puede solicitar a los investigadores que sus datos y muestras sean retirados del estudio.

En caso de que los resultados de los análisis mostraran algún hallazgo importante para su salud, su médico sería informado para que se lo comunicara.

**Confidencialidad**

La información obtenida en el estudio será confidencial, de acuerdo con lo que establece la Ley Orgánica 15/99 de Protección de datos de Carácter Personal (LOPD). Nadie, excepto los miembros del estudio y el Comité Ético de Investigación Clínica, tendrá acceso a su información. Los resultados obtenidos sólo se podrán publicar de forma anónima y nunca de forma individual.

- |  |                          |                          |
|--|--------------------------|--------------------------|
|  | <b>SÍ</b>                | <b>NO</b>                |
| 1. Entiendo que una muestra de mi ADN y ARN será almacenada y analizada como una parte de este estudio   | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Acepto que los investigadores almacenen mi ADN y ARN y lo analicen en el futuro. Entiendo que será utilizado específicamente para esta investigación y no será utilizado para otros propósitos. | <input type="checkbox"/> | <input type="checkbox"/> |

Nombre y apellidos del participante	Fecha	Firma
Nombre y apellidos del entrevistador/investigador	Fecha	Firma

Apreciamos sinceramente su cooperación en este proyecto de investigación.

Este protocolo ha sido revisado y aprobado por el Comité Ético del IMAS y de los centros participantes.

