

UNIVERSIDAD DE SALAMANCA

FACULTAD DE MEDICINA

DEPARTAMENTO DE MEDICINA



TESIS DOCTORAL

**CÁNCER COLORRECTAL EN EL ADULTO JOVEN,
UNA ENTIDAD DIFERENTE**

JÉSICA PÉREZ GARCÍA

2019



EL DR. D. ROGELIO GONZÁLEZ SARMIENTO, CATEDRÁTICO DEL DEPARTAMENTO DE MEDICINA DE LA UNIVERSIDAD DE SALAMANCA

CERTIFICA:

Que el trabajo titulado “Cáncer colorrectal en el adulto joven, una entidad diferente”, que presenta la licenciada en Ciencias Químicas Dña. Jérica Pérez García ha sido realizado bajo su dirección en la Unidad de Medicina Molecular del Departamento de Medicina y reúne, a su juicio, originalidad y contenidos suficientes para que sea presentado ante el tribunal correspondiente y optar al título de Doctora por la Universidad de Salamanca.

Y para que así conste y a los efectos oportunos, expido el presente certificado en Salamanca a 4 de junio de 2019.

Fdo.: Dr. Rogelio González Sarmiento

El presente trabajo ha sido financiado por el Proyecto FIS-FEDE PI16/01920

À mi niña, mi vida

Agradecimientos

*-En primer lugar, al **Prof. Dr. Rogelio González Sarmiento**, director de este trabajo de investigación. Son muchas las palabras de agradecimiento que podría dedicarle y, probablemente no serían suficientes. Gracias por la orientación y ayuda, por la paciencia y amistad, por haberme brindado la oportunidad de formar parte de este proyecto. Gracias por enseñarme tanto. Sin duda, mi mejor Maestro.*

*-Al **Dr. José Perea** y al **Dr. Juan Luis García** por su inestimable ayuda y su permanente disponibilidad. Gracias por contagiarme vuestra ilusión en este proyecto. Sin vosotros este trabajo no hubiera sido posible.*

*-Al **Servicio de Anatomía Patológica del Hospital Universitario 12 de Octubre (Madrid)**, al **Servicio de Anatomía Patológica** y al **Servicio de Digestivo del Hospital Universitario de Salamanca** y, de igual manera, al **Servicio de Anatomía Patológica del Hospital Nuestra Señora de Sonsoles (Complejo Asistencial de Ávila)** por proporcionarnos las muestras de tumores utilizadas en este trabajo y por su colaboración con el proyecto. También agradecer al **Servicio de Digestivo del Hospital Universitario de Salamanca***

*-A todos mis **compañeros del laboratorio** tanto de la Facultad de Medicina como del Centro de Investigación del Cáncer. Es imposible nombraros a todos los que habéis pasado por aquí y a todos los que estais. Son tantos recuerdos y tantas anécdotas... Una de las mejores cosas de trabajar en este laboratorio: el haber conocido gente maravillosa, algunos de vosotros más que compañeros de trabajo grandes amigos. De manera especial, agradecer a todas las personas que han formado parte del equipo "**colon**": Rosalía, contigo empezó todo, gracias por confiar en mi; Laurita y Yusep, gracias por los buenos ratos vividos juntos y, a todos los compañeros que nos han ayudado a que los estudios salgan a tiempo. Agradecer a todos los integrantes del grupo "**agro-querada**" porque, aunque de vez en cuando expulse a alguno un par de minutos, sois increíbles y espero que esta amistad no se pierda y que nuestras casas rurales prosigan mucho tiempo más. (La siguiente invitamos Javi y yo (J&J corporation)). Mon, no me puedo olvidar de ti, ¡¡gracias por tu apoyo, por tus animos y por darme a mi futuro yerno!!*

*-**Eva**, que decirte... ¡¡gracias por todo!! Gracias por tus animos, tu apoyo, tus consejos y, sobre todo, gracias por tu amistad. Tantas horas hemos pasado juntas, tantos momentos buenos y no tan buenos, pero siempre los hemos llevado de la mejor manera posible. ¡¡Gracias por ser la mejor compañera y tan buena amiga!!*

*-A **Nieves**, por todo el apoyo que me has ofrecido cuando lo he necesitado, porque haces mucho más fácil el trabajo difícil. A **Estrella**, por tu paciencia y tu permanente disponibilidad. Gracias a las dos por vuestro cariño y cercanía. También quiero agradecer de manera especial la inestimable ayuda de **Ana Belén** desde el primer momento en el que te incorporaste al laboratorio, gracias por ayudarme tanto.*

*-A todas las personas que he conocido durante todo este tiempo que, sin ser compañeros de laboratorio, también se han convertido en grandes amigos para mí. En especial a **Rosario Vidal**, gracias por escucharme, por ayudarme tanto, por estar cuando te he necesitado. ¡¡Gracias de corazón!!*

*-A mi **familia**, gracias a mis padres por todo lo que me habéis dado, por ser mi ejemplo a seguir. Gracias por vuestro sacrificio diario para conseguir que haya llegado hasta aquí. Os debo todo. **Mamá**, gracias por tu interminable paciencia conmigo, gracias por estar siempre ahí y, pese a que muchas veces me dijeras que parase ya, gracias por apoyarme hasta el final. Ójala sepa hacerlo con mi niña como tu lo has hecho conmigo. **Papá**, simplemente gracias por quererme tanto, gracias por ser la alegría de la casa. **Patri**, ¡¡uf!! ¿¿qué sería de mi vida sin ti?? No te digo nada porque de mí ya lo sabes todo, ¡¡te quiero hasta el infinito!! Gracias por hacerme con **David** el regalo más bonito que jamás imaginé, **Lola**.*

*-**David**, la mejor persona que he conocido, gracias por hacerme tan feliz. Gracias por apoyarme en todo y hacer mi vida tan bonita. Esta tesis también lleva tu nombre. ¡¡Te quiero!!*

*-**Carmen**, gracias por enseñarme a amar.*

Recuerda mirar las estrellas y no tus pies. Intenta dar sentido a lo que ves y pregúntate por lo que hace al universo existir. Sé curioso. Aunque la vida puede parecer difícil, siempre hay algo que puedes hacer y tener éxito. Lo importante es que no te rindas.

Stephen Hawking

ÍNDICES



Índice

ÍNDICES	II
Índice de figuras	VI
Índice de tablas	VIII
Tabla de abreviaturas	B
INTRODUCCIÓN	6
1. EL CÁNCER	27
2. EL CÁNCER DE COLON	28
2.1. CLASIFICACIÓN DEL CCR	30
2.2. EPIDEMIOLOGÍA DEL CCR	31
2.3. FACTORES DE RIESGO	33
2.3.1. FACTORES DE RIESGO RELACIONADOS CON LA FORMA DE VIDA	33
2.3.2. FACTORES DE RIESGO BIOLÓGICOS	34
2.3.3. FACTORES ASOCIADOS A OTRAS ENFERMEDADES	34
2.3.4. FACTORES DE RIESGO HEREDITARIO	35
3. SÍNDROMES HEREDITARIOS DE PREDISPOSICIÓN AL CCR	35
3.1. SÍNDROME DE LYNCH	35
3.2. POLIPOSIS	37
3.2.1. POLIPOSIS ADENOMATOSA FAMILIAR (PAF)	38
3.2.2. OTRAS VARIANTES DE LA POLIPOSIS	38
4. CARCINOGENÉISIS DEL CÁNCER COLORRECTAL	41
4.1. VÍA DE INESTABILIDAD CROMOSÓMICA (CIN)	42
4.2. VÍA DE INESTABILIDAD DE MICROSATÉLITES (MIS)	43
4.3. VÍA SERRADA O VÍA DEL FENOTIPO METILADOR DE ISLAS CpG (CIMP)	44
4.4. CLASIFICACIÓN MOLECULAR DEL CCR	46
5. CÁNCER COLORRECTAL EN JÓVEN VS. CÁNCER COLORRECTAL EN ADULTO	47
6. CÁNCER COLORRECTAL EN EL ADULTO JOVEN	48
6.1. CARACTERÍSTICAS MOLECULARES DEL CÁNCER COLORRECTAL EN EL ADULTO JÓVEN	48
6.2. PERFIL CROMOSÓMICO DEL EOCRC	50
7. DELECIÓN DE LA REGIÓN 16p13.12-p13.11	52
8. PAPEL DE <i>NOMO1</i> EN LA CARCINOGENÉISIS DEL CCR	55
8.1. ASPECTOS GENERALES SOBRE <i>NOMO1</i>	55
8.2. COMPLEJO PROTEÍCO NCLN/NOMO/TEMEM147	56
9. VÍA DE SEÑALIZACIÓN NODAL	58
9.1. ASPECTOS GENERALES DE LA VÍA NODAL	58
9.2. PAPEL DE LA VÍA NODAL EN CARCINOGENÉISIS	60
9.3. PAPEL DE LA VÍA NODAL EN EL CARCINOMA COLORRECTAL	61
HIPÓTESIS Y OBJETIVOS	63
PACIENTES Y MÉTODOS	67

1. PACIENTES-----	69
1.1. PROCEDENCIA Y CLASIFICACIÓN -----	69
1.2. RECOGIDA DE INFORMACIÓN-----	70
2. MANIPULACIÓN DE ÁCIDOS NUCLEICOS -----	72
2.1. EXTRACCIÓN DE DNA A PARTIR DE SANGRE PERIFÉRICA -----	72
2.2. EXTRACCIÓN DE DNA A PARTIR DE TEJIDO TUMORAL FRESCO -----	72
2.3. EXTRACCIÓN DE DNA A PARTIR DE TEJIDO TUMORAL PARAFINADO-----	72
2.4. EXTRACCIÓN DE DNA A PARTIR DE LÍNEAS CELULARES -----	73
2.5. EXTRACCIÓN DE RNA A PARTIR DE LÍNEAS CELULARES -----	73
3. CUANTIFICACIÓN DEL DNA EXTRAÍDO -----	74
4. AMPLIFICACIÓN DE FRAGMENTOS DE DNA MEDIANTE PCR-----	74
5. ELECTROFORESIS EN GEL DE AGAROSA-----	75
6. PURIFICACIÓN DE ÁCIDOS NUCLEICOS-----	76
6.1. PURIFICACIÓN DE DNA-----	76
6.2. PURIFICACIÓN DE RNA -----	76
7. SECUENCIACIÓN AUTOMÁTICA-----	76
8. HIBRIDACIÓN <i>IN SITU</i> POR FLUORESCENCIA (FISH)-----	77
9. PCR CUANTITATIVA-COMPARATIVA A TIEMPO REAL (q-PCR) -----	79
10. ESTUDIO DE LA LONGITUD TELOMÉRICA MEDIANTE qPCR-----	83
10.1. AMPLIFICACIÓN DE LOS FRAGMENTOS-----	83
10.2. ANÁLISIS ESTADÍSTICO -----	84
11. CULTIVOS CELULARES-----	85
12. TECNOLOGÍA CRISPR-CAS9 -----	86
12.1. DISEÑO Y MANIPULACIÓN DE LOS sgRNAs -----	87
12.2. CLONAJE DE LOS sgRNAs-----	89
12.3. TRANSFORMACIÓN DE BACTERIAS COMPETENTES <i>E.coli DH5</i> MEDIANTE CHOQUE TÉRMICO --	91
12.4. EXTRACCIÓN Y PURIFICACIÓN DE DNA PLASMÍDICO (MAXIPREP)-----	92
12.5. TRANSFECCIÓN DE LA LÍNEA CELULAR HT29 CON PX458-sgRNAs -----	94
12.5.1. CONTAJE CELULAR-----	94
12.5.2. TRANSFECCIÓN POR LIPOFECCIÓN-----	94
12.5.3. TRANSFECCIÓN POR ELECTROPORACIÓN -----	95
12.6. <i>CELL SORTING</i> MEDIANTE CITOMETRÍA DE FLUJO -----	96
13. ENSAYO DE VIABILIDAD CELULAR -----	97
14. ENSAYO DE MIGRACIÓN CELULAR-----	97
15. MICROARRAYS GENÓMICOS CON LA PLATAFORMA CYTOSCAN™-----	98
16. ANÁLISIS DE EXPRESIÓN DE PROTEÍNAS DE LA VÍA NODAL-----	100
16.1. EXTRACCIÓN DE PROTEÍNAS A PARTIR DE LÍNEAS CELULARES -----	100
16.2. WESTERN BLOT-----	100
17. ARRAYS DE EXPRESIÓN (WT PLUS DE AFFYMETRIX)-----	102
18. ANÁLISIS DE LOS GENES IMPLICADOS EN LA VIA NODAL MEDIANTE SECUENCIACIÓN MASIVA -----	103

18.1. DISEÑO DEL PANEL -----	103
18.2. MEDIDA DE LA CALIDAD DEL DNA-----	104
18.3. PREPARACIÓN DE LA LIBRERÍA -----	105
18.3.1. HIBRIDACIÓN -----	105
18.3.2. PURIFICACIÓN -----	106
18.3.3. EXTENSIÓN Y LIGACIÓN-----	106
18.3.4. AMPLIFICACIÓN DE LA LIBRERÍA-----	107
18.3.5. PURIFICACIÓN DE LA LIBRERÍA-----	108
18.4. SECUENCIACIÓN EN EL EQUIPO MiSeq (ILLUMINA)-----	108
18.5. ANÁLISIS DE LAS VARIANTES-----	108
18.5.1. PROGRAMAS PARA EL ESTUDIO <i>IN SÍLICO</i> DE MUTACIONES NO DESCRITAS-----	109
18.5.2. ANÁLISIS DE SITIOS REGULADORES DE <i>SPLICING</i> -----	111
19. BASES DE DATOS -----	111
RESULTADOS -----	113
1. ESTUDIO DE LA REGIÓN 16p13.12-p13.1-----	115
1.1. ESTUDIO DE LA REGIÓN 16p13.12-p13.11 EN PACIENTES CON CCR-----	115
1.2. HIBRIDACIÓN <i>IN SITU</i> POR FLUORESCENCIA DE LA REGIÓN 16p13.12-p13.11 -----	116
1.3. COMPARACIÓN DE LAS CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS, MOLECULARES Y FAMILIARES ENTRE DIFERENTES SUBGRUPOS-----	116
1.3.1. GRUPO GLOBAL DE PACIENTES CON PÉRDIDA DE LA REGIÓN 16p13.12-p13.11 vs. GRUPO GLOBAL DE PACIENTES SIN PÉRDIDA DE LA REGIÓN 16p13.12-p13.11 -----	116
1.3.2. PACIENTES JÓVENES CON PÉRDIDA DE LA REGIÓN 16p13.12-p13.11 vs. PACIENTES JÓVENES SIN PÉRDIDA DE LA REGIÓN 16p13.12-p13.11-----	119
1.3.3. ANÁLISIS DE LA LONGITUD TELOMÉRICA EN LOS TUMORES DE LOS DIFERENTES SUBGRUPOS DE PACIENTES-----	122
1.4. ESTUDIO DEL GEN <i>NOMO1</i> EN LA REGIÓN 16p13.12-p13.11 -----	124
1.4.1. EXISTENCIA DE UN ÚNICO GEN <i>NOMO</i> -----	124
1.4.2. ESTUDIO DEL GEN <i>NOMO</i> EN CCR -----	127
1.4.3. LA DELECIÓN DE <i>NOMO</i> ES SOMÁTICA-----	131
1.4.4. ESTADO DEL GEN <i>NOMO</i> EN PÓLIPOS ADENOMATOSOS -----	132
1.4.5. ESTADO DEL GEN <i>NOMO</i> EN LA REGIÓN PERITUMORAL -----	132
1.4.6. ESTUDIO DEL GEN <i>NOMO</i> EN OTROS TUMORES-----	133
2. ESTUDIO DEL PAPEL DE <i>NOMO</i> EN LA VÍA NODAL-----	134
2.1. GENERACIÓN DE UNA LÍNEA CELULAR <i>KNOCKOUT</i> PARA <i>NOMO</i> MEDIANTE LA TÉCNICA DE EDICIÓN GÉNICA CRISPR/CAS9-----	135
2.1.1. <i>SCREENING</i> DE COLONIAS POR SECUENCIACIÓN DEL PLÁSMIDO PX458-----	135
2.1.2. TRANSFECCIÓN DE LA LÍNEA CELULAR HT29-----	136
2.1.3. GENERACIÓN DE UNA LÍNEA CELULAR <i>KNOCKOUT</i> PARA <i>NOMO</i> -----	138
2.2. CARACTERIZACIÓN FENOTÍPICA DE LOS CLONES <i>KNOCKOUT</i> PARA <i>NOMO</i> OBTENIDOS -----	140
2.2.1. ESTUDIO DE VIABILIDAD CELULAR-----	141
2.2.2. FORMACIÓN DE ESFERAS -----	141

2.2.3. ESTUDIO DE LA CAPACIDAD MIGRATORIA CELULAR-----	142
2.3. ESTUDIO DE ALTERACIONES <i>OFF-TARGET</i> EN LA LÍNEA <i>KNOCKOUT</i> PARA <i>NOMO</i> MEDIANTE aCGH -----	144
2.4. ESTUDIO DE LA EXPRESIÓN DE LAS PROTEINAS DEL COMPLEJO NCLN/ <i>NOMO</i> /TEMEM147-----	146
2.5. ESTUDIO DE LA EXPRESIÓN DE LAS PROTEINAS INTEGRANTES DE LA VÍA NODAL-----	147
3. ESTUDIO DEL TRANSCRIPTOMA EN AUSENCIA DE <i>NOMO</i> -----	150
4. ESTUDIO GENÉTICO DE LOS GENES IMPLICADOS EN LA VÍA DE SEÑALIZACIÓN NODAL Y del complejo PROTEICO <i>nomo/ncl/temem147</i> MEDIANTE SECUENCIACIÓN MASIVA (NGS) -----	152
4.1. ASPECTOS GENERALES -----	152
4.2. ANÁLISIS DE LOS DATOS -----	154
4.1.1. VARIANTES NEUTRALES-----	155
4.1.2. VARIANTES DE SIGNIFICADO CLÍNICO DESCONOCIDO-----	158
4.1.3. MUTACIONES PATOGENÉTICAS-----	165
DISCUSIÓN -----	171
1. LA REGIÓN 16p13.12-13.11 ESTÁ DELEACIONADA EN CCR -----	173
2. EL GEN <i>NOMO</i> ESTÁ IMPLICADO EN EOCRC-----	176
3. NCLN Y <i>NOMO</i> SON ESENCIALES PARA MANTENER LA INTEGRIDAD DEL COMPLEJO -----	180
4. LA VÍA NODAL NO SE MODIFICA EN AUSENCIA DE <i>NOMO</i> -----	181
5. LOS GENES IMPLICADOS EN LA VÍA DE SEÑALIZACIÓN NODAL ESTÁN MUTADOS EN CCR -----	183
CONCLUSIONES -----	193
BIBLIOGRAFÍA -----	197
ANEXOS -----	217
Anexo 1. Datos de la media de los valores de C _T obtenidos para las curvas de amplificación tanto de <i>TEL1</i> (telómero) como de <i>36B4</i> (control endógeno) para cada tumor analizado. -----	219
Anexo 2. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes jóvenes (qPCR). -----	221
Anexo 3. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes jóvenes (qPCR). -----	223
Anexo 4. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes ancianos (qPCR).-----	227
Anexo 5. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes de edades comprendidas entre los 46 y los 69 años (grupo intermedio) (qPCR). -----	229
Anexo 6. Resultados normalizados de la cantidad de DNA (RQ) amplificada para sangre periférica en pacientes jóvenes con delección homocigota en el tumor (qPCR). -----	231
Anexo 7. Resultados normalizados de la cantidad de DNA (RQ) amplificada en pólipos adenomatosos en pacientes jóvenes con delección homocigota en el tumor (qPCR). -----	233
Anexo 8. Resultados normalizados de la cantidad de DNA (RQ) amplificada en muestra peritumoral en pacientes jóvenes con delección homocigota en el tumor (qPCR). -----	235
Anexo 9. Resultados normalizados de la cantidad de DNA (RQ) amplificada en muestras de otros tumores procedentes de pacientes jóvenes (qPCR). -----	237
Anexo 10. Valores de expresión de los genes integrantes de la vía NOD. -----	239
Anexo 11. Listado de artículos publicados. -----	241

Índice de figuras

Figura 1. Tumores más frecuentemente diagnosticados en el mundo en el año 2018	27
Figura 2. Anatomía del colon	28
Figura 3. Cáncer colorrectal localizado en el revestimiento interno del colon.....	29
Figura 4. Número estimado de casos incidentes de cáncer en España.....	33
Figura 5. Síndrome de Peutz-Jeghers: manchas pigmentarias en piel y mucosas	39
Figura 6. Esquema secuencia adenoma-carcinoma.....	43
Figura 7. Hipermetilación de la región promotora y silenciamiento global.....	45
Figura 8. Perfil molecular del CCR en el adulto joven	49
Figura 9. Resultados de la segmentación de la región 16p.13.12-p13.11 para seis de las muestras analizadas	52
Figura 10. Esquema de la vía de señalización Nodal	58
Figura 11. Representación esquemática de la vía Nodal y del posible papel del complejo Nicalina/Nomo	60
Figura 12. Programa de amplificación estándar utilizado.....	75
Figura 13. Secuencia de nucleótidos correspondiente a un fragmento del gen <i>NOMO</i>	77
Figura 14. Gráfica de la curva de fusión del producto de amplificación	80
Figura 15. Esquema representativo del programa de termociclador para la cuantificación relativa de fragmentos de DNA mediante PCR a tiempo real cuantitativa comparativa	82
Figura 16. Esquema representativo del programa del termociclador para la amplificación de telómeros mediante PCR cuantitativa	84
Figura 17. Gráficas resultado de la amplificación de los telómeros.....	84
Figura 18. Ilustración dónde se reflejan las tres fases del mecanismo de defensa de microorganismos llevado a cabo por CRISPR-Cas	86
Figura 19. Esquema de la región diana del gen <i>NOMO</i> donde aparecen marcados los tres sgRNAs seleccionados	88
Figura 20. Esquema parcial del vector PX458	89
Figura 21. Esquema de los pasos a seguir para la extracción de DNA empleando columnas.....	92
Figura 22. Esquema de trabajo de la plataforma de microarrays genómicos Cytoscan 750k.....	99
Figura 23. Esquema y reactivos de las diferentes etapas a seguir en la preparación de la librería	105
Figura 24. Montaje para la asignación de los index a las muestras	107
Figura 25. Perfil cromosómico del cromosoma 16 de dos tumores incluidos en el estudio.....	115
Figura 26. Confirmación mediante FISH de la delección de la región 16p13.12-p13.11	116
Figura 27. Diagrama de Kaplan-Meier de la supervivencia global y la supervivencia libre de enfermedad de acuerdo con el estado de la región 16p.	118
Figura 28. Representación gráfica de la longitud telomérica media de los tumores según la edad de diagnóstico y el estado de la región 16p13.12-p13.1.....	123
Figura 29. Alineamiento de la secuencia de nucleótidos amplificada para dos fragmentos diferentes de los genes <i>NOMO1</i> , <i>NOMO2</i> y <i>NOMO3</i>	125
Figura 30. Representación gráfica de las barras de normalización correspondientes a la cantidad de DNA amplificada para cada fragmento de diferentes genes incluidos en el estudio.....	126
Figura 31. q-PCR correspondiente al análisis de un fragmento de <i>NOMO</i> en los 5 tumores	128
Figura 32. Estado del gen <i>NOMO</i> en los tumores EOCRC y LOCRC dependiendo del estado de la región 16p13.12-p13.11.....	129
Figura 33. Diagrama correspondiente a las barras de normalización de la cantidad de DNA amplificado del fragmento de <i>NOMO</i> para 5 de las muestras de sangre periférica incluidas en el estudio.....	131

Figura 34. Diagrama correspondiente a las barras de normalización de la cantidad de DNA amplificada del fragmento de NOMO para tres de las muestras de DNA extraído de región peritumoral	132
Figura 35. Barras de normalización de la cantidad de DNA amplificada del fragmento de NOMO para las líneas celulares HT29 y HCT116	134
Figura 36. Esquema del plásmido pSpCas9(BB)-2A-GFP (PX458).....	135
Figura 37. Secuencias de nucleótidos de cada sgRNA insertado en el plásmido PX458.....	136
Figura 38. Imágenes en contraste de fases y fluorescencia de células HT29 transfectadas con PX458-sgRNA.....	137
Figura 39. Fragmento de la secuencia de nucleótidos correspondiente a parte del exón 3 y del intrón 3 del gen NOMO.....	139
Figura 40. Barras de normalización de la cantidad de DNA amplificada del fragmento de NOMO para los clones obtenidos mediante combinación de guías.....	140
Figura 41. Expresión de la proteína NOMO en las líneas celulares HT29 sin transfectar (HT29 NORMAL), en la HT29 transfectada con el plásmido vacío (PX458 VACÍO) y en la línea knockout para NOMO (KNOCKOUT NOMO).....	140
Figura 42. Efecto del del bloque del gen NOMO sobre la proliferación celular.....	141
Figura 43. Imagen del crecimiento de las células de los diferentes clones obtenidos.....	142
Figura 44. Fotos tomadas a las 0h, 24h y 48h de la migración celular para los dos clones <i>knockout</i> para NOMO (clon 1 y clon 2) y los dos clones control (HT29 normal y PX458) correspondientes a una de las réplicas.....	142
Figura 45. Análisis de los datos de migración.....	143
Figura 46. Representación esquemática de las alteraciones en el número de copias detectadas para los tres clones analizados en los diferentes cromosomas.....	144
Figura 47. Expresión de las proteínas NCLN y TEMEM147 en los tres clones.....	147
Figura 48. Expresión de las proteínas ALK4 y ActR11 en los tres clones.....	148
Figura 49. Expresión de la proteína CRIPTO en los tres clones.....	149
Figura 50. Expresión de las proteínas SMADs: SMAD2, SMAD2-P, SMAD3, SMAD3-P y SMAD4 en los tres clones.....	150
Figura 51. Vías de señalización alteradas considerando todos los genes desregulados ante el bloqueo de NOMO.....	152

Índice de tablas

Tabla 1. Clasificación de Astler y Coller.....	30
Tabla 2. Clasificación TNM para el CCR, 7ª Ed. AJCC.....	31
Tabla 3. Riesgo de tumores en síndrome de Lynch.....	36
Tabla 4. Síndromes polipósicos que predisponen al CCR.....	37
Tabla 5. Diferencias clínico-patológicas y moleculares entre el CCR diagnosticado en el adulto joven y el CCR diagnosticado en el adulto anciano.....	47
Tabla 6. Resumen de las características clínicas, anatomopatológicas y familiares más importantes de los pacientes incluidos en el estudio de aCGH.....	71
Tabla 7. Oligonucleótidos empleados para la amplificación de la región de interés del gen NOMO.....	75
Tabla 8. Oligonucleótidos utilizados para la amplificación de las regiones analizadas mediante PCR cuantitativa en los ensayos de pérdida de heterocigosidad (LOH).....	81
Tabla 9. Oligonucleótidos utilizados para la amplificación de los telómeros mediante PCR cuantitativa.....	83
Tabla 10. Líneas celulares utilizadas y resumen de sus mutaciones.....	85
Tabla 11. Secuencias dianas seleccionada y sus oligos correspondientes.....	89
Tabla 12. Reactivos necesarios en cada etapa del clonaje de los sgRNAs.....	90
Tabla 13. Condiciones de temperatura y tiempo para cada etapa de la clonación de las guías en el vector PX458.....	91
Tabla 14. Etapas que seguir para la extracción y purificación de DNA plasmídico (Maxiprep).....	93
Tabla 15. Anticuerpos usados en este trabajo en los experimentos de western blot.....	102
Tabla 16. Cantidad de DNA necesaria dependiendo de la calidad.....	104
Tabla 17. Condiciones adecuadas para llevar a cabo la hibridación de los oligos con las regiones específicas de interés.....	106
Tabla 18. Condiciones para llevar a cabo la extensión y ligación.....	107
Tabla 19. Condiciones para llevar a cabo la amplificación de librerías.....	107
Tabla 20. Características clínicas, patológicas y familiares de todos los casos de cáncer de colon (grupo global) sin tener en cuenta la edad.....	117
Tabla 21. Características clínicas, patológicas y familiares de los pacientes diagnosticados a una edad igual o inferior a los 45 años, con y sin la alteración en el brazo corto del cromosoma 16.....	120
Tabla 22. Frecuencia alélica de los nucleótidos que difieren en la secuencia nucleotídica de los tres genes (<i>NOMO1</i> , <i>NOMO2</i> y <i>NOMO3</i>).....	127
Tabla 23. Resumen del estado del gen NOMO en todos los tumores analizados en función de la edad de diagnóstico.....	130
Tabla 24. Estado de NOMO en tumores EO CRC en función de si los tumores son MSS o MIS.....	131
Tabla 25. Estado del gen NOMO en diferentes tipos de tumores.....	134
Tabla 26. Informe aportado por el Servicio de Separación Celular del Centro de Investigación del Cáncer de Salamanca.....	138
Tabla 27. Genes con diferencias estadísticamente significativas en su expresión al inactivar NOMO.....	151
Tabla 28. Genes incluidos en el panel diseñado para el estudio genético de los genes implicados en la vía Nodal.....	153
Tabla 29. Muestras secuenciadas de manera correcta en el estudio de mutaciones en genes implicados en la vía Nodal.....	154
Tabla 30. Listado de variantes consideradas como variantes neutrales identificadas en los diferentes genes que componen el panel de genes implicado en la vía Nodal.....	157
Tabla 31. Listado de variantes descritas como neutrales por los programas de predicción.....	160
Tabla 32. Predicciones realizadas mediante el programa ESE FINDER en las variantes probablemente neutrales.....	161
Tabla 33. Listado de variantes descritas como neutrales por unos programas de predicción utilizados y como patogénicas por otros programas.....	162

Tabla 34. Predicciones realizadas mediante el programa ESE FINDER en variantes de significado incierto.....	163
Tabla 35. Listado de variantes descritas como patogénicas por los dos programas de predicción utilizados.....	164
Tabla 36. Predicciones realizadas mediante el programa ESE FINDER en variantes probablemente patogénicas.....	165
Tabla 37. Listado de mutaciones patogénicas identificadas en los tumores analizados.....	167
Tabla 38. Número de tumores correspondientes a cada uno de los subgrupos establecidos en este trabajo con mutación patogénica en los diferentes genes implicados en la vía de señalización Nodal.....	169

TABLA DE ABREVIATURAS



TABLA DE ABREVIATURAS

- **A**: adenina
- **ACTRII**: *ARF2 Actin Related Protein 2 Homolog*
- **ACVRIB**: *activin receptor type 1B precursor*
- **APC**: *adenomatous polyposis coli*
- **APS**: persulfato de amonio
- **Array CGH**: Arrays basados en la CGH
- **ATP**: *adenosin triphosphate*
- **BAC**: cromosoma artificial de bacteria
- **BER**: *base excision repair*
- **bp**: par de bases
- **BSA**: *bovine serum albumin*
- **C**: citosina
- **CCHNP**: cáncer colorrectal hereditario no polipósico
- **CCR**: cáncer colorrectal
- **CGH**: hibridación genómica comparada
- **CIAA**: cloroformo: alcohol isoamílico
- **cm**: centímetro
- **CIMP**: fenotipo metilador de islas CpG
- **CIN**: *chromosome instability*
- **CN**: *copy number*
- **CNA**: *copy number alteration*
- **CRISPR**: *clustered regularly interspaced short palindromic repeats*
- **CSCs**: *cancer stem cells*
- **Ct**: ciclo umbral
- **DAPI**: *4',6-diaminido-2-phenilindole*
- **dbSNP**: *database of Single Nucleotide Polymorphisms*
- **DCC**: *deleted in colorectal carcinoma*
- **DMEM**: *Dulbecco's Modified Eagle's Medium*
- **DMSO**: dimethyl sulfoxide
- **DNA**: *desoxyribonucleic acid*
- **dNTP**: *deoxynucleotide triphosphate*
- **DSBs**: *double-strand breaks*
- **EDTA**: ácido etilendiaminotetraacético
- **EGFP**: *green fluorescent protein*
- **EOCRC**: *early onset colorrectal cancer*
- **EtOH**: etanol
- **FBS**: *fetal bovine serum*
- **FFPE**: formalin-Fixed Paraffin-Embedded
- **FISH**: hibridación *in situ* por fluorescencia
- **GPR78**: *binding immunoglobulin protein (BiP)*
- **HDR**: *homology directed repair*
- **HMM**: *Hidden Markov Model*
- **IGF1**: factor de crecimiento tipo insulínico 1
- **IGF2**: factor de crecimiento tipo insulínico 2
- **IMC**: índice de masa corporal
- **kb**: kilobase
- **kDa**: Kilodalton
- **kg**: kilogramo
- **KRAS**: *Kirsten Rat Sarcome*
- **Lefty**: *left-right determination factors*
- **LEMD3**: *LEM Domain Containing protein 3*
- **LOCRC**: *late onset colorrectal cancer*
- **LOH**: *loss of Heterozygosity*
- **M**: molar
- **MAPD**: *median absolute pairwise difference*
- **MAPK**: *mitogen-activated protein kinase*
- **mg**: miligramo
- **MIS**: inestabilidad de microsatélite

TABLA DE ABREVIATURAS

- **ml**: mililitro
- **ML**: millones de lecturas
- **MLH1**: *MutL homolog 1*
- **mM**: millimolar
- **MMR**: *Mismatch Repair*
- **MMS**: estabilidad de microsatélite
- **MSH2**: *MutS homolog 2*
- **MSH6**: *MutS homolog 6*
- **MTT**: Bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolio
- **MUTYH**: *MutY DNA glicosilasa*
- **MSS**: estabilidad de microsatélite
- **μl**: microlitro
- **μg**: microgramo
- **μm**: microméetro
- **μM**: micromolar
- **NCLN**: *Nicastrin like protein (nicalin)*
- **NER**: *nucleotide excision repair*
- **ng**: nanogramos
- **NGS**: *Next Generation Sequencing*
- **NHEJ**: non-homologous end joining
- **nm**: nanometro
- **NOMO**: modulador nodal
- **OMS**: organización mundial de la salud
- **PAF**: poliposis adenomatosa familiar
- **PAFA**: poliposis adenomatosa familiar atenuada
- **PAM**: *protospacer adjacent motif*
- **PBS**: Phosphate Buffered Saline
- **PCR**: reacción de la polimerasa en cadena
- **PI3K**: Phosphatidylinositol 3-kinase
- **PKD1**: *polycystic kidney disease 1*
- **pM**: picomolar
- **pmol**: picomol
- **PMS2**: *Postmeiotic segregation increased 2*
- **PNKT4**: polinucleótido quinasa T4
- **PVDF**: *polyvinylidene difluoride*
- **qPCR**: PCR cuantitativa-comparativa a tiempo real
- **RAR**: regiones recurrentes alteradas
- **RBFOX1**: *Fox-1 homolog A*
- **RE**: retículo endoplasmático
- **REDECAN**: red española de registros de cáncer
- **RNA**: *ribonucleic acid*
- **rpm**: revoluciones por minuto
- **SD**: desviación estándar
- **SDS**: sodium dodecyl sulphate
- **SDS-PAGE**: *sodium dodecyl sulfate polyacrylamide gel electrophoresis*
- **SEOM**: sociedad española de oncología médica
- **sgRNAs**: *single guide RNA*
- **SL**: síndrome de Lynch
- **SMAD**: *mothers against decantaplegic homolog*
- **SNP**: *single nucleotide polymorphism*
- **SP**: sangre periférica
- **STR**: *short tandem repeat*
- **STS**: *steryl-sulfatase precursor*
- **TBE**: *tris borato, EDTA*
- **TBS**: *tris buffered Saline*
- **TBS-T**: *tris buffered saline + tween20*
- **TDGF1**: *teratocarcinoma-derived growth factor 1 precursor*
- **TEMED**: *tetramethylethylenediamine*
- **TERT**: *telomerase reverse transcriptase*
- **TERC**: *telomerase RNA component*

TABLA DE ABREVIATURAS

- **TGF- β** : *transforming growth factor-beta*

- **TP53**: *tumor protein 53*

- **Tm**: temperatura de fusión

INTRODUCCIÓN



1. EL CÁNCER

El término cáncer engloba a un conjunto de enfermedades relacionadas por el crecimiento descontrolado y en cantidades excesivas de las células. Las células cancerígenas también se pueden diseminar hasta otras partes del cuerpo a través del torrente sanguíneo y el sistema linfático. La mayoría de los cánceres adquieren el nombre del órgano o tejido donde se originan, y se producen como consecuencia de una serie de alteraciones genéticas y epigenéticas progresivas (Rubin *et al.* 2011).

Los últimos datos publicados por la SEOM Sociedad Española de Oncología Médica (SEOM) indican que el cáncer sigue constituyendo una de las principales causas de morbi-mortalidad del mundo habiendo aumentado desde los 14 millones de casos estimados en el año 2012 a los 18,1 millones en 2018. Las estimaciones poblacionales indican que el número de casos nuevos aumentará en las dos próximas décadas, alcanzando los 29,5 millones en 2040 (últimos datos disponibles a nivel mundial estimados por los proyectos EUCAN y GLOBOCAN, de la OMS). A nivel mundial, en el año 2018, los datos aportados por GLOBOCAN indican que el cáncer con mayor incidencia fue el cáncer de pulmón con 2.093.876 casos seguido por el cáncer de mama con 2.088.489 casos y en tercer lugar el cáncer colorrectal con un total de 1.849.518 casos (Figura 1).

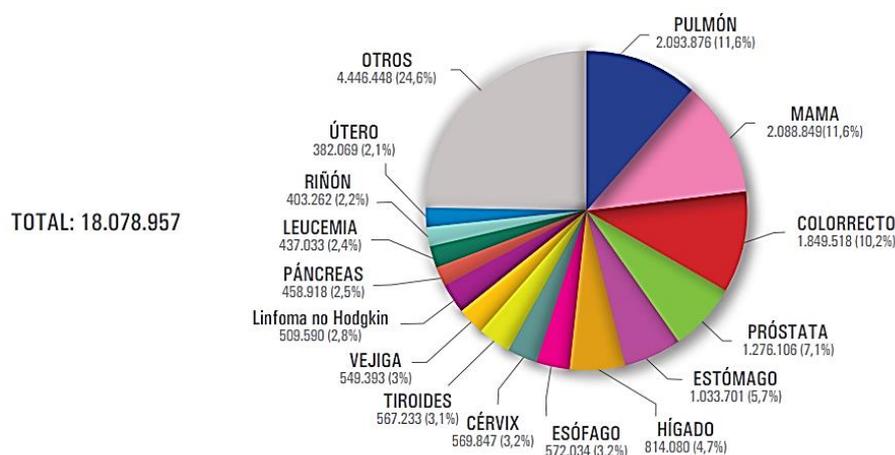


Figura 1. Tumores más frecuentemente diagnosticados en el mundo en el año 2018. Estimación para el año 2018, ambos sexos. Fuente: GLOBOCAN 2018

El número de tumores diagnosticados en España en el año 2019 alcanzará los 277.234, según las estimaciones de REDECAN, en comparación con los 247.771 casos diagnosticados en el año 2015: 148.827 en varones y 98.944 en mujeres. Los cánceres más frecuentes diagnosticados en España en 2019 serán los de colon y recto (44.937 nuevos casos), próstata (34.394), mama (32.536), pulmón (29.503) y vejiga urinaria (23.819). A mucha distancia, los

siguientes cánceres más frecuentes serán los linfomas no Hodgkin y los cánceres de cavidad oral y faringe, páncreas y estómago.

2. EL CÁNCER DE COLON

El colon es la parte del intestino que está comprendida entre el íleon y el recto. Está situado inmediatamente después del ciego y mide aproximadamente 1,5 metros. Entre sus principales funciones destaca la de absorber agua, nutrientes y otras sustancias como electrolitos, minerales y vitaminas de los alimentos. Anatómicamente, el colon se divide en colon ascendente o derecho, colon transverso, colon descendente o izquierdo, sigma y recto (Figura 2). Las capas que forman la pared del colon son: mucosa, submucosa, capa muscular (interna circular y externa longitudinal) y serosa. Además, esta porción del intestino está expuesta a diversos agentes fisiológicos, biológicos y químicos que incrementan el riesgo a desarrollar ciertas enfermedades incluyendo el cáncer (Arévalo *et al.* 2015)

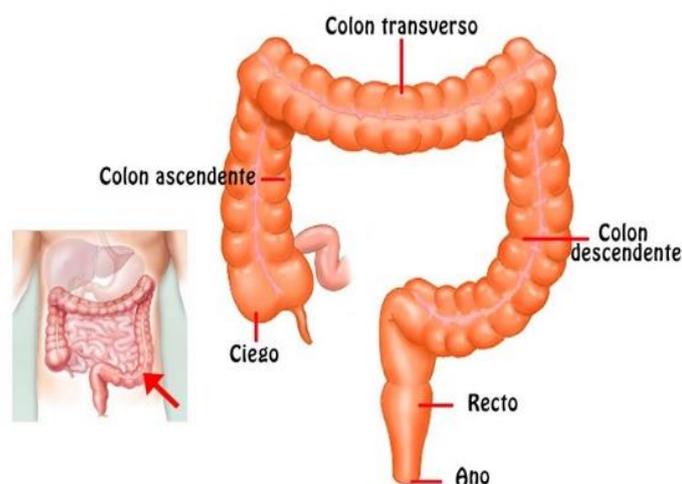


Figura 2. Anatomía del colon.

El cáncer colorrectal (CCR) engloba a un conjunto de neoplasias localizadas en el colon o en el recto cuya histología típica corresponde a un adenocarcinoma (Schneider *et al.* 2014). La mayoría de los tumores colorrectales comienzan como un crecimiento en el revestimiento interno del colon o del recto denominados pólipos (Figura 3).

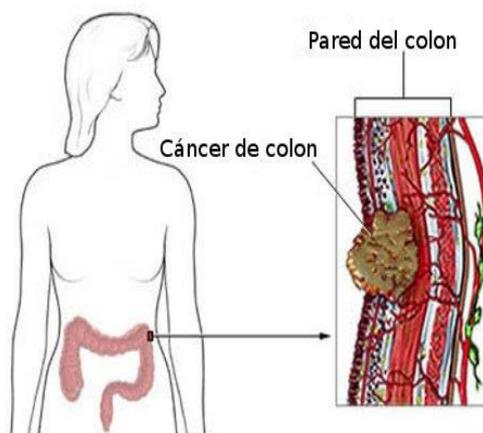


Figura 3. Cáncer colorrectal localizado en el revestimiento interno del colon.

Algunos tipos de pólipos pueden convertirse en CCR con el paso del tiempo (proceso que en la mayor parte de los casos requiere varios años), pero no todos los pólipos se convierten en cáncer (Marley *et al.* 2016). La probabilidad de que un pólipo se transforme en canceroso depende del tipo de pólipo. Los tres tipos principales de pólipos son:

-Pólipos adenomatosos (adenomas): son los más frecuentes y los de mayor potencialidad maligna. Morfológicamente se clasifican en sésiles, pediculados y planos o deprimidos. Estos pólipos pueden mostrar alto o bajo grado de displasia y ser clasificados como tubulares, vellosos o túbulovellosos. Se inician en las glándulas productoras de moco que recubren el colon y el recto (Valastyan *et al.* 2011).

Si un pólipo adenomatoso eventualmente se vuelve canceroso, se forma un tumor en la pared del recto o del colon y, posteriormente, puede invadir los vasos sanguíneos o vasos linfáticos aumentando la posibilidad de metastatizar a otros sitios anatómicos (Valastyan *et al.* 2011). De estos tumores, la gran mayoría (más del 95%) se clasifican como adenocarcinomas (<https://www.cancer.net/cancer-types/colorec-tal-cancer/overview>). La agresividad del tumor está directamente relacionada con el número de pólipos, el tamaño, el grado de displasia y la proporción de componente vellosos (Bas *et al.* 2015, Silva *et al.* 2014).

-Pólipos serrados: los pólipos serrados han sido objeto de atención recientemente, ya que tienen potencial de malignización, considerándose actualmente que alrededor del 35 % de los carcinomas colorrectales se originan a partir de adenomas serrados. Son adenomas que, en la evaluación microscópica, presentan una arquitectura serrada; es decir con pliegues intraluminales al interior de las criptas y en la superficie, lo que genera un aspecto estrellado o dentado. (Rex *et al.* 2012, Sweetser *et al.* 2013).

-Pólipos inflamatorios o pólipos hiperplásicos: los pólipos hiperplásicos, también llamados metaplásicos, se originan por una apoptosis retrasada o fallida en las células epiteliales que lo conforman, por lo que técnicamente no se produce una hiperplasia real. Recientemente, se han identificado en estos pólipos mutaciones en genes reguladores del ciclo celular como *KRAS* o *BRAF*. Endoscópicamente son lesiones pequeñas (menos de 5 mm), sésiles, localizadas en la cresta de los pliegues, generalmente en rectosigmoides (casi el 90%). La mayoría son asintomáticos, siendo un hallazgo incidental durante la colonoscopia. No presentan displasia, carecen de riesgo de malignización y no requieren seguimiento. (Jass *et al.* 2007).

Existen otros cánceres menos comunes que también se desarrollan en la región colorrectal que incluyen tumores carcinoides (que comienzan en células intestinales productoras de hormonas), tumores del estroma gastrointestinal (que se forman en células colónicas especializadas conocidas como células intersticiales de Cajal), linfomas (cánceres del sistema inmunitario que se forman en el colon o recto) y sarcomas (que generalmente comienzan en los vasos sanguíneos pero ocasionalmente se forman en las paredes colorrectales) (Li *et al.* 2014).

2.1. CLASIFICACIÓN DEL CCR

Tradicionalmente, los tumores se han clasificado en función del estadio clínico y de sus características histológicas. En 1932, Dukes propuso la primera clasificación pronóstico para el CCR en función de la capacidad invasiva del tumor. Posteriormente, fueron Astler y Coller los encargados de modificar dicha clasificación que ha sido utilizada hasta hace pocos años. Estos, establecieron 4 estadios diferentes en función de la invasión del tumor y de la metástasis (Tabla 1) (Astler *et al.* 1953).

Tabla 1. Clasificación de Astler y Coller.

Estadios	Extensión Tumoral
A	Tumor limitado a mucosa
B1	Tumor invade hasta muscular propia
B2	Tumor atraviesa la capa muscular hasta la serosa
C1	Igual que B1 con afectación metastásica ganglionar
C2	Igual que B2 con afectación metastásica ganglionar
D	Metástasis a distancia

En el caso del CCR, el sistema que codifica la extensión del tumor primario (T), los ganglios linfáticos regionales (N) y las metástasis a distancia (M) de la Unión Internacional de Control del Cáncer y la del *American Joint Committee on Cancer* es el más utilizado (Edge *et al.*

2010). La clasificación TNM es un sistema dual y comprende una clasificación clínica o pretratamiento y una patológica o posquirúrgica (Tabla 2) (Labianca *et al.* 2010).

Surgió por primera vez en 1977 y desde entonces ha sufrido siete revisiones las cuales han sido fundamentales para abordar las mejoras en oncología, incluidos los avances en la detección temprana, el manejo del paciente, el tratamiento y el descubrimiento de nuevos factores pronósticos y predictivos (Edge *et al.* 2009; Henson *et al.* 2017).

Tabla 2. Clasificación TNM para el CCR, 7ª Ed. AJCC

T	Tumor Primario
T _x	No puede ser valorado
T ₀	No hay evidencia de tumor primario
T _{is}	Carcinoma <i>in situ</i> (intraepitelial/invasión lamina propia)
T ₁	Invade submucosa
T ₂	Invade muscular propia
T ₃	Invade tejidos pericorectales
T _{4a}	Penetra peritoneo visceral
T _{4b}	Invade órganos/estructuras adyacentes
N	GANGLIOS REGIONALES
N ₀	No puede ser valorado
N ₁	Metástasis en 1-3 ganglios regionales
N ₂	Metástasis en 4 o más ganglios linfáticos regionales
M	METÁSTASIS A DISTANCIA
M ₀	No evidencia de metástasis a distancia
M ₁	Metástasis a distancia

Históricamente, la estadificación del cáncer y la planificación del tratamiento se basaban únicamente en la extensión anatómica del tumor. La anatomía sigue siendo un factor pronóstico clave para el cáncer, y la estadificación basada en la anatomía seguirá siendo de gran importancia. Sin embargo, el conocimiento específico cada vez mayor de la biología de las células tumorales permite obtener información pronóstica que complementa y, en algunos casos, es más relevante que la extensión anatómica. Por ello, esta información biológica debe incorporarse en los modelos de pronóstico del cáncer y en la nomenclatura TNM (Edge *et al.* 2010).

2.2. EPIDEMIOLOGÍA DEL CCR

El CCR es la tercera neoplasia maligna que se diagnostica con mayor frecuencia y la cuarta causa de muerte por cáncer en todo el mundo, cuenta con más de 1,8 millones de nuevos

casos diagnosticados y 700.000 muertes por esta enfermedad al año (Orive et al. 2019). Su incidencia tiene clara variación en relación con áreas geográficas. Así, zonas con un nivel mayor de desarrollo, como es el caso de Europa occidental o EE.UU. y Canadá, tienen tasas de incidencia más elevadas, mientras que zonas en vías de desarrollo, como Asia y África, las tasas de incidencia son menores (Boyle et al. 2000). Cabe destacar que, en las dos últimas décadas, en la mayoría de los países desarrollados tanto la incidencia como la tasa de muerte anual, ha disminuido alrededor del 2% para el CCR. Esta rápida disminución se ha atribuido en gran parte al aumento de los programas de cribado entre los individuos con edades igual o superior a los 50 años, lo que permite una detección temprana para esta enfermedad. Por esta razón, el *European Code Against Cancer* y el *American College of Gastroenterology* en los Estados Unidos, recomiendan que la población participe en programas de detección precoz del CCR a partir de 50 años de edad (*European Colorectal Cancer Screening Guidelines Working Group*; Quirke et al. 2012).

En Europa, el CCR es el segundo tipo de tumor más diagnosticado en mujeres después del cáncer de mama y el tercero más común en varones tras los tumores de próstata y pulmón. Después del cáncer de pulmón, es la causa más común de muerte por cáncer en Europa (Driss et al. 2015). La incidencia de este tipo de tumor es ligeramente superior en el sexo masculino que en el femenino (relación 1,4:1). En 2012, se estimó que 241.600 hombres europeos fueron diagnosticados con CCR, y que 113.200 murieron a causa de esta enfermedad. En el caso de las mujeres, se registraron 205.200 casos de CCR y 101.500 muertes relacionadas durante ese mismo año (Ferlay et al. 2013).

En España, los datos de incidencia proporcionados por REDECAN (red española de registros de cáncer; <http://redecana.org/es/index.cfm>) en 2017, revelaron que el CCR fue el tumor más diagnosticado en la población general y el segundo en mortalidad. Fue el tercer tumor más frecuente en varones después del de próstata y pulmón con 20.621 casos diagnosticados durante ese año. En el caso del sexo femenino, el CCR fue el segundo tumor más frecuente tras el de mama con una incidencia de 13.711 nuevos casos diagnosticados durante el año 2017 (“Las cifras del cáncer en España 2018” (SEOM)). Los últimos datos publicados por REDECAN indican que, en España, durante el año 2019 el tumor más diagnosticado en la población será el colorrectal con un total de 44.937 nuevos casos (30.606 de colon y 14.331 de recto). Será el segundo más frecuente en el sexo masculino después del de próstata con 26.746 casos y también el segundo más frecuente en el sexo femenino después del de mama con 11.191 casos (REDECAN) (Figura 4).

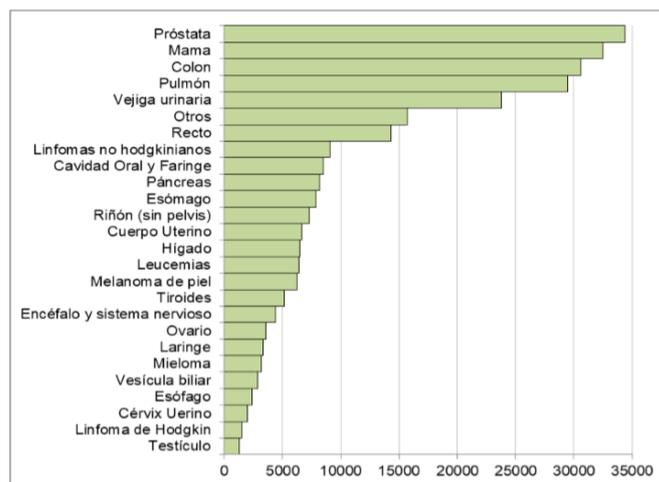


Figura 4. Número estimado de casos incidentes de cáncer en España por tipo tumoral, 2019. Ambos sexos. (Fuente:REDECAN).

2.3. FACTORES DE RIESGO

Los factores de riesgo del CCR se pueden dividir en los siguientes grupos:

2.3.1. FACTORES DE RIESGO RELACIONADOS CON LA FORMA DE VIDA

-Sobrepeso y sedentarismo: la Agencia Internacional para la Investigación del Cáncer (IARC) estableció que existen evidencias suficientes de que el sobrepeso y la inactividad física aumentan el riesgo de padecer cáncer de colon; sin embargo, la relación no está tan clara para el cáncer de recto (IARC, 2002). Se ha planteado la hipótesis de que la obesidad, especialmente la obesidad abdominal y el sedentarismo causan resistencia a la insulina e hiperinsulinemia crónica. La hiperinsulinemia crónica reduce la concentración de las proteínas de unión al factor de crecimiento de tipo insulínico IGF-1 y 2 y al aumento de los niveles tisulares de IGF-1, que desempeña un papel fundamental en el desarrollo y la progresión de los tumores (Bianchini *et al.* 2002; Renehan *et al.* 2008; Bardou *et al.* 2013). Una mayor actividad física puede reducir el riesgo de cáncer de colon y recto ya que reduce el tiempo del tránsito gastrointestinal acortando el tiempo que los carcinógenos están en contacto con las células de la mucosa colónica (Harris *et al.* 2009).

Cuando el índice de masa corporal (IMC) aumenta de 23 kg/m² a 30 kg/m², el riesgo de padecer cáncer de colon aumenta linealmente en aproximadamente el 50% en hombres y en el 25% en mujeres, mientras que el aumento en los niveles de actividad física (ya sea en intensidad, duración o frecuencia) se asocia con una reducción aproximada del 40% en el riesgo de desarrollar cáncer de colon (Arnold *et al.* 2015; Larsson *et al.* 2007; Moghaddam *et al.* 2007).

-Dieta: diferentes publicaciones han demostrado que el consumo de frutas, verduras, fibra y también un aumento en el consumo de productos lácteos, podría tener un pequeño papel preventivo en el desarrollo del CCR (Park *et al.* 2005; Solera *et al.* 2007). Además, también hay estudios sobre la relación entre el consumo de carne roja (ovino, bovino y porcino) y el desarrollo de CCR que muestran resultados dispares, probablemente por la inclusión en algunos de ellos de carnes procesadas o manufacturada y por la distinta forma de preparación de las carnes. Los estudios más recientes con cohortes amplias sí muestran una asociación débil entre consumo de carne roja y el desarrollo de CCR (Norat *et al.* 2015).

-Tabaco y alcohol: es bien conocido que el consumo excesivo de alcohol es un factor de riesgo, tanto en hombres como en mujeres, para el desarrollo de cáncer de colon y de recto (Parajuli *et al.* 2014, Song *et al.* 2015). Además, el tabaco también se ha postulado como un factor de riesgo del desarrollo de pólipos colónicos de cualquier histología, así como del riesgo de adenomas avanzados (Botteri *et al.* 2008).

2.3.2. FACTORES DE RIESGO BIOLÓGICOS

Dentro de estos factores se incluyen las características físicas, como el sexo, la raza o grupo étnico y la edad. Los factores más evidentes implicados en CCR son:

-Edad: estudios recientes han demostrado que el 90% de los individuos con CCR presentan, en el momento del diagnóstico, edades superiores a los 50 años alcanzando su máxima incidencia entre los 65 y los 74 años. El diagnóstico a edades inferiores o iguales a los 45 años es infrecuente y representa menos del 10% de todos los casos (Laish *et al.* 2019; Libutti *et al.* 2015; Hartz *et al.* 2012; Saez *et al.* 2017)).

-Sexo: la incidencia de este tipo de tumor es ligeramente superior en el sexo masculino que en el femenino (relación 1,4:1), sin embargo, no hay ninguna evidencia científica que manifieste un aumento del riesgo a desarrollar CCR por pertenecer a un sexo o al otro (Libutti *et al.* 2015).

2.3.3. FACTORES ASOCIADOS A OTRAS ENFERMEDADES

Se ha descrito que algunas enfermedades que afectan al tracto gastrointestinal inferior, como la enfermedad inflamatoria intestinal en sus dos variantes: enfermedad de Crohn y colitis ulcerosa, incrementan el riesgo de desarrollar cáncer colorrectal (Bae *et al.* 2014; Cunningham *et al.* 2010). El riesgo aumenta con el tiempo de evolución y la severidad de la enfermedad.

La presencia de pólipos también incrementa el riesgo de desarrollar CCR. El riesgo depende del tamaño y del tipo de pólipo. El adenoma veloso y el adenoma veloso-tubular son los que más potencial maligno presentan (20 y 10%, respectivamente). Los pólipos mayores de 2 cm. se asocian también con un alto potencial maligno; así como los pólipos múltiples, que tienen un potencial de malignización dos veces mayor que un pólipo aislado (Amersi *et al.* 2015).

Además, en pacientes con una historia personal de CCR, el desarrollo de cáncer colorrectal metacrónico en los primeros 5 años de seguimiento es del 1.5-3%.

2.3.4. FACTORES DE RIESGO HEREDITARIO

Aunque la mayoría de los casos de CCR son esporádicos, diferentes estudios estiman que el CCR familiar, definido por la presencia de dos o más familiares de primer grado afectados con este tumor, representan más del 20% de todos los casos (Lichtenstein *et al.* 2000; Aaltonen *et al.* 2007). Sin embargo, tan sólo el 5% de los tumores colorrectales tienen una predisposición hereditaria bien definida mediante un patrón de herencia mendeliano (Pérez *et al.* 2010).

Las personas con antecedentes en su familia de CCR tienen un riesgo aumentado a desarrollar esta enfermedad. Además, este riesgo aumenta en los individuos que proceden de familias con síndromes hereditarios, como la poliposis adenomatosa familiar (PAF) o el síndrome de Lynch (SL). En estos casos, el riesgo acumulado a lo largo de la vida en portadores de mutación en genes de alta penetrancia es muy alto (50-90%) (Valle *et al.* 2014).

3. SÍNDROMES HEREDITARIOS DE PREDISPOSICIÓN AL CCR

3.1. SÍNDROME DE LYNCH

El SL es la forma de cáncer colorrectal hereditario más común, manifestándose en un 2-4% de todos los casos de cáncer colorrectal que se diagnostican en la actualidad (Lynch *et al.* 2009). Los pacientes diagnosticados con este síndrome desarrollan CCR a una edad media de 45 años, aproximadamente 20 años antes que la población general. En el CCHNP tiene lugar una aceleración de la carcinogénesis, de modo que la transformación de un diminuto adenoma a carcinoma tiene lugar en 2-3 años, proceso que en la mayoría de los CCR puede llevar mucho más tiempo, incluso hasta 10 años (Perea *et al.* 2015).

El SL constituye una enfermedad hereditaria autosómica dominante en la que la penetrancia de las mutaciones es de hasta el 80%. Estas mutaciones germinales se localizan principalmente en los genes *MLH1*, *MSH2*, *MSH6*, *PMS2* y el extremo 3' UTR del gen *EPCAM*,

también conocidos como genes de reparación de los errores de emparejamiento o genes MMR (*mismatch repair*) (Lynch *et al.* 2009). Dichas mutaciones interfieren en las tareas de replicación y provocan el mal funcionamiento del sistema MMR, favoreciendo la acumulación de errores en la secuencia de DNA. Los errores acumulados pueden afectar tanto a regiones codificantes de genes relacionados con la progresión tumoral (*TGFβ RII*, *IGF1R*, *MSH3* y *BAX*) (Pinheiro *et al.* 2015, Sameer *et al.* 2014) como a unas secuencias altamente repetitivas denominadas microsatélites, siendo éstas últimas, las responsables del fenómeno molecular característico del síndrome: la inestabilidad de microsatélites (MIS) (Stigliano *et al.* 2014).

El tumor más frecuentemente asociado al SL es el cáncer colorrectal; sin embargo, se han descrito multitud de tumores extracolónicos asociados a este síndrome (Tabla 3) (Pinheiro *et al.* 2015, Schneider *et al.* 2014). La frecuencia de aparición de cada uno de ellos es variable y las diferencias clínicas fenotípicas del SL tienen que ver tanto con la alteración molecular en cada caso, así como con la penetrancia variable que presenta este síndrome (Lynch *et al.* 2009). Así, mutaciones en *MLH1* se relacionan con mayor riesgo de desarrollar CCR a una edad precoz, mutaciones en *MSH2* se relacionan con mayor riesgo de desarrollar tumores extracolónicos, mutaciones en *MSH6* se relacionan con mayor riesgo de desarrollar cáncer endometrial, y mutaciones en *PMS2* se relacionan con un aumento de riesgo de desarrollar cáncer endometrial y CCR (Balmaña *et al.* 2013).

Tabla 3. Riesgo de tumores en síndrome de Lynch.

Tipo tumor	Riesgo (%)
Colorrectal (varón)	65-80
Colorrectal (mujer)	30-70
Endometrio	40-60
Ovario	9-12
Gástrico	2-13
Vías urinarias	4-6
Cerebral	4
Vesícula biliar y vías biliares, páncreas	2
Intestino delgado	5
Tumores glándulas sebáceas	0-40

El diagnóstico de SL es complejo dada su heterogeneidad; por este motivo, el Grupo Colaborativo Internacional sobre CCHNP (ahora conocido como *International Society of Gastrointestinal Hereditary Tumours – INSIGHT*) propuso en 1990 un conjunto de criterios para detectar a los sujetos con SL conocidos como criterios de Ámsterdam (Vasen *et al.* 1991).

Estos criterios permitieron utilizar la misma terminología por todos los investigadores; sin embargo, también fueron criticados por ser demasiado estrictos, de modo que excluían a familias con pocos miembros, no incluían tumores extracolónicos y tampoco las características de los tumores colorrectales. Por este motivo, en 1999, se establecieron unos nuevos criterios, denominados criterios de Ámsterdam modificados, (entre los que se incluyeron los tumores extracolónicos) (Vasen *et al.* 1999).

En 1996, se propusieron las guías de Bethesda, que describen casi todas las condiciones clínicas en las que existe sospecha de SL. Diversos estudios han puesto de manifiesto que estos criterios resultan útiles para la identificación de familias que deberán someterse a estudios genéticos para la detección de mutaciones en los genes CCHNP. Recientemente, los criterios utilizados para el diagnóstico de SL son los criterios de Bethesda actualizados (Rodríguez-Bigas *et al.* 1997; Umar *et al.* 2004).

3.2. POLIPOSIS

Las poliposis hereditarias constituyen un grupo de enfermedades con características muy variables (Tabla 4) (Blanco *et al.* 2010).

Tabla 4. Síndromes polipósicos que predisponen al CCR.

Nombre	Gen	Herencia	Tipo de poliposis
Poliposis Adenomatosa Familiar (PAF)	<i>APC</i>	Autosómico dominante	Poliposis adenomatosa
Poliposis Adenomatosa Familiar Atenuada (PAFA)	<i>APC</i>	Autosómico dominante	Poliposis adenomatosa
Poliposis Adenomatosa asociada a <i>MUTYH</i> (PAM)	<i>MUTYH</i>	Autosómico recesivo	Poliposis adenomatosa
Síndrome de Gardner	<i>APC</i>	Autosómico dominante	Poliposis adenomatosa
Síndrome de Turcot	<i>APC, MLH1 y MSH2</i>	Autosómico dominante	Poliposis adenomatosa
Síndrome de Peutz-Jeghers	<i>STK11/LKB1</i>	Autosómico dominante	Poliposis hamartomatosa
Poliposis juvenil familiar	<i>SMAD4 y BRMP1A</i>	Autosómico dominante	Poliposis hamartomatosa
Síndrome de Cowden	<i>PTEN</i>	Autosómico dominante	Otras poliposis
Poliposis mixta hereditaria	Se desconoce el gen	Autosómico dominante	Otras poliposis

3.2.1. POLIPOSIS ADENOMATOSA FAMILIAR (PAF)

La poliposis adenomatosa familiar es un síndrome autosómico dominante que afecta por igual a varones y a mujeres. Esta variante de la poliposis es la más importante. Los pólipos suelen aparecer a partir de los 20 años de edad del paciente. Supone menos del 1% del total de CCR y aunque se hereda de acuerdo a un patrón de herencia autosómico dominante, la tercera parte de los casos no presenta antecedentes familiares de la enfermedad. Su prevalencia es de 1 por cada 10.000-20.000 individuos y con una incidencia estimada en la población española de 2.8 por cada 100.000 habitantes (Laurent *et al.* 2011).

El fenotipo de esta enfermedad se caracteriza por la aparición de cientos a miles de pólipos adenomatosos que tapizan el epitelio del colon y el recto. Los primeros pólipos surgen durante la infancia o la adolescencia y, sin tratamiento quirúrgico, alguno de ellos acaba degenerando en adenocarcinomas durante la edad adulta (Balmaña *et al.* 2013). En principio, las manifestaciones de la enfermedad quedan limitadas a las regiones del colon y el recto; sin embargo, existen casos en los que se observan pólipos en otras zonas del tracto intestinal, por ejemplo, en el duodeno, en el estómago o en la ampolla de Váter.

La PAF también puede presentar manifestaciones extracolónicas: tumores desmoides, hipertrofia congénita del epitelio pigmentario de la retina (*congenital hypertrophy of retinal pigment epithelium*, CHPRE), osteomas, anomalías dentarias, quistes epidermoides y otras neoplasias benignas y malignas de tejidos blandos como el cáncer papilar o folicular de tiroides o el hepatoblastoma en niños (Gryfe *et al.* 2009).

La aparición del síndrome es consecuencia de mutaciones germinales en el gen supresor de tumores *APC* (*adenomatous polyposis coli*). El 80-90% de las familias con PAF, presentan una mutación en el gen *APC* en la línea germinal (Gryfe y col. 2009). La penetrancia de las mutaciones encontradas en este gen es del 100%. El 20% de los casos diagnosticados de poliposis adenomatosa familiar presentan una mutación de *novo* en el gen *APC* (Schneider *et al.* 2014).

3.2.2. OTRAS VARIANTES DE LA POLIPOSIS

-Poliposis adenomatosa familiar atenuada (PAFA): se presenta en un 10% de los pacientes con poliposis adenomatosa. Se caracteriza por la presencia de menos de 100 pólipos en la superficie colorrectal. La edad de aparición de los pólipos o el cáncer suele retrasarse unos 15 años con respecto a la PAF clásica (Giang *et al.* 2018). Un 10% de las familias con poliposis atenuada presenta mutaciones en el gen *APC*. Las mutaciones en *APC* asociadas a esta variante

de la enfermedad aparecen localizadas principalmente en tres regiones del gen: el extremo 5', el exón 9 y en el extremo 3' (Sarvepalli *et al.* 2018).

-Poliposis Adenomatosa Asociada a *MUTYH* (PAM): El 30% de las familias con poliposis atenuada no presentan una mutación en el gen *APC*, sino que tienen mutaciones en el gen *MUTYH* y presenta un patrón de herencia autosómico recesivo (Tieu *et al.* 2016).

-Síndrome de Turcot: se caracteriza por el desarrollo de neoplasias en el sistema nervioso central, que incluyen meduloblastomas, astrocitomas y ependimomas. La mayor parte de estos pacientes también desarrollan características típicas de los pacientes afectados de PAF y un 70% de las familias presentan mutaciones en *APC*. No obstante, hay pacientes con este síndrome que también presentan alteraciones en los genes reparadores del DNA (Sahnane *et al.* 2016; Velter *et al.* 2017).

-Síndrome de Gardner: es una variante de la FAP clásica que incluye las mismas manifestaciones extracolónicas. Actualmente, el síndrome de Gardner y la poliposis adenomatosa familiar se consideran diferentes grados de severidad de la misma enfermedad. En ambos síndromes se han encontrado las mismas alteraciones en el gen *APC*.

-Síndrome de Peutz-Jeghers: es una enfermedad autosómica dominante que se caracteriza por la aparición en la infancia de pólipos hamartomatosos difusos en el intestino y una pigmentación característica en las mucosas de la zona perioral y oral, en la piel de la cara, de los antebrazos, palmas de las manos y los pies, y en la zona perianal (Figura 5) (Henne-ton *et al.* 2017). Los pacientes afectados tienen una alta predisposición a desarrollar neoplasias gastrointestinales y extraintestinales. Entre el 50% y el 90% de los pacientes desarrollan cáncer y un 40% de ellos desarrolla el tumor en el colon. El gen responsable del síndrome de Peutz-Jeghers es el gen supresor de tumores *STK11*, que se ha encontrado mutado en el 50-70% de los pacientes (Koh *et al.* 2016).



Figura 5. Síndrome de Peutz-Jeghers: manchas pigmentarias en piel y mucosas.

Síndrome PTEN-Hamartoma-Tumor: el síndrome de PTEN-Hamartoma-tumor engloba a los síndromes de Cowden, Bannayan-Riley-Ruvalcaba, Proteus y Proteus-like. Todos ellos están causados por mutaciones en el gen PTEN/MMAC1, y se caracterizan por la presencia de múltiples hamartomas que afectan a cualquier zona del cuerpo (Eissing *et al.* 2019).

El síndrome de Cowden es difícil de reconocer, se caracteriza por la presencia de varios hamartomas en diferentes sitios del cuerpo. Los individuos afectados presentan predisposición a desarrollar diversos cánceres, en particular cáncer de mama, tiroides o en de endometrio.

Los individuos afectados por el síndrome Bannayan-Riley-Ruvalcaba presentan macrocefalia, pólipos hamartomatosos en el intestino y tumores adiposos benignos en la piel (lipomas). Además, antes y después del nacimiento, estos individuos muestran un crecimiento exagerado.

El síndrome Proteus es una enfermedad rara que se caracteriza por el crecimiento desproporcionado de algunas partes del cuerpo, principalmente, el tejido óseo, la piel, el sistema nervioso central, los ojos y el tejido conectivo. El término Proteus-like se utiliza para describir aquellos que muestran características similares al síndrome Proteus pero que carecen de los criterios de diagnósticos específicos de esta enfermedad, para el síndrome de Cowden y el síndrome Bannayan-Riley-Ruvalcaba (Biesecker *et al.* 2012).

-Poliposis juvenil familiar: es el síndrome de poliposis hamartomatosa más frecuente. Se hereda mediante un patrón de herencia mendeliano autosómico dominante con una penetrancia variable. Dependiendo de las series, el 20-50% de los afectados presentan historia familiar. La edad media de inicio son los 18 años. Estos pólipos tienen un gran potencial de malignización y pueden desarrollar tumores de los cuales un alto porcentaje aparecen en el colon o el recto. Se han descrito mutaciones en los genes supresores de tumores de la vía de señalización *TGF- β* , *SMAD4* y *BMPR1*, asociados a esta enfermedad (Hsiao *et al.* 2016; Wood *et al.* 2017).

-Síndrome de poliposis mixta hereditaria: se caracteriza por la presencia de una variedad de diferentes tumores colorrectales entre los que se incluyen pólipos adenomatosos, hiperplásicos y hamartomatosos, y carcinomas. Parece afectar solo al colon y no se han descrito manifestaciones extraintestinales (Heald *et al.* 2015).

4. CARCINOGENESIS DEL CÁNCER COLORRECTAL

La mayoría de los CCR se desarrollan a través de una serie de eventos secuenciales en la mucosa colónica normal que pasa por las fases de mucosa en riesgo, adenoma, adenocarcinoma y que puede terminar en metástasis. La progresión morfológica de la mucosa normal a adenoma y de esta a tumor maligno es una manifestación de mutaciones del DNA que afectan a genes reparadores, a oncogenes y a genes supresores de tumores entre otros. La acumulación de estas alteraciones genéticas en el epitelio colónico requiere algunos años, normalmente décadas, lo que está de acuerdo con la edad media de diagnóstico de los pacientes afectados de CCR (alrededor de 65-70 años según las series) (Ogino *et al.* 2008).

Los primeros en describir la existencia de una cascada de alteraciones moleculares como responsables del desarrollo del CCR fueron Fearon y Vogelstein (Fearon y Vogelstein, 1990), que establecieron la secuencia adenoma-carcinoma que llega hasta nuestros días y es la base del modelo de carcinogénesis del CCR más frecuente. Este modelo, describe la carcinogénesis colorrectal y explica la transición del epitelio normal hacia el carcinoma esporádico. Las alteraciones se originan por diversas mutaciones adquiridas en células somáticas, que ocasionan la inactivación de genes supresores de tumores como *APC*, *DCC* y *p53* y la amplificación de oncogenes como *K-RAS* y *β -catenina*.

En las últimas dos décadas, dos descubrimientos moleculares han posibilitado un nuevo enfoque: 1) el hallazgo de la inestabilidad de microsatélites (MIS) causada, como se ha indicado anteriormente, por defectos en los genes reparadores de errores en el DNA (MMR) y, 2) el descubrimiento del papel de la epigenética (en particular, la hipermetilación de las regiones promotoras) en el silenciamiento de la función génica (Jass *et al.* 2002; Guillermo *et al.* 2014). La metilación en la región promotora de múltiples genes da lugar al denominado fenotipo metilador de islas CpG (*CpG island methylator phenotype* o CIMP). Los pacientes con tumores CIMP tienen características clínicas y patológicas diferentes al resto de pacientes con CCR (Al-Sohaily *et al.* 2012).

Las alteraciones genéticas y epigenéticas actúan desregulando las rutas de señalización que participan en el metabolismo celular, la proliferación, la diferenciación, la supervivencia y la apoptosis (Al-Sohaily *et al.* 2012). Existen diferentes mutaciones genéticas que se han relacionado con la carcinogénesis colorrectal, aunque, la función precisa de muchos de estos genes en los que se han localizado estas mutaciones, en la iniciación y progresión de la enfermedad aún no se ha confirmado (Guillermo *et al.* 2014).

Hasta la fecha, se han descrito tres vías moleculares diferentes para explicar el desarrollo del CCR: la vía de la inestabilidad cromosómica (*chromosome instability* o CIN), la vía de inestabilidad de microsatélites (MIS) y, la vía del fenotipo metilador de islas CpG (CIMP). Todas ellas no son mutuamente excluyentes ya que algunos tumores exhiben características de más de una vía (Al-Sohaily *et al.* 2012).

4.1. VÍA DE INESTABILIDAD CROMOSÓMICA (CIN)

Es la vía más común en la carcinogénesis del CCR esporádico: entre el 70% y el 85% de los casos se originan por la vía CIN o tradicional, también llamada vía supresora (Worthley *et al.* 2007). Este tipo de tumores se caracterizan por la ganancia o pérdida de cromosomas enteros o de regiones cromosómicas que albergan genes *drivers* en el proceso de la carcinogénesis colorrectal. La vía CIN resulta de defectos en la segregación cromosómica con aneuploidía posterior, disfunción telomérica o defectos en los mecanismos de respuesta a daño en el DNA. La consecuencia es un desequilibrio en el número de cromosomas (aneuploidía), amplificaciones genómicas cromosómicas y una alta frecuencia de falta de heterocigosidad (Loss of Heterozygosity, LOH) (Arvelo *et al.* 2015, Osorio *et al.* 2014). Junto con estas anomalías cariotípicas se produce la acumulación de mutaciones en oncogenes y genes supresores de tumores (Kanwal *et al.* 2016).

El desarrollo del CCR por la vía CIN es el mecanismo más acorde con el modelo genético propuesto por Fearon y Vogelstein. Se inicia por mutaciones en el gen *APC*, que producen una proteína defectuosa que impide la degradación de la β -catenina y conduce a la formación de un adenoma benigno en el epitelio intestinal. Las mutaciones somáticas en *APC* se observan en el 60-80% de CCR, así como en un gran porcentaje de las lesiones precursoras (adenomas), lo que indica que la mutación en *APC* es un evento temprano en el proceso de tumorigénesis. Posteriormente, ocurre una segunda mutación en el gen *KRAS*, que aumenta la tasa de proliferación y causa la transformación celular hacia un adenoma avanzado. *KRAS* está mutado en el 30- 60% de los CCR y de los adenomas grandes. Se propone que este gen puede desempeñar un papel importante en la transición desde adenoma hacia carcinoma a través de la activación de otras señales (Wang *et al.* 2006). Finalmente, ocurren mutaciones en otros genes como: *DCC*, *SMAD* y *p53*, hasta que progresa a CCR. La pérdida de un segmento del cromosoma 17p (donde se encuentra *p53*) se encontró en el 75% de los CCR, pero no en los adenomas, lo que sugiere que la pérdida del mismo es un acontecimiento tardío en el proceso de tumorigénesis colorrectal. También es común la pérdida del cromosoma 18q, donde se

encuentran múltiples genes supresores tumorales. Esta delección del cromosoma 18q fue identificada en hasta el 70% de los CCR primarios. Los genes en este cromosoma incluyen: *DCC*, *SMAD2* y *SMAD4*. La pérdida de heterocigosidad de 18q y la inactivación secundaria del otro alelo podría ser otra alternativa final en la transición a carcinoma (Matsubara *et al.* 2012). Estas alteraciones le confieren una ventaja de crecimiento a la célula y le proporcionan un alto potencial de metástasis (Figura 6) (Martin *et al.* 2010).

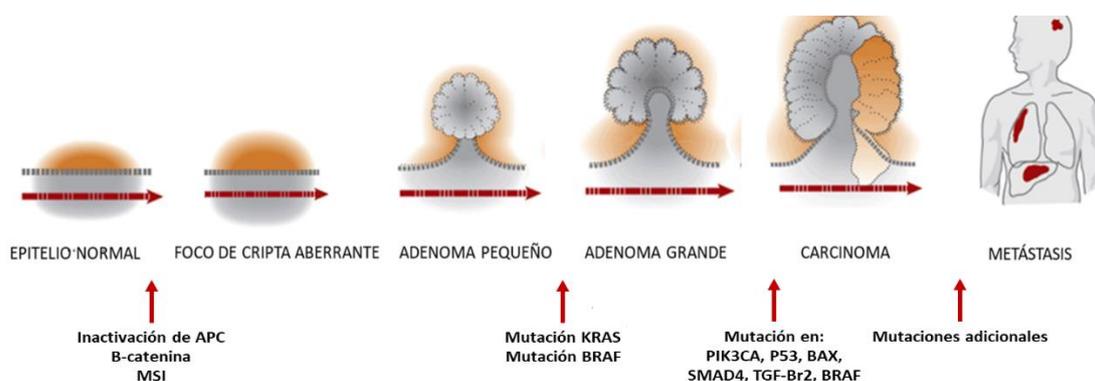


Figura 6. Esquema secuencia adenoma-carcinoma. Diagrama de la acumulación de alteraciones que tienen lugar en la mucosa colónica normal, que pasa por las fases de mucosa en riesgo, adenoma, adenocarcinoma y que puede terminar en metástasis (Modificada de basicmedicalkey).

4.2. VÍA DE INESTABILIDAD DE MICROSATÉLITES (MIS)

La MIS se conoce también como la vía mutadora y ocurre en el 10% de los casos de CCR. Los microsatélites son secuencias de nucleótidos cortas, repetitivas, que se extienden a lo largo de todo el genoma. El hecho de que estas secuencias sean repetitivas es el responsable de que exista una tasa de errores durante la replicación del DNA más elevada que en el resto de las secuencias. Los genes que contienen secuencias repetidas dentro de la región codificante, como *APC*, *MLH1*, *TGF/BR11* y *BAX*, pueden presentar también este tipo de errores.

El sistema de reparación de errores del DNA, MMR, reconoce y repara los desajustes de pares de bases que se producen durante la replicación del mismo (Guinney *et al.* 2015). La inestabilidad de los microsatélites es un reflejo de la incapacidad del sistema MMR para corregir estos errores. Los principales integrantes del sistema MMR en los que se han observado mutaciones son las proteínas *MLH1* y *MSH2*. La pérdida de función de dichas proteínas provoca, además de la MIS, una acumulación de mutaciones somáticas generalizada a lo largo del genoma, en oncogenes y genes supresores de tumores que juegan un papel fundamental en la iniciación y progresión del cáncer (Nayak *et al.* 2018).

Como ya se ha indicado anteriormente, mutaciones en la línea germinal en alguno de los genes MMR aumentan de forma significativa el riesgo a desarrollar SL. Estas mutaciones germinales están presentes en cerca del 3% de todos los casos de CCR, mientras que la mutación somática o el silenciamiento por hipermetilación de genes MMR representa alrededor del 20% de los CCR esporádicos (Roepman *et al.* 2016). El fenotipo MIS en el SL requiere de una doble inactivación bialélica de los genes de la maquinaria MMR. La primera inactivación es la mutación en la línea germinal y la segunda inactivación puede responder a tres posibles mecanismos: mutaciones somáticas, LOH o a una alteración epigenética (Guinney *et al.* 2015).

Finalmente, los análisis de MIS se realizan con un panel de marcadores STR (*short tandem repeat*), conocido como panel de Bethesda, entre los que se encuentran los polinucleótidos de adenina BAT25, BAT26 y BAT40, además de los dinucleótidos repetidos D5S346, D2S123, D17S250. Los tumores se clasifican como MIS alta (MIS-H), cuando muestran dos o más marcadores inestables y MIS baja (MIS-L) cuando tienen inestable un solo marcador; y estables (MSS) cuando no se encuentra inestabilidad en ninguno de los STR (Nardoen *et al.* 2010). Los análisis de MIS son útiles para el cribado molecular de individuos con sospecha de SL (Wong *et al.* 2010), puesto que este síndrome se relaciona con una alta frecuencia de tumores MIS-H, mientras que los CCR esporádicos tienen un porcentaje menor cercano al 20% (Poulogiannis *et al.* 2010).

4.3. VÍA SERRADA O VÍA DEL FENOTIPO METILADOR DE ISLAS CpG (CIMP)

Las alteraciones epigenéticas producen cambios en la expresión génica o en su función, sin cambiar la secuencia de DNA. La aparición del CCR puede ocurrir por mecanismos epigenéticos que inactivan la expresión de diversos genes. El mecanismo implicado mejor conocido es la metilación del DNA, aunque existen otras, como las modificaciones en las histonas (Matsubara *et al.* 2012). La metilación de genes en la región promotora resulta en el silenciamiento de estos; por lo tanto, supone un mecanismo alternativo para la pérdida de la función de los genes supresores de tumores (Guillermo *et al.* 2014). La metilación del DNA es la segunda vía genética más común en el CCR esporádico, se presenta aproximadamente en el 15% de los casos y consiste en la unión de un grupo metilo en el carbono 5 de la citosina por acción de las enzimas DNA metiltransferasas, lo que genera la 5 metil-citosina, también denominada la quinta base.

Las modificaciones por metilación ocurren en las citocinas de los dinucleótidos CpG, conocidas como islas CpG, que se localizan en los promotores de casi la mitad de todos los genes

y se encuentran generalmente no metiladas en los genes activos (Godai *et al.* 2009). En las células cancerosas, la hipermetilación de las islas CpG de los promotores bloquea la expresión de genes, como los supresores de tumores y los de reparación del DNA (Figura 7) (Anderson 2014; Yamane *et al.* 2014). Este fenómeno de hipermetilación se conoce como fenotipo metilador de las islas CpG o CIMP. Por el contrario, la hipometilación de los promotores de algunos genes ocurre en estadios avanzados del cáncer y se caracteriza por el aumento en la transcripción génica (Kim *et al.* 2018).

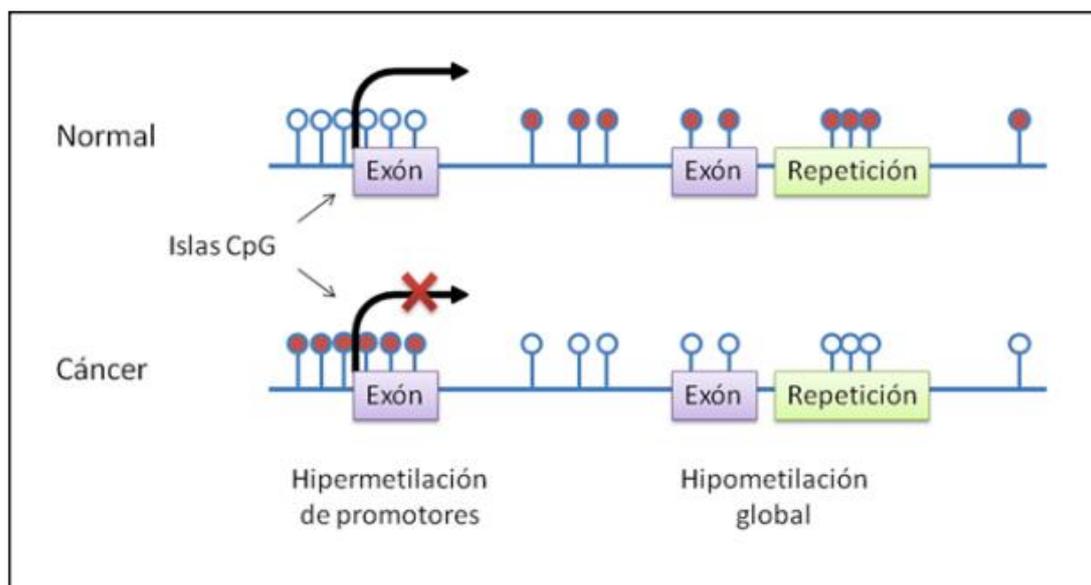


Figura 7. Hipermetilación de genes promotores y silenciamiento global (D' Ottavio *et al.* 2014).

La presencia de CIMP en el tumor se puede identificar por inmunohistoquímica y, en consecuencia, se pueden clasificar los tumores en: CIMP alto o CIMP-H (*high*) y CIMP bajo o CIMP-L (*low*) (D' Ottavio *et al.* 2014). El CIMP-H se observa en el 15-20% de los CCR esporádicos y presenta características que difieren de los tumores con CIMP-L (Yang *et al.* 2017). La identificación de tumores con CIMP-H es más común en mujeres y pacientes mayores, el origen del tumor suele ser en el colon proximal (colon derecho); además, cerca del 15% del CCR con CIMP-H se relaciona con la metilación de islas CpG en el promotor del gen *MLH1*. Se ha observado también que la mitad de todos los tumores de CCR con CIMP-H presentan MIS-H debido a la inactivación del gen de reparación *MLH1*; en estos casos se concluye que existe una relación clara entre la vía de inestabilidad de microsatélite y la vía del fenotipo metilador (Martinez-Ureña *et al.* 2013). Histológicamente, los tumores CIMP-H son a menudo poco diferenciados, de histología mucinosa o en anillo de sello, con inestabilidad de microsatélites, y

presentan BRAF mutado (protooncogén similar a K-Ras). Las lesiones precursoras son generalmente los pólipos serrados sésiles (Arvelo *et al.* 2015).

4.4. CLASIFICACIÓN MOLECULAR DEL CCR

En el CCR no se ha implementado todavía una clasificación molecular estandarizada internacionalmente, sin embargo, existen diferentes aspectos patogénicos que podrían servir de base (Haq *et al.* 2009). En 2015 se publicó *The Consensus Molecular Subtypes of Colorectal Cancer* (Guinney *et al.* 2015) en el que se identificaron cuatro subtipos moleculares. Esta clasificación se basa en la superposición de las tres vías. Partiendo del estado MIS y CIMP, parece probable la existencia de al menos cuatro grupos homogéneos con características anatomoclínicas y moleculares diferentes, en los que además parece haber correlación entre las categorías MIS y CIMP (tumores MIS-High se asocian mayoritariamente con genotipos CIMP-High y tumores MIS-Low se asocian mayoritariamente con genotipos CIMP-Low/0) (Leonardo *et al.* 2018; Ogino y Goel, 2008):

- 1- **Grupo 1 (CMS1): CIMP-H / MIS-H** (14% de los CCR); estos tumores muestran inestabilidad de microsatélites, fenotipo hipermutado, *BRAF* mutado y metilación en la región del promotor del gen *MLH1*. Estos tumores fueron más frecuentes en el sexo femenino, mayor localización en el colon proximal, se suele diagnosticar en estadios avanzados y con peor supervivencia después de la recaída.
- 2- **Grupo 2: CIMP-L / MIS-H** (5% de los CCR); se genera por mutaciones en los genes del sistema MMR y está presente en el cáncer colorrectal hereditario no polipósico (SL). También presentan una localización mayoritaria en el colon proximal y características mucinosas pero, no están pobremente diferenciados.
- 3- **Grupo 3: CIMP-H / MSS** (5-10% de los CCR); se originan a partir de pólipos serrados y se caracteriza por presentar *BRAF* mutado (c.1799T>A) y la metilación de múltiples genes sin incluir *MLH1*. Son tumores pobremente diferenciados con características de células en anillo de sello, se asocian con mujeres ancianas y la localización del tumor predomina en el colon derecho.
- 4- **Grupo 4: CIMP-L / MSS** (75% de los CCR); presente en adenomas convencionales, estos tumores se caracterizan por tener un elevado número de ganancias y pérdidas de material genético. Estos tumores se presentan comúnmente en el lado izquierdo. Los pacientes tuvieron mejores tasas de supervivencia después de la recaída.

5. CÁNCER COLORRECTAL EN JÓVEN VS. CÁNCER COLORRECTAL EN ADULTO

Estudios recientes han establecido diferencias clínicas, genéticas y biológicas según la edad de inicio del CCR, lo que sugiere que, la edad, debería de ser un criterio importante para la subclasificación del CCR (Pillozzi *et al.* 2015). Se pueden establecer dos entidades diferentes: el CCR de inicio temprano (EOCRC (*early onset colorrectal cancer*)) y el CCR de inicio tardío (LOCRC (*late onset colorrectal cáncer*)) con puntos de corte variables en la literatura que oscilan entre los 45 o 50 años para el EOCRC y 70 o 80 años para el LOCRC (Perea *et al.* 2013).

Se asume que el CCR es una enfermedad del adulto mayor, con un 85-90% de casos diagnosticados después de los 55 años; no obstante, puede afectar a toda la población (Koh *et al.* 2015). En países desarrollados se ha evidenciado una disminución de su incidencia en las últimas décadas; sin embargo, últimamente la prevalencia en adolescentes y adultos jóvenes ha ido en aumento (Creaney *et al.* 2014, Siegel *et al.* 2009). En Estados Unidos, la incidencia de CCR en adultos jóvenes durante 1992 a 2005, aumentó en 1,5% para varones y 1,6% para mujeres (Siegel *et al.* 2009).

En la actualidad, el EOCRC representa una enfermedad muy heterogénea en la que el componente hereditario puede estar presente, aunque la mayoría de los casos son esporádicos (Arriba *et al.* 2016). Diferentes alteraciones moleculares contribuyen a la "heterogeneidad" del EOCRC y existen pruebas que sugieren que, en comparación con el LOCRC, el CCR de inicio temprano puede tener un perfil molecular diferente (Tabla 5). El aumento en la incidencia del EOCRC junto con la naturaleza agresiva de estos tumores, enfatizan no solo la importancia de conocer los factores de riesgo en este grupo de edad, sino también la importancia de la evaluación temprana en individuos jóvenes con síntomas de la enfermedad (Silla *et al.* 2014).

Tabla 5. Diferencias clínico-patológicas y moleculares entre el CCR diagnosticado en el adulto joven y el CCR diagnosticado en el adulto anciano.

Características clínicas y moleculares	EOCRC (≤50años)	LOCRC (>70años)
Colon proximal		X
Colon distal y recto	X	
Sincrónicos y metacrónicos	X	
Etapas III y IV en el momento del diagnóstico	X	
Mucinoso/células en anillo de sello y pobremente diferenciado	X	
Mayoritariamente MSS	X	
MIS debido a la metilación de la región promotora del gen MLH1		X
CIN		X
CIMP (-)	X	
CIMP (+)		X
MSS y CIN (-) (MACS)	X	

Diferentes estudios han demostrado que los individuos con CCR diagnosticado a edades tempranas tienen características histológicas diferentes a los individuos con edades más avanzadas como, por ejemplo, la localización del tumor y el estadio de este al diagnóstico. En comparación con el LOCRC, el EOCRC se presenta con mayor frecuencia en el colon distal y el recto (Berger *et al.* 2006). El EOCRC también se asocia con un mayor porcentaje de tumores sincrónicos y metacrónicos. Un estudio realizado por Liang *et al.*, que evaluó las características clínico-patológicas y moleculares de los EOCRC, mostró una mayor incidencia de tumores sincrónicos (5,8% frente a 1,2%, $p = 0,007$) y metacrónicos (4,0% frente a 1,6%, $p = 0,023$) en individuos jóvenes (menores de 40 años) en comparación con individuos mayores (Liang *et al.* 2003). Según datos de la *National Cancer Database*, en comparación con el CCR de inicio tardío, el EOCRC presentó con más frecuencia una histología mucinosa y de células en anillo de sello (12.6% vs 10.8%, $P < 0,001$) y escasa o ninguna diferenciación (20.4% vs 18 %, $p < 0,001$) (Ahn *et al.* 2014). La razón de estas diferencias histológicas es desconocida, pero las diferencias en la biología molecular de estos tumores pueden ser una posible explicación. También se ha comprobado que el diagnóstico de la enfermedad en estadios avanzados es más común en el EOCRC. Los datos del SEER (*Surveillance, Epidemiology, and End Results*) mostraron que los individuos jóvenes (20-40 años) con CCR tienen una supervivencia general inferior a 5 años en comparación con los individuos mayores (60-80 años) (61.5% frente a 64.9%; $p = 0,02$) (Liang *et al.* 2003). Sin embargo, las tasas de supervivencia en el EOCRC son iguales o incluso superiores a las del LOCRC (Yeo *et al.* 2013).

6. CÁNCER COLORRECTAL EN EL ADULTO JOVEN

6.1. CARACTERÍSTICAS MOLECULARES DEL CÁNCER COLORRECTAL EN EL ADULTO JÓVEN

La patogénesis del EOCRC está bien caracterizada en individuos con SL, en los que, existe una mutación en línea germinal en algún gen del sistema MMR, que se relaciona con un aumento en la susceptibilidad a desarrollar cáncer de colon (Perea *et al.* 2014). Sin embargo, el conocimiento acerca de las características moleculares del EOCRC esporádico es bastante más limitado (Yeo *et al.* 2013). Estudios recientes han mostrado que el CCR esporádico diagnosticado en los pacientes jóvenes puede tener un perfil molecular diferente al CCR vinculado al SL (Ballester *et al.* 2016). El CCR esporádico de inicio precoz puede atribuirse al efecto acumulativo de múltiples variantes genéticas que muestran una penetrancia variable (Stigliano *et al.* 2014).

Una mejor comprensión de estos mecanismos moleculares ayudará a diseñar estrategias específicas de prevención y seguimiento de la enfermedad.

Los tumores MIS en la población joven están relacionados con el SL y, rara vez se relacionan con la inactivación epigenética del gen *MLH1* (Antelo *et al.* 2012). Estudios recientes han demostrado que la proporción de MIS encontrada en el CCR de inicio joven varía de 19.7% a 41.0% dependiendo de la edad de inicio (Ahnen *et al.* 2014). La elevada proporción de tumores MIS en la población joven se ha atribuido al elevado número de pacientes con SL en este grupo de edad. Un estudio realizado por Yiu *et al.* mostró que los tumores con MIS en la población mayor (mayores de 60 años) estaban asociados con la inactivación de *MLH1* (83%) y con la metilación de la región promotora de este gen (62%). Por el contrario, los tumores MIS de inicio precoz (<45 años) se asociaron con la inactivación del gen *MSH2* (Yiu *et al.* 2005).

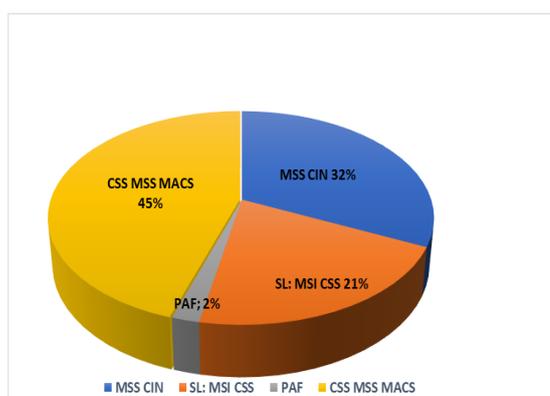


Figura 8. Perfil molecular del CCR en el adulto joven. (Modificado de Ballester *et al.* 2016)

Sin embargo, la MIS no explica la mayoría de los casos de CCR de inicio temprano, porque todavía hay aproximadamente un 85% de estos tumores que exhiben MMS. Los tumores MSS diagnosticados en pacientes jóvenes están mal caracterizados. Diferentes estudios han demostrado que estos tumores son notablemente diferentes de los MSS de inicio tardío. La ubicación del tumor en el colon izquierdo, la baja frecuencia de otras neoplasias primarias y un componente familiar importante son características significativas del MSS-EOCRC (Perea *et al.* 2014).

También se ha propuesto la existencia de un subconjunto específico de tumores CCR de inicio temprano denominados MACs. Los tumores MACs son aquellos tumores MSS que, debiendo presentar CIN, no lo hacen (Osorio *et al.* 2014). Son por tanto tumores MSS sin CIN que tampoco pertenecen estrictamente a la vía serrada. Pueden representar hasta el 30% de los CCR esporádicos, y sus diferentes implicaciones clínicas y pronósticas podrían ser reflejo de vías de carcinogénesis todavía no identificadas (Osorio *et al.* 2014). Este tipo de tumores se localizan

con mayor frecuencia en el colon distal y el recto, tienen características histológicas asociadas con un pronóstico desfavorable, presentan metástasis en el momento del diagnóstico y muestran una recurrencia temprana de la enfermedad y una supervivencia más baja que los pacientes con MIS o CIN (Banerjee *et al.* 2009). Además, estudios recientes han confirmado que este subconjunto de tumores presenta CIMP-L, rara vez se asocian con mutaciones en *BRAF*, tienen ausencia de expresión de *MLH1* y parecen tener un patrón diferente de hipometilación cuando se comparan con los MIS y CIN (Silla *et al.* 2014).

Por otra parte, los CCR diagnosticados en pacientes jóvenes tienen una mayor tasa de CIMP-L. Sin embargo, dentro de los pacientes jóvenes con SL una proporción mayor presentan CIMP-H cuando se comparan con aquellos pacientes con LS que desarrollan CCR a una edad mayor. De hecho, existe una asociación entre tumores EOCRC con CIMP-H y mutaciones germinales en los genes MMR. Por el contrario, los tumores LOCRC con CIMP-H son más frecuentemente tumores esporádicos con MIS (Perea *et al.* 2014).

6.2. PERFIL CROMOSÓMICO DEL EOCRC

La incorporación de herramientas como el CGH (*comparative genomic hybridization*) al estudio del CCR ha permitido identificar diferentes regiones cromosómicas y genes implicados en diferentes vías metabólicas (WNT, RAS-MAPK, PI3K, TGF- β) frecuentemente alterados en la enfermedad (Arvelo *et al.* 2015, Bass *et al.* 2011).

En el caso del CCR, existen regiones cromosómicas con pérdida o ganancia de material genético que parecen estar afectadas de forma muy repetida, por lo que han sido denominadas “regiones recurrentemente alteradas” o RAR (*recurrently altered regions*). De hecho, la elevada frecuencia de aparición de algunas de estas alteraciones ha llevado a sugerir su utilidad como marcadores pronósticos; sin embargo, a día de hoy continúa sin existir una visión integrada acerca del significado que estos cambios pueden tener sobre la carcinogénesis colorrectal (Arriba *et al.* 2015). Dentro de estas regiones RAR se incluyen ganancias de material genético en 7p (que contiene *EGFR*), 8q (que contiene *MYC*), 13q y 20q (que contiene *ZNF217*, *CYP24* y *Aurora2Kinasa*), y pérdidas en 8p (que contiene *FGFR1*), 17p (que contiene *TP53*) y 18p/q (que contienen *SMAD4* y *DCC*, respectivamente) (Cancer Genome Atlas Network, 2012; Arriba *et al.* 2015).

Existen numerosos estudios en los que se han correlacionado las características clínico-patológicas y el perfil molecular en los tumores de aparición tardía, sin embargo, esta relación apenas está estudiada en los tumores EOCRC, posiblemente debido a la baja frecuencia de CCR

en personas jóvenes. Existen indicios que otorgan credibilidad a la suposición de que los diferentes fenotipos observados en personas jóvenes y ancianas puedan tener su origen en mecanismos moleculares también diferentes: por un lado, ambos tumores difieren en cuanto a sus características clinicopatológicas y, por el otro, la carcinogénesis asociada a tumores esporádicos de inicio temprano parece implicar un subtipo de la vía CIN, sin participación de la vía CIMP y con ausencia de mutaciones en *BRAF* (Kirzin *et al.* 2014). De ser esto así, la comparación y búsqueda de diferencias entre los perfiles genómicos de ambos tipos de tumores y posterior indagación de los genes localizados en las regiones diferenciales, podría proporcionar información acerca de nuevas vías de carcinogénesis específicas de cada grupo de edad (Arriba *et al.* 2015).

Un estudio reciente, en el que colaboró nuestro grupo, en el que se realizó un análisis exhaustivo de las alteraciones en el número de copias de DNA (CNAs) que se producen en dos grupos de tumores procedentes de pacientes con diferentes edades (menores de 45 años y mayores de 70 años) concluyó que el 97,3% de todos los tumores, sin tener en cuenta ningún criterio de edad, mostró un elevado índice de inestabilidad genómica. Los tumores de inicio precoz se caracterizaron por un menor número de segmentos alterados, pero una mayor cifra de cambios recurrentes, donde se llegaron a alcanzar porcentajes de frecuencia del 61%, mientras que, en los ancianos, ningún cambio recurrente superó una frecuencia del 47%. Entre las alteraciones recurrentes observadas, hubo una primera fracción que resultó común a los dos grupos, lo cual puede significar la existencia de genes que proporcionan cierta ventaja adaptativa a las células tumorales independientemente de la edad de inicio de la enfermedad. En cuanto a la segunda fracción de las alteraciones, presentes en los dos grupos, pero con recurrencias estadísticamente diferentes, permitió establecer tres categorías en función del grado de disparidad observado entre los porcentajes de frecuencia vinculados a cada población (Arriba *et al.* 2016):

a) Alteraciones muy frecuentes en jóvenes y muy infrecuentes en ancianos: grado de disparidad muy alto. Incluyeron ganancias en 1p35.2-p35.1 y 3p21.31- p21.1, y pérdidas en 1p35.2-p35.1, 7q11.21, 9p13.1-q21.1, 16p13.11 y 16p12.3.

b) Alteraciones más frecuentes en jóvenes que en ancianos: grado de disparidad alto. Esta categoría únicamente incluyó pérdidas que se localizaron en 1p36.33, 1p12-p11.2, 1q21.1, 1q21.1-q21.2, 3p21.31-p21.1, 7q22.1, 7q35 y 14q11.1- q11.2.

c) Alteraciones más frecuentes en ancianos que en jóvenes: grado de disparidad alto. Sólo hubo tres segmentos que demostrasen mayoría significativa entre la población anciana.

Todos correspondieron a ganancias y se encontraron localizados en el cromosoma 7 (7q11.23, 7q22.1 y 7q11.22).

Aunque por el momento no se conocen con detalle los mecanismos moleculares que subyacen tras el CCR en el adulto joven, su carcinogénesis parece estar vinculada a regiones cromosómicas que no son las más habituales en el desarrollo del CCR. Los resultados obtenidos en el trabajo de Arriba *et al.* indican que las ganancias en 1p35.2-p35.1 y 3p21.31-p21.1 y las pérdidas en 1p36, 1p12, 1q21, 9p13, 14q11, 16p13 y 16p12 son más frecuentes en los tumores de inicio temprano, mientras que las ganancias en 7q11 y 7q22 predominan en los de inicio tardío (Arriba *et al.* 2016).

Teniendo en cuenta todas estas diferencias a nivel cromosómico entre ambos grupos de tumores, parece razonable pensar que tales diferencias puedan tener su origen en desviaciones ocurridas durante alguna etapa de su proceso de formación y que, por tanto, en el CCR están implicados distintos mecanismos de carcinogénesis en función de la edad (Berg *et al.* 2010).

7. DELECIÓN DE LA REGIÓN 16p13.12-p13.11

Los análisis de todos estos datos publicados en el trabajo de Arriba *et al* reveló una delección recurrente en el brazo corto del cromosoma 16, concretamente en la región 16p13.12-16p13.11 que se podía presentar como delección única en esta región o junto con otras alteraciones cromosómicas, bien sean ganancias o pérdidas de material genético (Figura 9) (Perea *et al.* 2017).

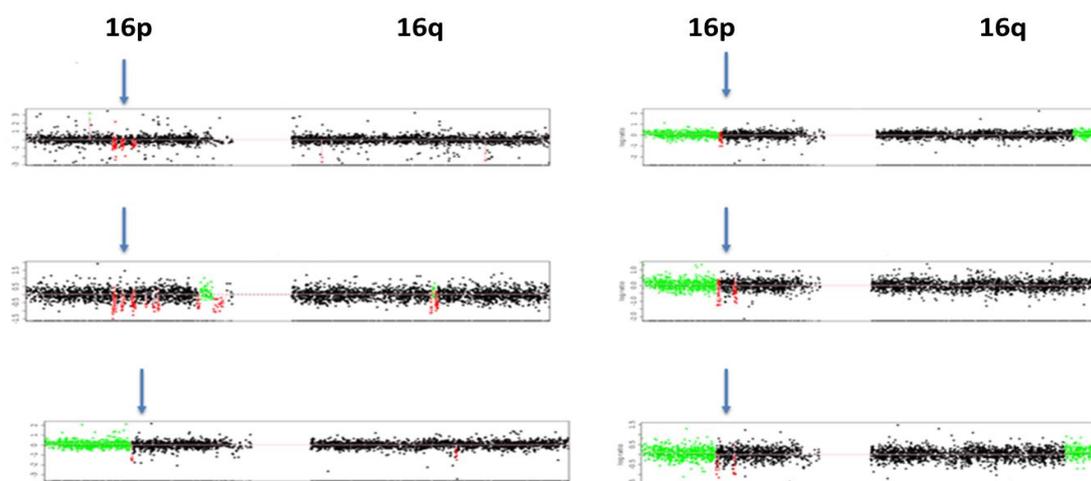


Figura 9. Resultados de la segmentación de la región 16p13.12-p13.11 para seis de las muestras analizadas. El análisis de este cromosoma muestra una pequeña delección en el brazo 16p. Las relaciones logarítmicas normalizadas de las intensidades de la sonda se representan como puntos de diferente color. Las ganancias de material genómico se muestran en verde, las pérdidas se muestran en rojo y los números de copia normales se muestran en negro (Perea *et al.* 2017).

Esta pequeña zona que presentaba pérdida de material genético era más común en el grupo de los jóvenes que en el de los ancianos (33% vs. 16,3%) (Perea *et al.* 2017).

Se han descrito varias enfermedades asociadas a esta región cromosómica. Las microdeleciones en la región 16p13.11 hacen que las personas que las presentan puedan tener retraso en el desarrollo, discapacidad intelectual con o sin múltiples anomalías congénitas asociadas que afectan al corazón, riñón, huesos, dientes y dedos tanto de las manos como de los pies (Ramalingam *et al.* 2011). También se observaron algunos hallazgos aislados como microcefalia, estatura pequeña y estrabismo. Además, existen otras enfermedades asociadas a la presencia de microduplicaciones en esta región, aunque inicialmente se consideraban como variantes benignas raras, la evidencia acumulada indica que estas duplicaciones son muy comunes en pacientes con autismo, esquizofrenia, epilepsia y trastorno por déficit de atención e hiperactividad (Ingason *et al.* 2011).

La región 16p13.12-p13.11 es un *hotspot* genómico particularmente rico en repeticiones de pequeñas secuencias de DNA (LCRs *low copy repeats*) altamente homólogas que aumentan la probabilidad de alteraciones en el número de copias a través de recombinaciones homólogas no alélicas. Estas recombinaciones son, probablemente, el mecanismo subyacente que conduce a la formación de microdeleciones / microduplicaciones (Ramalingam *et al.* 2011). Esta región puede subdividirse en tres fragmentos de secuencias de copia única denominados intervalo I, II y III cada uno de los cuales están flanqueados por secuencias LCRs con una homología entre las regiones codificantes de cada intervalo del 99% (Ingason *et al.* 2011). Las deleciones típicas en esta región tienen un tamaño muy variable (de 0.8 a 3.3 Mb) y pueden llegar a abarcar los tres intervalos (Tropeano *et al.* 2013). El intervalo II representa la región crítica respecto a las variaciones genómicas ya que es donde se han identificado la mayoría de las microdeleciones descritas hasta el momento (Ingason *et al.* 2011).

La región correspondiente a la deleción encontrada en el 23,3% de los tumores de colon analizados en el trabajo de Arriba *et al.* se localiza en el intervalo II. Esta región abarca un conjunto de genes donde, entre otros, se incluyen:

-*ABCC6*: codifica MRP6, un miembro de la familia de proteínas ABC (*ATP-Binding-Cassette*) con un papel fisiológico desconocido (Bakircioglu *et al.* 2011). La mayoría de los miembros de esta familia de proteínas están involucrados en el transporte transmembrana. El gen *ABCC6* humano y sus dos pseudogenes comparten un 99% de homología en sus secuencias de nucleótidos. La pérdida de función de la proteína MRP6 se asocia con el desarrollo de

pseudoxantoma elástico (PXE), un trastorno hereditario recesivo que afecta a los tejidos elásticos (Tamás *et al.* 2005).

-*PDXDC1*: codifica la proteína *Pyridoxal-Dependent Decarboxylase Domain-Containing Protein 1*. Apenas se conoce nada acerca de esta proteína. Se trata de una enzima que podría metabolizar los neurotransmisores de catecolamina. Determinados estudios han demostrado que esta proteína está altamente expresada en el hipocampo y que sus niveles de expresión están inversamente correlacionados con el PPI (*acoustic prepulse inhibition*) por lo que podría ser un buen objetivo para investigación como tratamiento potencial contra la esquizofrenia (Feldcamp *et al.* 2017).

-*NTAN1*: representa otro gen candidato importante en el locus 16p13.11. Codifica la amilasa N-terminal específica de la asparagina, una enzima que participa en la regulación de la vida media de las proteínas. La inactivación del gen *NTAN1* en ratones se ha asociado con características neurológicas anormales, como el comportamiento social alterado, el aprendizaje alterado y la memoria alterada (Kettunen *et al.* 2012).

-*BFAR*: regulador bifuncional de la apoptosis. La proteína que codifica es una proteína multidominio que se descubrió por primera vez como un inhibidor de la apoptosis celular inducida por BAX. Además, la proteína BFAR contiene un dominio que media la unión de la caspasa-8. Por tanto, se propuso que esta proteína actúa como nexo de unión entre los componentes de las vías de apoptosis extrínseca e intrínseca. Estudios más recientes sugieren que las funciones biológicas de la proteína BFAR probablemente sean relevantes para la supresión de la muerte celular y el mantenimiento de la supervivencia de las neuronas (Li *et al.* 2012).

-*NPIPA2*: se trata de un miembro de una familia de genes específicos de primates que codifica una proteína que contiene un dominio transmembrana y se encuentra en la envoltura nuclear en células cultivadas.

-*PLA₂G₁₀*: este gen codifica la fosfolipasa secretora A2. Forma parte de una familia de enzimas relacionadas estructuralmente que se encuentran tanto en los mamíferos como en los venenos de insectos y serpientes. Estas enzimas catalizan la hidrólisis de los glicerofosfolípidos produciendo ácidos grasos libres y lisofosfolípidos. Se encargan de generar precursores limitantes de la velocidad en la biosíntesis de diversos tipos de lípidos biológicamente activos, incluidas las prostaglandinas, los ácidos grasos hidroxilados, los leucotrienos, los tromboxanos y el factor activador de plaquetas (Kettunen *et al.* 2012).

-*RRN3*: codifica el factor de iniciación de la transcripción específico de la RNA polimerasa. Es un factor de transcripción asociado a la RNA polimerasa I que es esencial para la transcripción del DNA ribosomal (rDNA). El modelo generalmente aceptado es que *RRN3* funciona como un puente de unión entre la RNA polimerasa I, concretamente con la subunidad Pol I, y el factor de selectividad del promotor SL1 formando un complejo de preiniciación de la transcripción en los promotores de los genes que codifican los rRNA (rDNA) (Stepanchick *et al.* 2013).

- *PKD1P6*: se trata de un pseudogen del gen *PKD1*. Las enfermedades asociadas con *PKD1P6* incluyen la enfermedad hepática poliquística y la enfermedad renal poliquística. (Xu *et al.* 2018).

-*NOMO1*: Modulador Nodal 1. Este gen codifica una proteína que, aunque su función no está muy clara, se sabe que forma un complejo proteico, junto con la proteína NICALINA (NCLN) y TEMEM147, que actúa como antagonista de la vía Nodal (Zhang, *et al.* 2015). Hay evidencias de una elevada actividad de esta vía en determinados tipos de tumores. Por ejemplo, el incremento en los niveles de expresión de las proteínas que actúan como ligandos de la vía Nodal en el melanoma maligno está correlacionado con la progresión mientras que la inhibición en la actividad de la vía disminuye la tumorigénesis (Topczewska *et al.* 2006).

Dado que también se han realizado estudios en los que se pone de manifiesto el importante papel que desempeña la vía de señalización Nodal en el desarrollo y progresión del cáncer colorrectal, nos centramos en el estudio de *NOMO1*, que funciona como antagonista de esta vía de señalización (Gong *et al.* 2014).

8. PAPEL DE *NOMO1* EN LA CARCINOGENÉISIS DEL CCR

8.1. ASPECTOS GENERALES SOBRE *NOMO1*

Aunque apenas se conoce nada acerca de este gen, se sabe que es uno de los tres genes, *NOMO1*, *NOMO2* y *NOMO3*, prácticamente idénticos que se localizan en una región de duplicaciones dentro del brazo corto del cromosoma 16 (Zhang *et al.* 2015). Se ha propuesto que estos tres genes codifican tres proteínas estrechamente relacionadas entre sí, con un tamaño idéntico y con una homología en su región codificante de aproximadamente un 99,6% y que, podrían llegar a desarrollar la misma función (Martin *et al.* 2004).

La proteína codificada por *NOMO1* es una proteína transmembrana altamente conservada, conocida también como pM5, con un tamaño de 130 kDa (Haffner *et al.* 2004).

Dicha proteína está formada por un total de 1.222 aminoácidos repartidos en tres dominios: un dominio extracelular que comprende del aminoácido 32 al 1.155, un dominio transmembrana helicoidal que comprende los siguientes 21 aminoácidos y, finalmente, un dominio citoplasmático que abarca desde el aminoácido 1.177 hasta el 1.222 (Gene Cards; <http://www.genecards.org/>). Estudios recientes revelan que este gen regula la expresión de otros genes implicados en el desarrollo óseo y es esencial para el desarrollo del cartílago en el pez cebra. Su ausencia produce condrodisplasia (Cao *et al.* 2018).

Esta proteína ha sido identificada como parte de un complejo proteico que participa en la vía de señalización Nodal, la cual es esencial para la especificación del eje del cuerpo primario, así como para la formación del mesodermo y del endodermo durante el desarrollo embrionario (Haffner *et al.* 2007; Dettmer *et al.* 2010; Shen *et al.* 2007).

8.2. COMPLEJO PROTEÍCO NCLN/NOMO/TEMEM147

Hay evidencias acerca de la existencia de un complejo proteico del que forma parte, entre otras, la proteína NOMO1 que se localiza en la membrana del retículo endoplásmico (RE) (Dettmer *et al.* 2010). Este complejo tiene una estructura similar al que forma la γ -secretasa, el cual es esencial para el procesamiento proteolítico de varias proteínas transmembrana incluida la proteína precursora β -amiloide asociada con la enfermedad de Alzheimer (Salnikov *et al.* 2018). Este complejo proteico formado por la γ -secretasa lo comprenden, además, cuatro proteínas transmembrana incluidas la Presinilina1, el componente con actividad enzimática; la Nicastrina, proteína transmembrana tipo I que parece estar involucrada en la maduración del complejo γ -secretasa (Johnson *et al.* 2017), PEN2 y APH-1 cuyas funciones aún no están muy establecidas (Zhu *et al.* 2017).

Posteriormente, se identificó en el RE una proteína transmembrana con una estructura muy similar a la Nicastrina, denominada Nicalina (NCLN, *Nicastrin like protein*) que también pertenece a la familia de los receptores aminopeptidasa/transferrina (TfR) (Kamat *et al.* 2014). Esta proteína tiene un tamaño de 60 KDa y forma parte de un complejo proteico de membrana de alto peso molecular que, en principio, no parece guardar relación funcional con el complejo de la γ -secretasa aunque tengan una estructura prácticamente idéntica (Haffner *et al.* 2004). Inicialmente, se propuso que este complejo cuyo tamaño oscila entre los 200 y 220 KDa, estaba formado por las proteínas transmembrana NCLN (60KDa) y NOMO1 (130KDa). Lo que no está muy claro es la manera en la que se forma el complejo. Sin embargo, se ha demostrado que la proteína NCLN juega un papel crucial en la formación del mismo, controlando los niveles de la

proteína NOMO1. El bloqueo de NCLN en diferentes líneas celulares supone un descenso en la expresión de NOMO1. De manera similar, la expresión de NCLN se ve reducida en células con NOMO1 bloqueado lo cual, sugiere que ambas proteínas son inestables cuando no están unidas entre sí, por lo que la formación del complejo tiene un efecto estabilizador, al igual que ocurre con los componentes del complejo γ -secretasa (Dettmer *et al.* 2010). También se ha observado que un aumento en los niveles de expresión de NCLN resulta en un incremento en los niveles de expresión de NOMO1, sin embargo, la sobreexpresión de NOMO1 no afecta a los niveles de expresión de NCLN (Thinakaran *et al.* 1997). Por lo que se puede pensar, que la proteína NCLN estabiliza a NOMO1 al formar el complejo siendo NCLN el factor limitante en la formación del complejo proteico (Haffner *et al.* 2007).

Teniendo en cuenta la discrepancia que existía entre el tamaño molecular del complejo NCLN/NOMO1 (200-220KDa) y el tamaño molecular de cada proteína monomérica, NCLN 60KDa y NOMO1 130KDa, se pensó en la existencia de un tercer componente. Poco después se confirmó la presencia de la proteína transmembrana TMEM147 con un tamaño molecular de 22KDa. Al igual que ocurre con NCLN y con NOMO1, esta proteína está altamente conservada en metazoos, se localiza en el RE y su distribución en tejidos humanos es idéntica a la distribución de los otros dos componentes del complejo (Haffner *et al.* 2007). Además, se ha demostrado que NCLN no sólo controla los niveles de expresión de NOMO1, sino que también controla los niveles de TMEM147, lo que confirma su papel como factor limitante en la formación del complejo (Dettmer *et al.* 2010).

Diferentes estudios confirman que la formación del complejo ocurre de manera secuencial. En una etapa inicial, NCLN interacciona con NOMO1 formando un subconjunto que se puede aislar bajo ciertas condiciones para, posteriormente, interactuar con TMEM147. Los diferentes dominios de la NCLN median las interacciones con las dos proteínas que se le unen. La unión con NOMO1 se establece a través del dominio luminal y las interacciones con TMEM147 a través del dominio transmembrana (Dettmer *et al.* 2010).

Este complejo proteico NCLN/ NOMO1/TMEM147 interfiere en las primeras fases del desarrollo embrionario y tiene un efecto antagonista en la vía de señalización Nodal (Haffne *et al.* 2004).

9. VÍA DE SEÑALIZACIÓN NODAL

9.1. ASPECTOS GENERALES DE LA VÍA NODAL

La vía de señalización Nodal desempeña un papel crucial en las primeras etapas del desarrollo embrionario, durante la formación del mesodermo y del endodermo, así como en la especificación de la asimetría izquierda y derecha. Esta vía también está implicada en el mantenimiento de la pluripotencia y la capacidad de renovación de las células *stem* embrionarias y juega un importante papel en carcinogénesis (Duan *et al.* 2015; Shen *et al.* 2007).

Los ligandos de la vía Nodal, conocidos como ligandos nodales, los cuales no están muy estudiados, son miembros de la súper familia del factor de crecimiento TGF- β que se unen a receptores serina/treonina quinasa tipo I y tipo II y activan la vía de los SMADs. Los receptores tipo I (*ALK4/ALK7*) y tipo II (*ActRII* o *ActRIIB*) activados fosforilan a las proteínas SMAD2 y/o SMAD3 citoplasmáticas que se activan e interaccionan con SMAD4 dando lugar a la formación de un complejo transcripcional que migra hacia el núcleo donde se asocia con otros reguladores transcripcionales como FOXH1, TP53 y MIXER para activar la transcripción de determinados genes (Figura 10) (Reissmann *et al.* 2001).

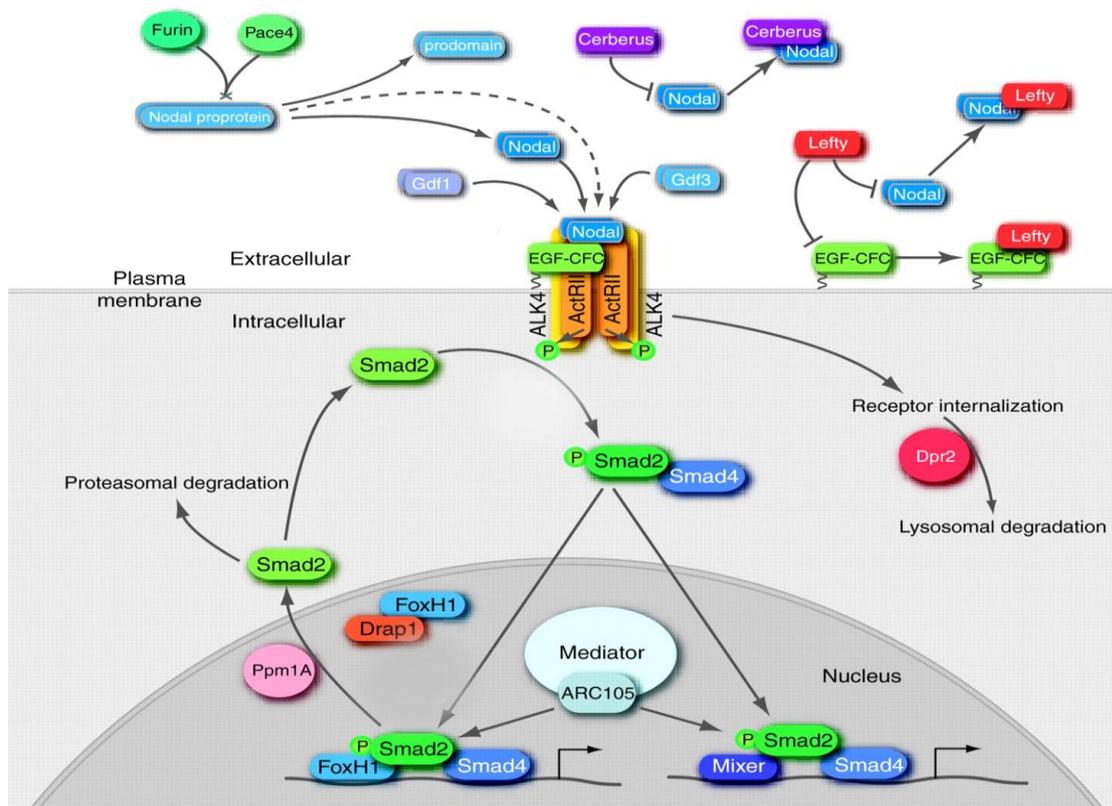


Figura 10. Esquema de la vía de señalización Nodal (Shen *et al.* 2007).

La vía de señalización Nodal necesita proteínas que actúan como correceptores. Se trata de pequeñas proteínas pertenecientes a la familia EGF-CFC, que se localizan en regiones extracelulares ricas en cisteínas y permanecen ancladas a la membrana plasmática a través de uniones glicosil-fosfatidilinositol (GPI), entre las que se encuentra CRIPTO (codificada por el gen *TDGF1*) (Shen *et al.* 2007; Bianco *et al.* 2003). Estas proteínas confieren especificidad a los ligandos nodales para su unión al complejo receptor ALK4/ActRII a través de interacciones proteicas. En ausencia de los correceptores EGF-CFC, los ligandos nodales no son capaces de activar la vía a pesar de mantener su capacidad para interactuar con el complejo receptor ALK4/ActRII (Dettmer *et al.* 2010). Se sabe que, además de participar en la activación de la vía Nodal, CRIPTO también puede activar otras vías implicadas en carcinogénesis, como la vía RAS/RAF/MAPK y la vía PI3K/AKT (Bianco *et al.* 2003).

Se conocen diversos inhibidores de la vía de señalización Nodal. Las proteínas LEFTY son proteínas extracelulares que antagonizan la señalización Nodal a través de interacciones con los correceptores EGF-CFC o con los ligandos nodales, que impiden la formación del complejo receptor. De este modo, las proteínas LEFTY no actúan como inhibidores competitivos de los receptores Alk4/ActRII (Shen *et al.* 2007; Bianco *et al.* 2003). Aunque las proteínas LEFTY son los inhibidores extracelulares de la vía Nodal más estudiados, existen otros como, por ejemplo, miembros de la familia de proteínas DAN/Cerberus que secuestran los ligandos nodales (Haffner *et al.* 2004).

Además de los inhibidores extracelulares mencionados anteriormente, existen otros reguladores de la vía Nodal, como la proteína transmembrana Tomorregulina1, que se une directamente al correceptor EGF-CFC (Harms *et al.* 2003). También, la proteína DRAP la cual se une a FOXH1 y previene su unión al DNA (Iratni *et al.* 2002). Y, por último, el complejo proteico NCLN/NOMO1/TMEM147, que se ha demostrado que actúa como antagonistas de la vía (Shen *et al.* 2007). Aunque el papel de este complejo en la vía Nodal no está muy estudiado, la ganancia o pérdida de función del complejo NCLN/NOMO1/TMEM147 da lugar a fenotipos similares a los descritos para mutantes de la vía de señalización Nodal. La expresión ectópica de cualquiera de estas tres proteínas, pero no de cada proteína por separado, causa ciclopía. Además, el bloqueo de la función de NOMO1 aumenta algunas de las actividades en las que interviene la vía Nodal y, al bloquear esta proteína, se invierte el efecto de los inhibidores nodales LEFTY. Por ello, se ha propuesto que el complejo NCLN/NOMO1/TMEM147 funciona como modulador de la vía Nodal actuando directamente sobre algunos integrantes de la vía, aunque, de momento, no se ha estudiado sobre cuáles actúa (Figura 11) (Dettmer *et al.* 2010).

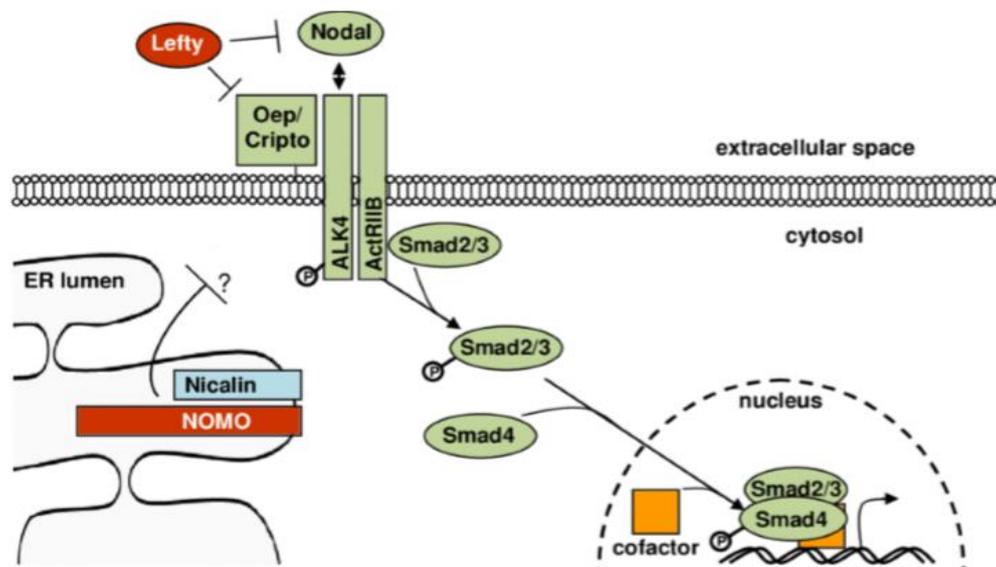


Figura 11. Representación esquemática de la vía Nodal y del posible papel del complejo Nicalina/Nomo en la regulación de esta (Dettmer *et al.* 2010).

9.2. PAPEL DE LA VÍA NODAL EN CARCINOGENESIS

Estudios recientes revelan la importancia de la vía Nodal en el mantenimiento de células madre (*stem*) embrionarias y su papel en carcinogénesis. Aunque la mayoría de los genes que intervienen en la vía Nodal se expresan raramente en adultos, hay evidencias de una elevada actividad de esta vía en determinados tipos de tumores (Quail *et al.* 2013).

Así, se ha descrito un incremento de los niveles de expresión de las proteínas que actúan como ligandos de la vía Nodal en el melanoma maligno, correlacionándose con la progresión del tumor; mientras que la inhibición en la actividad de la vía Nodal disminuye la tumorigénesis (Topczewska *et al.* 2006). Esta vía está implicada también en la vascularización del cáncer de mama y se ha relacionado con la tumorigénesis y metástasis en diferentes tipos de tumores como melanoma, cáncer de mama, colon, ovario, próstata, endometrio y páncreas. Además, se ha observado que las proteínas nodales promueven el desarrollo y la progresión de las *cancer stem cells* (CSCs) tumorales. Por todo esto, se ha pensado que algunos componentes de esta vía podrían tener un importante papel como posibles dianas terapéuticas en determinados tipos de tumores (Mancino *et al.* 2008; Yuehua *et al.* 2014).

La proteína de la vía nodal más estudiada en cáncer hasta el momento es CRIPTO, miembro de la familia de proteínas *TGF-CFC*, que actúa como correceptor esencial de la señalización Nodal al permitir que los ligandos nodales transmitan la señal a través del complejo receptor ALK4/ActRII (Yuehua *et al.* 2014). Esta proteína realiza funciones de regulación

relacionadas con el desarrollo embrionario y es necesaria para mantener la pluripotencia y la capacidad de autorrenovación de las células madre embrionarias. En el adulto, CRIPTO se expresa de manera anómala en una amplia variedad de tumores epiteliales como colon, mama, ovario, gástrico, pulmón, etc. (Bianco *et al.* 2003). La elevada frecuencia con la que aparece CRIPTO expresado en diferentes tipos de carcinomas y su baja expresión en tejido normal, junto con la observación de que la disminución de su expresión inhibe el crecimiento tumoral, sugieren que esta proteína podría ser una posible diana terapéutica contra el cáncer (Sandomenico *et al.* 2018).

9.3. PAPEL DE LA VÍA NODAL EN EL CARCINOMA COLORRECTAL

En diferentes líneas celulares de cáncer de colon (SW480, LOVO y HCT116) y en tejido tumoral humano se ha observado una sobreexpresión de ligandos nodales y de sus receptores, comprobándose que los niveles de expresión de las proteínas Nodales son mayores en tejido tumoral que en el tejido adyacente no tumoral. Esto sugiere que la sobreexpresión de determinados integrantes de la vía Nodal se relaciona con cáncer de colon (Yuehua *et al.* 2014). Los mecanismos por los cuales la vía de señalización Nodal puede facilitar la progresión del cáncer de colon no están muy estudiados, pero existen ensayos en ratones transgénicos que sugieren que Cripto1 podría inducir una transición epitelio-mesénquima (*EMT*). Sin embargo, hasta el momento no parece estar muy claro si los efectos oncogénicos de CRIPTO son dependientes de la actividad de la vía Nodal (Strizzi *et al.* 2004).

En el caso del carcinoma de colon, se ha observado un incremento en la expresión de CRIPTO en lesiones premalignas como en adenomas de colon, metaplasia intestinal o metaplasia de la mucosa gástrica. También se ha detectado en mucosa normal de colon en individuos con un aumento de riesgo, por historia familiar, de desarrollar CCR (Saeki *et al.* 1994). Además, la expresión de CRIPTO aumenta gradualmente conforme evoluciona el tumor: desde la aparición de las primeras lesiones premalignas hasta la formación del carcinoma. La disminución en los niveles de expresión de CRIPTO en líneas celulares de cáncer de colon, reduce drásticamente su crecimiento, lo que sugiere un papel oncogénico de esta proteína (Giorgio *et al.* 2014; Francescangeli *et al.* 2015).

CRIPTO tiene múltiples mecanismos de señalización que pueden contribuir al crecimiento del tumor. Una de las proteínas con las que actúa CRIPTO es GRP78, un regulador multifuncional de la homeostasis del RE, que también se ha implicado en gran medida en cáncer. Curiosamente, aunque GRP78 se localiza en el RE, también se expresa selectivamente en la

membrana plasmática de las células tumorales (Shani *et al.* 2007). Esta proteína forma un complejo con CRIPTO y sus interacciones son esenciales para la función de CRIPTO en las vías de señalización en las que está implicada. Como GPR78 es una chaperona involucrada en diferentes vías de señalización, una desregulación de su expresión podría tener un efecto mayor en el fenotipo del tumor que una reducción en los niveles de expresión de CRIPTO. En el caso del CCR, se ha observado que la expresión de la proteína GRP78 varía notablemente entre tejido normal de colon y tejido tumoral (Shani *et al.* 2007). Es decir, el efecto de CRIPTO en cáncer es complejo y depende fundamentalmente del contexto celular (Giorgio *et al.* 2014).

Por todo ello, consideramos que la exploración a nivel molecular de la vía Nodal y sus reguladores, como el complejo NCLN/NOMO/TMEM147, en tumores colorrectales y en otro tipo de tumores, pueden aportar información relevante en la búsqueda de nuevos marcadores tumorales y posibles nuevas dianas terapéuticas.

HIPÓTESIS Y OBJETIVOS



HIPÓTESIS

El cáncer colorrectal (CCR) es un tumor con una alta incidencia y que causa un elevado número de muertes en todo el mundo, siendo la edad y la interacción ambiente-gen los principales factores de riesgo. Surge como consecuencia de la acumulación de numerosos eventos moleculares que dan lugar a una proliferación descontrolada de las células del epitelio colónico.

Se asume que el CCR es una enfermedad del adulto mayor, con un 85-90% de casos diagnosticados después de los 55 años (LOCRC); no obstante, existe aproximadamente un 8-10% de pacientes diagnosticados antes de los 45 años (EOCRC). El diagnóstico de la enfermedad en individuos jóvenes se relaciona con fenotipos más agresivos del tumor y con peores resultados clínicos. Estas diferencias clínicas, junto con las diferentes alteraciones moleculares encontradas en ambos subgrupos de tumores, sugieren que pueden tener un perfil molecular diferente y, por ello, se ha planteado la hipótesis de que incluso puedan llegar a tener diferentes mecanismos moleculares de carcinogénesis.

Existen numerosos estudios en los que se han correlacionado las características clínico-patológicas y el perfil molecular en los tumores de aparición tardía, sin embargo, esta relación no se ha investigado completamente en los tumores EOCRC, posiblemente debido a la baja frecuencia de CCR en personas jóvenes.

Con el fin de establecer diferencias a nivel molecular entre ambos subgrupos, en un trabajo previo (Arriba *et al.* 2015) se definieron una serie de alteraciones cromosómicas recurrentes características. Se observó que, en la región 16p13.12-p13.11, existía una deleción recurrente que era más frecuente en el grupo EOCRC que en el LOCRC. En esta región cromosómica se localiza el gen *NOMO1*. Por ello postulamos que la región 16p13.12-p13.11 y el gen *NOMO1* podrían ser marcadores del EOCRC.

OBJETIVOS

Teniendo en cuenta todos estos antecedentes, en este trabajo nos propusimos los siguientes objetivos:

1.-Analizar el estado del gen *NOMO1* mediante qPCR en las muestras que presentan la deleción de la región 16p13.12-p13.11.

1.1- Estudiar si la deleción de *NOMO1* es germinal o somática.

1.2- Confirmar la deleción de este gen estudiando nuevos casos.

1.3- Ampliar el estudio de la delección de *NOMO1* a otros tipos de tumores en pacientes menores de 50 años con el fin de asociar esta alteración al EOCRC.

2.-Estudiar el papel de *NOMO* en la vía Nodal. Para ello:

2.1.-Generar una línea *knockout* para *NOMO* mediante el sistema de edición génica CRISPR/Cas9.

2.2- Establecer diferencias fenotípicas entre los clones *knockout* para *NOMO1* obtenidos y los clones utilizados como controles.

2.3.-Estudiar el efecto que la supresión de *NOMO* tiene sobre la expresión de los diferentes integrantes de la vía Nodal, así como en la expresión de las proteínas con las que forma un complejo proteico inhibidor de esta vía de señalización.

3.-Analizar el papel de *NOMO* en otras vías de señalización mediante un estudio del transcriptoma.

4.-Estudiar mediante ultrasecuenciación (NGS) los genes implicados en la vía Nodal y sus inhibidores en tumores de colon con y sin la delección de *NOMO1*.

PACIENTES Y MÉTODOS



1. PACIENTES

1.1. PROCEDENCIA Y CLASIFICACIÓN

Inicialmente, en el estudio de CGH arrays (aCGH) (Arriba et al. 2015), se analizaron muestras de tejido tumoral en bloques FFPE (formalin-fixed paraffin-embedded) de 146 pacientes procedentes del Hospital Universitario 12 de Octubre (Madrid), todos ellos diagnosticados y tratados de CCR primario entre enero de 2002 diciembre de 2009. Estos pacientes se clasificaron en dos grupos atendiendo a su edad en el momento de diagnóstico:

-Adultos jóvenes: pacientes diagnosticados a una edad igual o inferior a los 45 años (60 individuos).

-Adultos de edades intermedias: pacientes diagnosticados a una edad comprendida entre los 46 y los 69 años (67 individuos).

-Adultos ancianos: pacientes diagnosticados a una edad igual o superior a los 70 años (86 individuos).

Posteriormente, con el fin de confirmar los resultados obtenidos en este trabajo, se aumentó la serie de pacientes con 91 muestras tumorales: 41 procedentes de pacientes de diagnóstico temprano y 50 pertenecientes a pacientes diagnosticados a edades avanzadas.

En todos los casos, la cantidad de células tumores de cada muestra de tejido tumoral en bloques FFPE era superior al 70%.

Además de muestras de tumores de colon, en este trabajo, también estudiamos 16 pólipos adenomatosos, todos ellos procedentes de pacientes menores de 45 años. De estos 16 pólipos adenomatosos, 10 se conservaban en bloques FFPE y, los otros 6, los obtuvimos en tejido fresco.

En este trabajo, también se estudiaron otros tipos de tumores: 10 de endometrio, 22 tumores de mama, 9 de ovario, 15 glioblastomas y 13 tumores de cabeza y cuello todos ellos procedentes de pacientes con edad igual o inferior a 50 años. Estos tumores se conservaban en bloques FFPE.

Todos estos tumores proceden de tres instituciones diferentes: Hospital Universitario de Salamanca, Hospital Nuestra Señora de Sonsoles (Complejo Asistencial de Ávila) y Hospital Universitario 12 de Octubre (Madrid). Estas muestras fueron recogidas entre enero de 2002 y noviembre de 2018. Además del tejido tumoral, también se recogió sangre periférica de 13 de los pacientes. En este caso, las muestras se obtuvieron mediante venopunción antecubital, se recogieron en tubos de 10mL con EDTA y se conservaron a 4°C hasta su procesamiento.

Todas las muestras fueron obtenidas previo consentimiento informado, siguiendo las normas legales para estudios clínicos en España y las del comité de ética del Hospital Universitario 12 de Octubre, del Hospital Universitario de Salamanca y del Hospital Nuestra Señora de Sonsoles.

1.2. RECOGIDA DE INFORMACIÓN

Únicamente obtuvimos datos clínicos y datos clínico-patológicos de los 146 pacientes iniciales incluidos en el estudio de aCGH (Tabla 6).

-Edad en el momento de diagnóstico: para el grupo de adultos jóvenes la edad media de diagnóstico se estableció en $39,17 \pm 5,04$ años, todos ellos incluidos en un rango de 16 a 45 años, con máximo pico de incidencia a los 45 años. En el caso del grupo de adultos ancianos, la edad media de diagnóstico fue de $78,07 \pm 5,69$ años, todos ellos incluidos en un rango de 70 a 92 años, siendo en este caso el máximo pico de incidencia a los 70 años. Y, en el caso de los adultos de edades intermedias, la edad media de diagnóstico fue de $66,03 \pm 5,13$ años, todos ellos incluidos en un rango de 51 a 70 años, con máximo pico de incidencia a los 65 años.

De las muestras incluidas posteriormente, no se conocen las edades exactas del diagnóstico, únicamente se pudo saber si eran mayores de 70 años o menores de 45 años.

-Sexo: en el grupo inicial, contamos con un total de 81 muestras procedentes de varones, de las cuales 36 pertenecían a pacientes menores de 45 años y 45 a mayores de 70 años. En este mismo grupo, se contabilizó un total de 65 muestras de mujeres, 24 procedentes de menores de 45 años y 41 mayores de 70 años. Con respecto al grupo de muestras con las que se amplió el estudio, no se obtuvieron datos acerca del sexo de los pacientes.

-Localización tumoral: tanto en el grupo de los tumores diagnosticados a edades tempranas como en el grupo de ancianos, predominó el cáncer de colon frente al de recto (73,3% vs. 26,7% en jóvenes y 61,6% vs. 38,4% en el grupo de los ancianos) (Tabla 6). Además, en el grupo de jóvenes predominó la localización izquierda frente a la localización derecha (53,3% vs. 20%). Sin embargo, en la población adulta, predominaba la localización derecha, aunque este, no es tan acusado como en la población joven (26,7% vs. 34,9%).

-Grado de diferenciación tumoral y producción de moco: no hubo diferencias significativas en el grado de diferenciación tumoral ni en la frecuencia de neoplasias mucosecretoras o con células en “anillo de sello”.

-Estadificación tumoral: como se puede observar en la tabla 6, la distribución de los estadios tumorales fue más homogénea en la población joven que en la anciana. En la población joven fue más frecuente la detección de la enfermedad en las etapas iniciales (estadio I y II) y en la población anciana el estadio más común en el momento del diagnóstico fue el estadio II.

-Historia familiar: en el caso de la población joven, más del 66% de los pacientes diagnosticados con CCR tenían historia familiar de cáncer. Sin embargo, en los pacientes diagnosticados a edades superiores a los 70 años, menos del 25% presentaban historia familiar.

Tabla 6. Resumen de las características clínicas, anatomopatológicas y familiares más importantes de los pacientes incluidos en el estudio de aCGH. SD: desviación estándar.

	Pacientes menores de 45 años n (%)	Pacientes mayores de 70 años n (%)
Nº de pacientes	60 (100)	86 (100)
Media edad de diagnóstico (SD)	39,17 (5,04)	78,07 (5,69)
Sexo:		
hombre	36 (60)	45 (52,3)
mujer	24 (40)	41 (47,7)
Localización:		
colon derecho	12 (20,0)	30 (34,9)
colon izquierdo	32 (53,3)	23 (26,7)
recto	16 (26,7)	33 (38,4)
Diferenciación tumoral:		
pobre	4/49 (8,2)	5/71 (7,0)
Producción moco	14/49(28,6)	17/71 (23,9)
Células "anillo de sello"	2/49 (4,1)	2/71 (2,8)
Estadio TNM:		
I	17(28,3)	13 (15,1)
II	21 (35,0)	35 (40,7)
III	10 (16,7)	20 (23,3)
IV	12 (20,0)	18 (20,9)
Historia familiar de cáncer	40 (66,7)	21 (24,4)
Casos esporádicos	20 (33,3)	65 (75,6)

2. MANIPULACIÓN DE ÁCIDOS NUCLEICOS

La extracción de DNA se llevó a cabo en muestras de sangre periférica, en muestras de tejido tumoral fresco, tejido tumoral en parafina y en líneas celulares.

2.1. EXTRACCIÓN DE DNA A PARTIR DE SANGRE PERIFÉRICA

Para la extracción de DNA a partir de sangre periférica, se aisló la interfase de células nucleadas mediante centrifugación repetida y lisis osmótica de los eritrocitos con H₂O estéril. El botón celular, que contiene únicamente las células nucleadas, se resuspendió en tampón Fornace (0.25M sacarosa, 50mM Tris-HCl pH 7.5, 25mM KCl, 5mM MgCl₂) y se centrifugó a 15000 rpm durante 10 minutos. El botón celular obtenido se resuspendió nuevamente en tampón Fornace, al que se le añadió EDTA 10mM pH 8.0 para posibilitar la inactivación de las nucleasas, 50µg/mL de proteinasa K (ApliChem) para degradar las proteínas y SDS (dodecilsulfato sódico) a una concentración final del 1% para romper las membranas celulares. La mezcla se incubó a 55°C durante 8-16 horas. Tras ello se procedió a la extracción y purificación del DNA tratándolo con una mezcla de fenol tamponado a pH 8.0 y CIAA (cloroformo: alcohol isoamílico 24:1) y se sometió a centrifugación para separar las fases. El DNA que se encuentra en la fase acuosa se limpió de otros residuos celulares mezclándolo con CIAA y centrifugándolo nuevamente para separar las fases. Tras esto, se añadió etanol absoluto frío a la fase acuosa para precipitar el DNA. Tras un lavado con etanol al 70% para eliminar sales, finalmente se resuspendió en 200µl de H₂O estéril.

2.2. EXTRACCIÓN DE DNA A PARTIR DE TEJIDO TUMORAL FRESCO

En el caso del tejido tumoral fresco, en nuestro caso 7 pólipos adenomatosos, se tomaron muestras de 100-200 mg y se homogeneizaron en tampón Fornace con el homogeneizador *Polytron® System PT 1200 E*, (Thermo Fisher Scientific). Una vez homogeneizado el tejido se añadió EDTA, proteinasa K y SDS siguiendo el mismo procedimiento que para la extracción de DNA a partir de sangre periférica.

2.3. EXTRACCIÓN DE DNA A PARTIR DE TEJIDO TUMORAL PARAFINADO

En el caso del tejido tumoral parafinado, se realizaron 3-4 cortes de 10 µm de los bloques FFPE y se extrajo el DNA de la muestra utilizando el kit QIAamp DNA FFPE Tissue kit (Qiagen, Hilden, Germany). Para ello, en primer lugar, se eliminó la parafina añadiendo a la muestra 1 mL

de xileno. Se dejó la muestra en un baño a 70°C durante 30 minutos y cada 10 minutos se agitó rigurosamente (vórtex) para que se disuelva la parafina. Posteriormente, se centrifugó a 10000 rpm durante 5 minutos y se eliminó el sobrenadante. A continuación, se realizaron tres lavados con etanol para eliminar el xileno.

Una vez que se desparafinó la muestra, se añadieron 180 µl de tampón ATL y 20 µl de proteinasa K. La mezcla se incubó a 55°C durante 8-16 horas. Transcurrido este tiempo, se incubó la muestra a 90°C durante 1 hora. Después, se añadió a la muestra 200 µl de tampón AL y 200 µl de etanol absoluto. Se transfirió esta mezcla a la columna QIAamp MinElute (en un tubo de lavado de 2 ml) sin mojar el borde y se centrifuga a aproximadamente 6000 x g durante ≥1 min. Seguidamente se realizan dos lavados con 500 µl de tampón AW1 y seguidamente con 500 µl de tampón AW2. La elución final de la muestra se realizó en 40µl de H₂O estéril.

2.4. EXTRACCIÓN DE DNA A PARTIR DE LÍNEAS CELULARES

En el caso de las líneas celulares, las células se recogieron de las placas de cultivo usando 1.5 mL del agente de disociación TrypLE 1X (Gibco), se lavaron con PBS y se centrifugaron a 1500 rpm tres minutos. El precipitado celular se resuspendió en tampón Fornace, EDTA, proteinasa K y SDS, siguiendo el mismo procedimiento que para la extracción de DNA a partir de sangre periférica.

2.5. EXTRACCIÓN DE RNA A PARTIR DE LÍNEAS CELULARES

Para realizar la extracción del RNA, las células se colocaron en hielo y se lavaron con PBS frío. Se recogieron en 500 µl de TRizol (Life Technologies), por placa de 60 mm (para placa 100 mm, las células se recogen en 1ml de TRizol). La solución se transfirió a tubos nuevos, y a continuación se añadió 100 µl de cloroformo, se agitó en el vórtex durante 15 segundos y se dejó a temperatura ambiente durante 10-15 minutos. Posteriormente, se centrifugó a 12.000 g a 4°C durante 15 minutos. Se observaron tres fases diferenciadas: la fase inferior, que contiene proteínas; la fase intermedia, que contiene DNA, y la fase acuosa superior, que contiene RNA. Se trasladó la fase acuosa superior a otro tubo, y se procedió a la precipitación del RNA añadiendo 250 µl de isopropanol; se mezcló y se dejó a temperatura ambiente entre 5-10 minutos, tras lo cual se centrifugó a 12.000g a 4 °C durante 10 minutos. Después, se aspiró el sobrenadante, y se observó el pellet de RNA en el fondo del tubo, que se lavó con 1 ml de etanol al 75%. Se centrifugó a 12.000 g a 4°C durante 10 minutos, se eliminó el sobrenadante, y

se dejó secar al aire durante 5-10 minutos, evitando que el pellet se seque completamente. Se resuspendió el pellet en 40 µl de agua libre de RNAsas.

3. CUANTIFICACIÓN DEL DNA EXTRAÍDO

La concentración de DNA de las muestras de sangre periférica se determinó en un espectrofotómetro *NanoDrop*[®] 2000 (Thermo Fisher Scientific) midiendo la absorbancia de la muestra a 260 nm. Además, el cociente entre las absorbancias A260/A280 permitió determinar el grado de pureza de las muestras, estando los valores adecuados comprendidos entre 1.8-2.0. Si en algún caso la relación se salía de ese margen se procedió a realizar una nueva purificación del DNA, ya que esto indicaría la presencia de proteínas o solventes orgánicos si los valores son inferiores, mientras que valores superiores indicarían un exceso de RNA.

La concentración de DNA extraído de los tejidos tumorales parafinados se midió con un fluorómetro *Qubit*[®] 3.0 (Life Technologies). La concentración de las muestras procede de la señal que emite una molécula fluorescente sólo cuando se encuentra unida al DNA, minimizando el efecto de contaminantes o de DNA o RNA degradado y obteniendo medidas más fiables al ser un método más específico que las mediciones de absorbancia en UV²⁶⁰.

Todas las muestras de DNA se almacenaron en tubos Eppendorf[®] a -20°C, con el fin de evitar tanto la degradación progresiva del DNA como una posible contaminación.

4. AMPLIFICACIÓN DE FRAGMENTOS DE DNA MEDIANTE PCR

La reacción de amplificación PCR (polymerase chain reaction) se llevó a un volumen final de 25 µl utilizando 12.5 µL del sistema PCR Master Mix 2X (Promega) (22mM Tris-HCL a pH 8.4, 55 mM KCl, 1.65 mM de MgCl₂, 220 µM de dNTP, y 22U de Taq DNA polimerasa), 9.5 µl de agua libre de nucleasas, 1 µl de cada uno de los oligonucleótidos (*forward* y *reverse*) a una concentración final de 0.5 µM cada uno y 1µl de DNA molde a concentración de 100 ng/mL. Como control negativo se preparó en paralelo un tubo de reacción con todos los componentes excepto DNA.

Los oligonucleótidos específicos diseñados para la amplificación de la región de interés del gen *NOMO* se muestran en la Tabla 7.

Tabla 7. Oligonucleótidos empleados para la amplificación de la región de interés del gen NOMO.

Oligonucleótido sentido (5'-3')	Oligonucleótido anti-sentido (5'-3')
CAG TGC TCA GTA CCA TGT AG	GGG AGG AAT ACA AAC CCT C

La reacción de amplificación se llevó a cabo en un termociclador (Life Technologies-Invitrogen) utilizando el siguiente programa de amplificación (Figura 12).

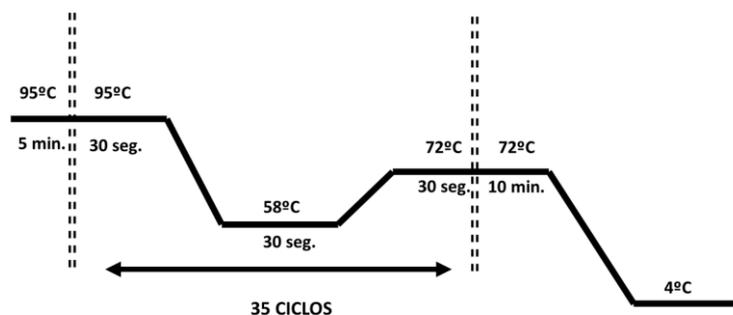


Figura 12. Programa de amplificación estándar utilizado.

5. ELECTROFORESIS EN GEL DE AGAROSA

El tamaño del producto amplificado se analizó mediante electroforesis en geles horizontales de agarosa al 2% en tampón TBE 0.5X (Tris 0.045M, ácido bórico 0.045 M, EDTA 1.0 mM pH 8.0) y teñido con bromuro de etidio. La muestra migró en paralelo con un marcador de tamaño molecular derivado del DNA del fago Φ X-174 cortado con la endonucleasa HaeIII que genera fragmentos de los siguientes tamaños: 1.353, 1.078, 0.872, 0.603, 0.310, 0.281, 0.271, 0.234, 0.194, 0.118, 0.072 kb.

Los DNAs extraídos de las muestras en parafina se cargaron en geles de agarosa al 1% en tampón TBE 0,5X para comprobar cualitativamente su calidad. Estas muestras migraron en paralelo con un marcador de tamaño molecular derivado del DNA del fago lambda Hind III- Φ X-174 cortado con la endonucleasa HaeIII que genera fragmentos de los siguientes tamaños: 22.130, 9.416, 6.557, 4.381, 2.322, 2.207, 1.353, 1.078, 0.872, 0.603, 0.310, 0.281, 0.271, 0.234, 0.194, 0.118, 0.072 kb.

La migración del DNA en el gel se monitorizó utilizando dos colorantes incluidos en el tampón de carga: el xileno cianol, que migra aproximadamente con los fragmentos de 5Kb en un gel de agarosa al 0.8%, y el azul de bromofenol, que migra aproximadamente con los fragmentos de 0.5Kb. La electroforesis se llevó a cabo con un potencial constante de 120 voltios

durante 40 minutos. Los resultados obtenidos fueron capturados gracias al programa Gel Doc 2000 (Biorad).

6. PURIFICACIÓN DE ÁCIDOS NUCLEICOS

6.1. PURIFICACIÓN DE DNA

La purificación de fragmentos de DNA procedentes de la amplificación fue realizada mediante PureLink® PCR Purification kit (Life Technologies-Invitrogen, California, U.S.A). Se mezcló el producto de PCR con 4 volúmenes de la solución de *Binding*. Se pasó esta mezcla por la columna y se centrifugó a 10000 rpm durante un minuto. Posteriormente, para limpiar el DNA, se añadió a la columna 500 µL de la solución *wash buffer*. A continuación, se centrifugó a 10000 rpm durante un minuto. Se decanta y se vuelve a centrifugar para secar la columna. Finalmente, el DNA pegado en la membrana se eluyó con 40 µL de *elution buffer*.

6.2. PURIFICACIÓN DE RNA

La purificación de RNA se llevó a cabo empleando el kit GenElute™ Total RNA Purification Kit (Sigma-Aldrich, St. Louis, USA). Se mezcló el RNA extraído con 350 µL de *buffer RL*. Vórtex durante 15 segundos. Seguidamente se añadieron 200 µL de etanol al 96% y de nuevo vórtex durante 15 segundos. Se añadió todo el volumen a la columna y se centrifugó 1 minuto a 3.500g (6.000 rpm) asegurándose de que todo el lisado pasa a través de la columna. Posteriormente se añadieron 400 µL de *Wash Solution A* a la columna y se centrifugó 1 minuto en las mismas condiciones que antes. Finalmente, se pasó la columna a un tubo limpio y se añadieron 50 µL de *Elution Solution A*. Centrifugar 2 minutos a 200g seguido por 1 minuto a 14.000g.

7. SECUENCIACIÓN AUTOMÁTICA

Para la secuenciación automática, las muestras se prepararon con 40-60ng de DNA purificado con 3 pmol del oligonucleótido sentido o antisentido correspondiente, llevado a un volumen final de 8 µL. La reacción de secuenciación por Sanger se llevó a cabo en el secuenciador automático ABI PRISM 377 *DNA Sequencer* (Applied Biosystems, Inc.) en la unidad de Genómica del Centro de Investigación del Cáncer (CSIC-Universidad de Salamanca), usando el kit *BigDye Terminator*® v.3.1.

La lectura y tratamiento de las secuencias automáticas se realizó con ayuda del programa *Chromas lite 2.01* (Figura 13). La determinación del grado de homología con las secuencias de referencia se llevó a cabo con el programa Serial Cloner.

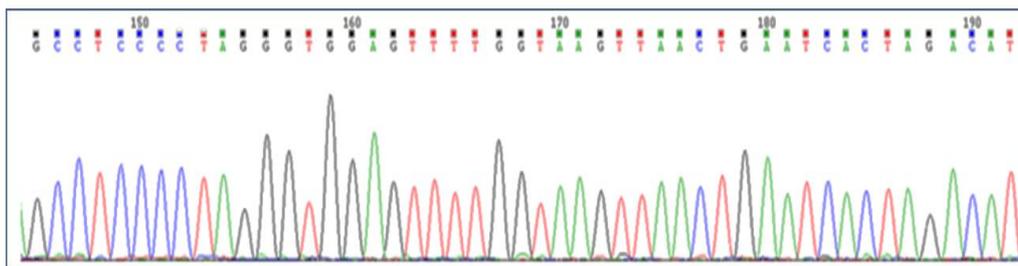


Figura 13. Secuencia de nucleótidos correspondiente a un fragmento del gen *NOMO*.

8. HIBRIDACIÓN *IN SITU* POR FLUORESCENCIA (FISH)

Para confirmar las ganancias y pérdidas de material genético detectadas por la técnica de aCGH, se llevó a cabo un análisis de hibridación *in situ* por fluorescencia. La técnica FISH, es una técnica citogenética de marcaje de cromosomas mediante la cual, estos son hibridados con sondas que emiten fluorescencia y permiten la visualización, distinción y estudio de los cromosomas así como de las anomalías que puedan presentar. Para este trabajo se utilizaron las sondas BAC y 354N7 que hibrida con la región 16q22.1 (bases 68.727.161-68.887.391) y CTD 2504F3 asignado a la región 16p13.1 (bases 15.982.491- 16.190.907) (González MB *et al.* 2004). Estas sondas fueron elegidas de la librería del Instituto “*Wellcome Trust Sanger*” (Cambridge, U.K). Todas estas sondas fueron hibridadas sobre metafases de individuos sanos con el fin de verificar su localización. Las hibridaciones se realizaron sobre portas incluidos en parafina. La FISH de las muestras en parafina se realizó sobre cortes de 4-5 μm de espesor. El grosor del corte es importante ya que tanto si es más fino como si es más grueso la FISH no será analizable. Si el tamaño de los cortes supera los 5 μm de espesor la desparafinización del tejido es más difícil, dando una hibridación ineficiente. De igual modo, si el tamaño es inferior a 4 μm los núcleos pueden aparecer incompletos produciéndose falsos positivos. Además, conviene eliminar las áreas en las que los núcleos están solapados o las zonas en las que el tejido está parcialmente aplastado.

Por tanto, teniendo en cuenta estas recomendaciones y basándonos en los resultados obtenidos al hibridar estas sondas en los controles normales, el punto de corte para la identificación de las alteraciones se delimitó en el límite >12% (3 señales del BAC para ganancia y 1 señal para pérdida).

El procedimiento a seguir para llevar a cabo esta técnica se compuso de 4 fases:

1-Marcaje directo de sondas no comerciales mediante *Nick Translation*

Para el marcaje directo de sondas, necesitamos preparar las siguientes soluciones:

- 0.2 mM *Spectrum Green* dUTP // 0.2 mM *Spectrum Orange* dUTP: 5 μ l 1 mM *Spectrum Green* dUTP // *Spectrum Orange* dUTP + 20 μ l H₂O esteril.

- 0.1 mM dTTP: 10 μ l 0.3 mM dTTP + 20 μ l H₂O esteril.

- 0.1 mM dNTP mix: 10 μ l 0.3mM dATP + 10 μ l 0.3mM dCTP + 10 μ l 0.3 mM dGTP.

Las cantidades de los reactivos utilizados están calculadas para realizar el marcaje de 1-3 μ g de DNA con lo que se obtiene un total de 50 μ l de mezcla de la reacción. Todo el proceso debe hacerse en frío: DNA [1 μ g-3 μ g] X μ l, 10xNT 5 μ l, dNTP 10 μ l, dTTP 5 μ l, *Spectrum Green/Orange* 2,5 μ l, H₂O esteril (ajustar hasta 50 μ l), enzima 5 μ l. Incubar 45 minutos a 15^oC y comprobar el tamaño de la sonda en un gel de agarosa al 1%. Si el tamaño de la sonda es el deseado, en torno a 300 pb, se inactiva la enzima incubando en el baño durante 10 minutos a 70^oC. Como resultado se obtiene la sonda marcada de forma directa a una concentración aproximada de 20 ng/ μ l. A continuación, se puede utilizar en la hibridación o se puede almacenar a - 20^oC.

2- Pretratamiento de portaobjetos

En primer lugar, hay que colocar los portaobjetos en una estufa a 65^oC hasta que se elimine la parafina. Al mismo tiempo, se atempera la solución de pepsina en un baño a 37^oC. Posteriormente, se introducen los portas en xilol durante 10 minutos. Se elimina el xilol y se repite este proceso de lavado durante tres veces más. Después, se realizan dos lavados con etanol absoluto durante 10 minutos. Para eliminar el etanol absoluto realizar dos lavados con etanol al 90% y al 70%, respectivamente, durante 5 minutos cada uno y para eliminar el etanol de 70% se lava con abundante H₂O destilada. Después de secar las preparaciones, se colocan los portaobjetos en un recipiente con tampón Tris-EDTA y se incuban durante 45 minutos en un baño a 99^oC evitando que la solución de Tris-EDTA hierva. Introducir los portas en la solución de pepsina y dejar digerir entre 25 y 45 minutos. El tiempo de digestión dependerá de cómo se haya eliminado la parafina del portaobjetos durante el tiempo que han estado en la estufa a 65^oC. Finalmente, se pasan los portas a la solución de Stop durante 5 minutos, se lavan con H₂O y se dejan secar a temperatura ambiente.

3-Hibridación

Se deposita la sonda en el portaobjetos, se superpone un cubreobjetos, se sella y se introducen los portaobjetos en la estación de hibridación HYbriter (Vysis) con el programa utilizado en nuestro trabajo para la hibridación de sondas en muestras parafinadas. El programa utilizado en este trabajo fue:

Desnaturalización a 83°C durante 7min.

Hibridación a 45°C durante 48 horas.

4- Lavados posthibridación y contratinción

En primer lugar, hay que calentar las soluciones de formamida y 2x SSC en un baño a 46°C. Después hay que eliminar el pegamento de sellado y, si es posible el cubreobjetos. Se realiza un lavado durante 5 minutos con formamida al 50% atemperada a 46°C. Seguidamente, se realiza un lavado durante 3 minutos en 2X SSC también a 46°C. Mantener en DAPI (4,6-diamidino-2-fenilindol) durante 2 minutos y añadir "Vectashield"® para que no se pierda la fluorescencia. Mantener las preparaciones a 4°C y protegidas de la luz.

5-Adquisición y análisis digital de las imágenes

Se realizó utilizando los siguientes equipos:

- Microscopio epifluorescente con rueda de filtros (DAPI, FITC, TRITC).
- Cámara CCD fría.
- Sistema de análisis de imagen Cytovision (Applied Imaging) o ISIS (Metasystem).

9. PCR CUANTITATIVA-COMPARATIVA A TIEMPO REAL (q-PCR)

Esta técnica permite detectar y evaluar la progresión de la reacción de amplificación en cada uno de sus ciclos. Se fundamenta en la adaptación a un termociclador de un lector de fluorescencia, que detecta la fluorescencia emitida por sondas o componentes intercalantes unidos al DNA. La monitorización de esta fluorescencia permite analizar la cinética de la PCR en cada uno de sus ciclos, determinando el número de copias del fragmento de interés al inicio de la reacción. Este valor se puede cuantificar con gran precisión a través del ciclo umbral (Ct).

El Ct es el número de ciclos necesarios para que se produzca un aumento de fluorescencia significativo con respecto a la señal base y es inversamente proporcional a la cantidad inicial de moléculas molde.

La curva de fusión (*melting*) del producto es otro parámetro que sirve en esta técnica para la caracterización del producto amplificado por el valor de temperatura de fusión (T_m). Esta curva se obtiene al finalizar la amplificación y se produce como consecuencia del decaimiento de la señal fluorescente debido a la fusión dependiente de la temperatura del producto. La derivada de la curva de fusión revela un pico máximo que corresponde a esa temperatura (Figura 14).

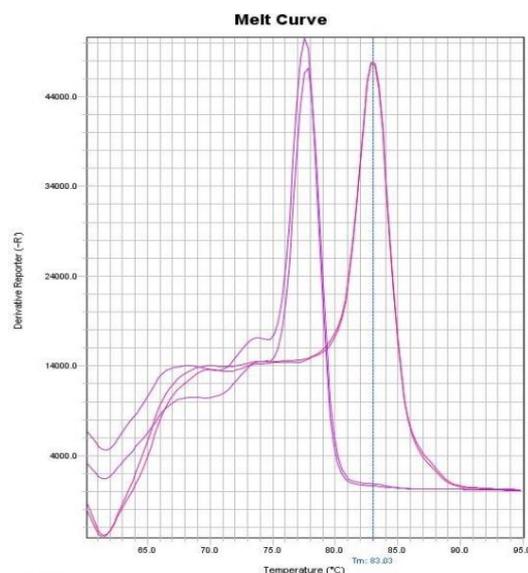


Figura 14. Gráfica de la curva de fusión del producto de amplificación. Se representa en el eje de abscisas la temperatura en grados centígrados y en el eje de ordenadas la fluorescencia en unidades arbitrarias. La T_m corresponde con el punto más alto de fluorescencia del pico de la curva.

Por tanto, el uso de la PCR a tiempo real cuantitativa comparativa, permite cuantificar la cantidad de DNA de un determinado fragmento comparando esta amplificación con otra muestra utilizada como control, realizada simultáneamente y siempre que ambas pertenezcan a la misma muestra. Para analizar estos cambios relativos se elige como control un fragmento de referencia que no varía su número de copias en la patología a estudiar. Comparando el Ct del fragmento de referencia con el Ct del fragmento problema en la muestra de los tumores y en la muestra control, se pueden determinar cambios relativos en la carga genética como pérdidas de heterocigosidad, grandes deleciones o amplificaciones en el gen de interés. De este modo, las cantidades relativas de los dos productos seguirán siendo las mismas si la cantidad de los fragmentos de DNA de partida son iguales; o, por el contrario, si hubiera una discrepancia en estos valores iniciales, la cantidad de los fragmentos de DNA amplificados mostraría diferente cinética y por tanto diferente fluorescencia emitida.

Por estas razones, esta técnica requiere la amplificación de varios fragmentos: el fragmento correspondiente a la secuencia de interés y el fragmento de referencia, que en nuestro caso, corresponde a un fragmento del exón 1 del gen de copia única *LEMD3* (Tabla 8). Además, fue necesario utilizar un control (DNA de una muestra en la que sepamos que no existe LOH) que se amplificó en iguales condiciones que las muestras de los tumores.

Tabla 8. Oligonucleótidos utilizados para la amplificación de las regiones analizadas mediante PCR cuantitativa en los ensayos de pérdida de heterocigosidad (LOH).

Región	Oligonucleótido sentido 5'-3'	Oligonucleótido antisentido 5'-3'
<i>LEMD3</i> (exón 1)	GCGGCTGCCGGGAGTCTAGACAG	GGTAGGTATGATTGGAGCCCG
<i>NOMO1</i> (exón/intrón 4)	AGCTCCATGTGGATGGAGTC	ACGGATGAAGTACAGAGTTC
<i>NOMO1</i> (exón 1)	CCGAGCATGCGCCTTAGTTC	ACCAGCATGTCCTCCGAGCC
<i>RBFOX1</i> (exón 2)	TCTGCATGGTGGCTCCTCAT	AAGGCTGAGCCATTGTGTCA
<i>PKD1</i> (exon 13)	GTGGAGGGAGGGACGCCAA	ACAGGGAAACCGAGGCTCAGAAA
<i>PKD1</i> (exon 30)	CTACAGGTGGGTGCCGTAGG	CGCCTTCCCTCTGGCTGC
<i>STS</i> (exon 5)	TCCTTTACAGGAAGATGAAG	CATTACCAACCTGATAGTTTT

Para lograr obtener cantidades iniciales idénticas de todas las muestras a analizar, los DNAs fueron cuantificados en el Qubit®. A continuación, se realizaron las diluciones pertinentes con la finalidad de dejar todos los DNAs a una concentración inicial de 15 ng/μl.

Las PCRs se realizaron en placas ópticas de 96 pocillos (MicroAmp® Fast Optical 96-Well Reaction Plate with Barcode 0.1 mL (Life Technologies-Invitrogen, California, U.S.A.). En cada placa se llevaron a cabo, por cada muestra, dos reacciones diferentes en pocillos diferentes: una reacción de amplificación del fragmento a estudiar y una reacción de amplificación del fragmento seleccionado del gen *LEMD3*. Se realizaron tres réplicas de cada amplificación para minimizar la variabilidad de concentración de las muestras.

Las placas fueron cubiertas con láminas transparentes ópticas para que se pudiese leer la fluorescencia a través de ellas (MicroAmp® Optical Adhesive Film, Life Technologies-Invitrogen, California, U.S.A.).

El equipo empleado para llevar a cabo la PCR a tiempo real fue StepOnePlus™ Real-Time PCR System (Life Technologies-Invitrogen, California, U.S.A.), compuesto por un termociclador y varios canales de lectura de fluorescencia. El programa utilizado para analizar los resultados fue el StepOne software 2.1 (Life Technologies-Invitrogen, California, U.S.A.).

El agente fluoróforo utilizado fue el SYBR Green incluido en la mezcla comercial SYBR® Green PCR Máster Mix (Roche, Basilea, Suiza). El SYBR Green es un agente intercalante que se une con gran afinidad al surco menor del DNA bicatenario, aumentando su fluorescencia unas 1000 veces.

La amplificación se llevó a cabo en un volumen de 20 μl con las siguientes condiciones: 10 μl de SYBR® Green PCR Máster Mix, 0.7 μl de cada oligonucleótido, 1.0 μl de DNA a una concentración de 15 $\text{ng}/\mu\text{l}$ y 9 μl de H_2O libre de nucleasas. También se realizaron controles negativos con todos los componentes de la reacción sin DNA para asegurar la ausencia de contaminación en las reacciones de amplificación.

El programa de amplificación utilizado para la obtención de las curvas de fusión se expone en la Figura 15.

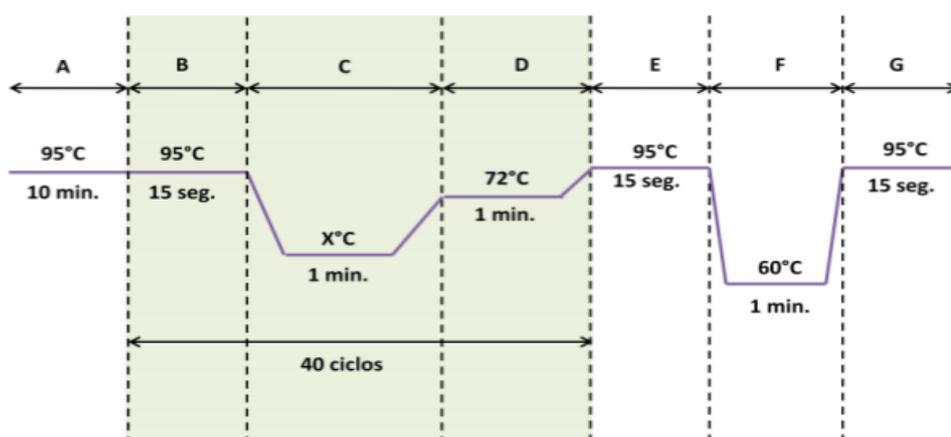


Figura 15. Esquema representativo del programa del termociclador para la cuantificación relativa de fragmentos de DNA mediante PCR a tiempo real cuantitativa comparativa. A es el tiempo de desnaturalización previa, BCD son los tramos de desnaturalización, anillamiento de los oligonucleótidos y elongación que se repiten 40 ciclos, y EFG es el programa para obtener las curvas de fusión (melting) de los fragmentos amplificados. Se incluyen las respectivas temperaturas de cada tramo. La temperatura de anillamiento X en este caso era de 58°C.

10. ESTUDIO DE LA LONGITUD TELOMÉRICA MEDIANTE qPCR

10.1. AMPLIFICACIÓN DE LOS FRAGMENTOS

El análisis de la medida de la longitud telomérica se llevó a cabo en un total de 91 muestras de DNA procedentes de tumores de colon diagnosticado tanto en pacientes jóvenes como en pacientes ancianos. No se pudo realizar el estudio en las 146 muestras que previamente se habían incluido en la técnica de CGH *arrays* por falta de material genómico.

El protocolo seguido fue distinto al descrito anteriormente en este trabajo para la PCR cuantitativa comparativa. En este caso, el gen de referencia utilizado fue el *36b4*. Los oligonucleótidos para la reacción de amplificación de los telómeros se muestran a continuación (Tabla 9).

Tabla 9. Oligonucleótidos utilizados para la amplificación de los telómeros mediante PCR cuantitativa.

Región	Oligonucleótido sentido 5'-3'	Oligonucleótido anti-sentido 5'-3'
36B4	CAGCAAGTGGGAAGGTG AATCC	CCCATTCTATCATCAACGGGTACAA
Tel1	GGCTTGCCTTACCCTTACCCTTACCCTTACCCT TACCCT	CGGTTTGGTTGGGTTGGGTTTGGGTTTGGGTTTGGGT T

La reacción de PCR cuantitativa comparativa relativa se llevó a cabo preparando dos mezclas de reactivos cuya única diferencia fueron los oligonucleótidos y la cantidad añadida de cada uno de ellos. Cada pocillo constó de 5 μL de SYBR[®] Green PCR Máster Mix (Roche, Basilea, Suiza), 0.3 μL de oligonucleótido sentido (10 μM), 0.9 μL ó 0.5 μL de oligonucleótido antisentido (10 μM) para la reacción de amplificación de telómeros o de *36b4* respectivamente 20 ng de DNA (4 μL por 5 ng/ μL) y ddH₂O hasta completar un volumen final de 10 μL .

De cada una de las muestras se realizaron tres réplicas por cada reacción de amplificación, es decir, tres réplicas para la amplificación de telómeros y tres réplicas para la reacción de amplificación del gen *36B4*. Se prepararon las 2 mezclas de reacción en recipientes separados y con una pipeta multicanal se añadieron 20 μL de mezcla de reacción para telómeros a cada pocillo de la mitad de una placa óptica de 96 pocillos y otros 20 μL de mezcla de reacción para el gen *36b4* a la otra mitad de la placa. Un pocillo de cada parte de la placa se utilizó como control negativo al añadirles sólo los reactivos y no contener ningún DNA.

Finalmente, tapamos la placa de 96 pocillos con un adhesivo óptico (*microampm optical adhesive film*), le aplicamos un spin y la dejamos reposar durante una hora a 4^oC con objeto de homogeneizar las mezclas lo máximo posible.

Una vez transcurrido este tiempo se realizó la reacción de amplificación de telómeros y 36B4 en el aparato StepOnePlus™ Real-Time PCR System (Life Technologies-Invitrogen, California, U.S.A.) siguiendo las condiciones mostradas en la Figura 16.

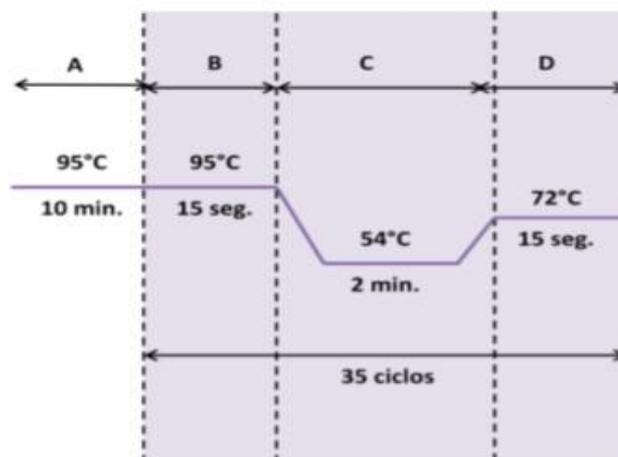


Figura 16. Esquema representativo del programa del termociclador para la amplificación de telómeros mediante PCR cuantitativa. A es el tiempo de desnaturalización previa y BCD son los tramos de desnaturalización, anillamiento de los oligonucleótidos y elongación que se repiten 35 ciclos. Se incluyen las respectivas temperaturas de cada tramo.

Terminado el proceso, se llevó a cabo el análisis de las amplificaciones con el mismo software que nos facilitó las gráficas de amplificación y los valores de las T_m para cada una de las reacciones (Figura 17).

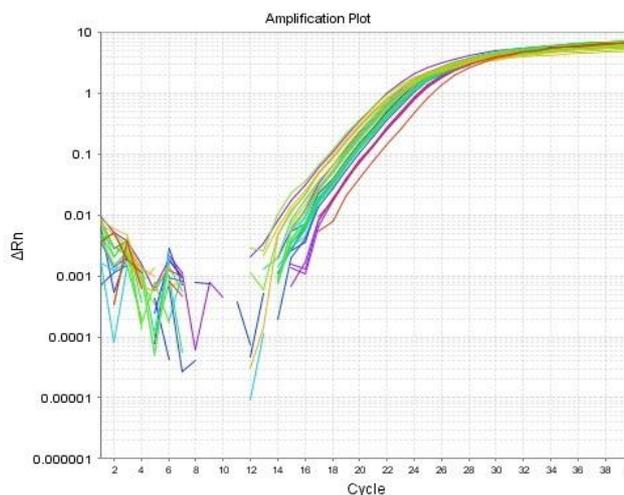


Figura 17. Gráficas resultado de la amplificación de los telómeros.

10.2. ANÁLISIS ESTADÍSTICO

El análisis estadístico de la medida de la longitud telomérica se realizó gracias a los programas StepOne software 2.1 (Life Technologies, California, U.S.A.) y GenEx 5.3.6 Enterprise (MultiD Analyses AB, Göteborg, Suecia).

El programa GenEx 5.3.6 Enterprise permite agrupar un elevado número de resultados obtenidos en diferentes placas de PCR a tiempo real cuantitativa comparativa relativa, elaborando un estudio conjunto previa exportación de los datos desde el software StepOne al software GenEx.

GenEx es un programa adecuado para el análisis de datos de PCR cuantitativa a tiempo real comparativa relativa ya que permite realizar una corrección de las frecuencias de amplificación del fragmento a estudiar y del gen reportero. Además, normaliza las distintas réplicas de una misma muestra y, a su vez, normaliza todos los resultados con el DNA de referencia presente en cada placa. Todo ello teniendo en cuenta la eficiencia de nuestra PCR según los datos obtenidos en la puesta a punto de la técnica.

11. CULTIVOS CELULARES

En este trabajo se seleccionaron dos líneas de cáncer de colon cuyas características se detallan en la tabla 10.

Tabla 10. Líneas celulares utilizadas y resumen de sus mutaciones.

Línea celular	ATCC® nº	Tipo tumoral	Tumorig.	Genes mutados	Mutaciones
HT29	HTB-38™	Adenocarcinoma colorrectal	SI	APC BRAF PI3KCA SMAD4 TP53	c.2557G>T; p.E853* c.4666_4667insA;p.T1556fs*3 c.1799T>A; p.V600E c.1345C>A; p.P449T c.931C>T; p.Q311X c.818G>A; p.R273H
HCT116	CCL-247™	Adenocarcinoma colorrectal	SI	CDKN2A CTNNB1 KRAS MLH1 PI3KCA	c.68_69insG;p.R24fs*20 c.133_135delTCT;p.S45del c.38G>A;p.G13D c.755C>A;p.S252* c.140A>G; p.H1047R

Todas las líneas celulares se cultivaron en un incubador (Cultek S.L.U., Madrid, España) a 37°C en una atmósfera húmeda con un 5% de CO₂. Para el cultivo se utilizó el medio Dubelcco's Modified Eagle Medium (DMEM) (*Sigma-Aldrich*) con glucosa (4.5 g/l), L-glutamina y L-piruvato. El medio de cultivo se suplementó con suero fetal bovino (FBS) al 10% y antibióticos al 1% (penicilina/estreptomina) (*Invitrogen, Life Technologies, California, Estados Unidos*).

12. TECNOLOGÍA CRISPR-CAS9

CRISPR (clustered regularly interspaced short palindromic repeats) es una región del DNA de algunas bacterias que actúa como mecanismo inmunitario frente a los virus. Concretamente ese conjunto de secuencias de DNA tiene la capacidad de reconocer virus invasores en la bacteria. Al hacerlo, despliega sobre ellos un enzima especial que se encarga de trocearlos y utiliza esos fragmentos para inmunizar a la bacteria frente a tales virus. Para que este mecanismo de defensa y ahora también nuevo sistema de edición genómica lleve a cabo los cortes en el DNA, se requieren únicamente dos componentes: la proteína Cas9 y el RNA guía. La enzima Cas9 se caracteriza por tener una arquitectura bilobular en la que uno de los lóbulos (REC) se encarga del reconocimiento de la diana y el otro (NUC) posee la actividad nucleasa (Nishimasu *et al.* 2014). Este último lóbulo se compone de dos dominios nucleasa RuvC y HNH, y de un dominio de reconocimiento de la secuencia PAM (dominio PI). La otra mitad del sistema CRISPR-Cas9 de diseño consiste en un componente de RNA dúplex, formado por el crRNA y la molécula de crRNA transactivador o tracrRNA, denominado de forma general al conjunto como RNA guía (Figura 18) (Cong *et al.* 2015).

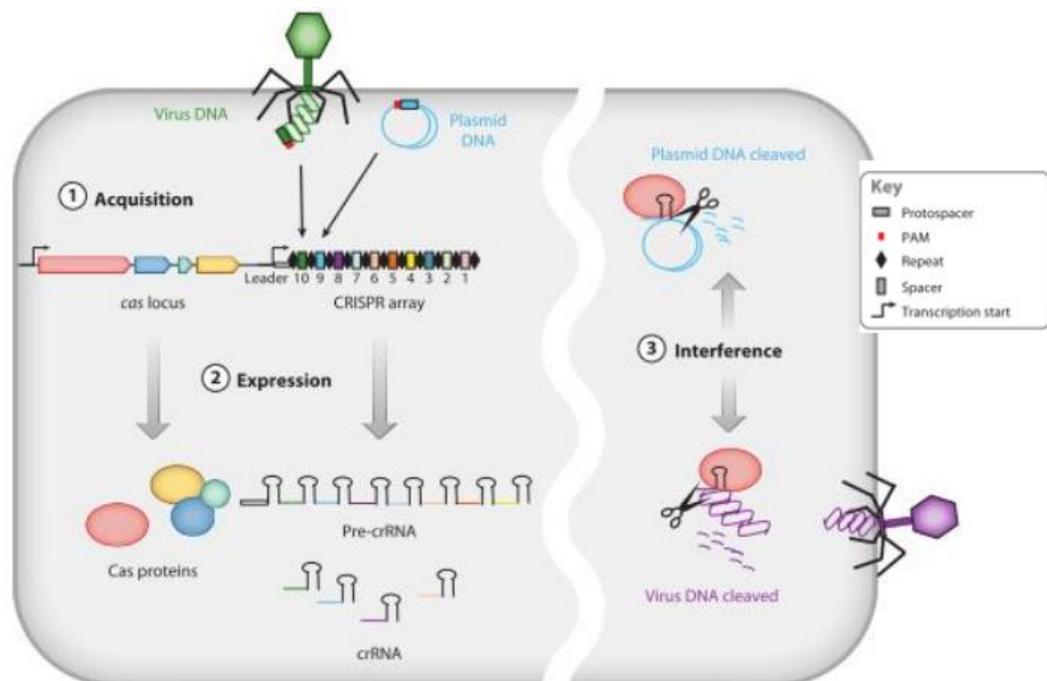


Figura 18. Ilustración dónde se reflejan las tres fases del mecanismo de defensa de microorganismos llevado a cabo por CRISPR-Cas. 1- Proceso de adquisición dónde se identifica el material genético exógeno y se integra el protospaciador en forma de espaciador. 2- Expresión de espaciadores primero en forma de precursor y, tras esto, en forma de crRNAs individuales. 3- Interferencia del material genético invasivo mediante la unión del crRNA que señala el corte llevado a cabo por proteínas Cas (Cong *et al.* 2015).

La proteína Cas9 es la encargada de realizar el corte en las hebras del DNA bicatenario (DSB) en la región diana. Una vez producido este corte, la maquinaria celular de reparación del daño en el DNA se pone en marcha. Dicho mecanismo de reparación está presente de forma natural en la gran mayoría de tipos celulares y organismos, y puede seguir dos vías de reparación, la de unión de extremos no homólogos (NHEJ) o la vía de reparación directa por homología (HDR), según se encuentre o no un molde.

Mediante la vía de unión de extremos no homólogos, los dos fragmentos bicatenarios de DNA generados son religados, produciéndose por ello modificaciones en forma de inserciones o deleciones (InDel) que, si se producen en el exón diana, pueden acabar generando un knock-out por cambios en la pauta de lectura o por la aparición de codones de parada prematuros (Ran F. A. *et al.* 2013).

12.1. DISEÑO Y MANIPULACIÓN DE LOS sgRNAs

Para el diseño de los sgRNAs (*single guide RNA*), en primer lugar, se debe realizar una búsqueda bioinformática en la base de datos genómicos Ensembl (http://www.ensembl.org/Homo_Sapiens/Gene/Sequence), para así poder obtener la secuencia deseada del gen *NOMO*. En nuestro trabajo, se seleccionó como secuencia diana el exón 3 y sus regiones intrónicas adyacentes. El motivo de seleccionar este fragmento del gen fue por dos razones principales: la primera que se encuentra al principio de toda la secuencia del gen, por lo que se tiene suficiente longitud de secuencia posterior para generar un codón de parada prematuro y, por tanto, dar lugar a una proteína truncada; y la segunda, que la última base de su secuencia codificante forma parte de un codón cuyas dos bases restantes se encuentran en el siguiente exón. Por tanto, si se elimina este fragmento del exón 3, se romperá el marco de lectura de la secuencia codificante y dará lugar a una proteína *NOMO* truncada (Figura 19).

El diseño de los sgRNAs lo realizamos utilizando la herramienta informática <http://tools.genomengineering.org>. Esta herramienta de diseño proporciona las secuencias para todos los oligonucleótidos necesarios para preparar las construcciones de sgRNA, analizar la eficacia en la modificación del objetivo y evaluar el potencial de corte en regiones fuera del objetivo. Para poner en marcha el diseño de los sgRNAs, una vez dentro del portal web de la herramienta, se rellenaron los campos requeridos y se adjuntó la secuencia de nucleótidos de la región diana del gen *NOMO* (exón 3) en formato *FASTA*. El tamaño de la secuencia debe de estar comprendido entre 23 y 250pb. Posteriormente, se abre una nueva ventana con los

resultados donde aparece un mapa de la región seleccionada del gen con los posibles sgRNAs que se pueden utilizar. También aparece un listado de todos los sgRNAs posibles que se encuentran en la región junto con sus parámetros de eficiencia y especificidad (*score*).

En este trabajo, seleccionamos tres sgRNAs que presentaban un score óptimo para el experimento (Figura 19).

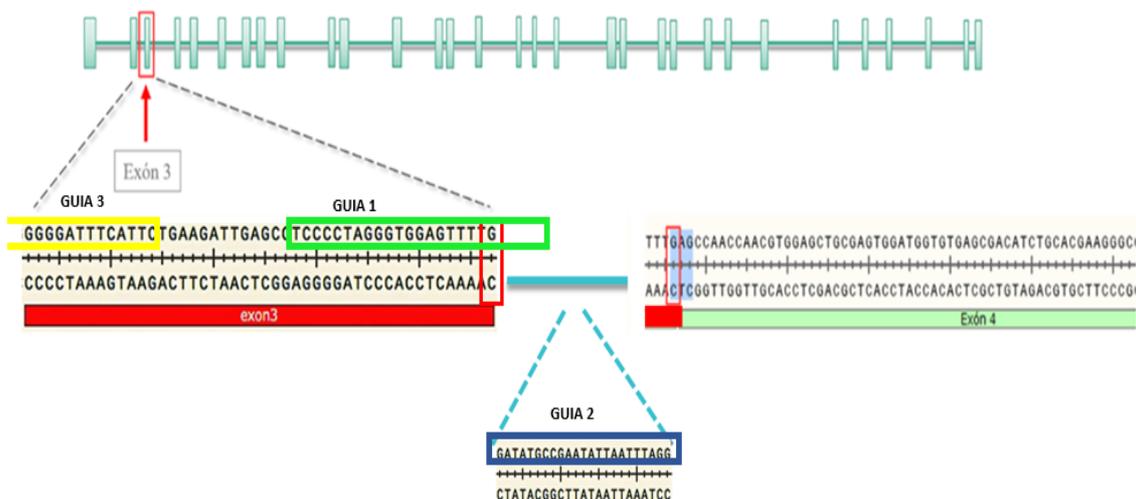


Figura 19. Esquema de la región diana del gen NOMO donde aparecen marcados los tres sgRNAs seleccionados. Se han identificado como guía 1 (verde), guía 2 (gris) y guía 3 (amarillo). Estas secuencias de nucleótidos deben de llevar en el extremo 3' una secuencia PAM (NGG).

Aunque la secuencia PAM no sea un componente del sistema CRISPR/Cas, es inviable usar esta herramienta sin conocer su participación en la misma. La secuencia PAM, cuyo acrónimo deriva de los términos en inglés *protospacer adjacent motif*, se compone de una serie de nucleótidos localizados en la región 3' de la hebra de DNA diana que se une al sgRNA a pocos nucleótidos de la secuencia diana. La presencia de esta secuencia implica la identificación de la diana por parte de Cas9, pues la endonucleasa asociada al sgRNA reconoce, en primer lugar, por el dominio PI a la secuencia PAM, produciéndose la unión DNA-Cas9; posteriormente se separan las dos hebras y la herramienta se desplaza *up-stream* de la hebra en busca de su diana (Stemberg *et al.* 2014). La secuencia nucleotídica de PAM, puede variar dependiendo del microorganismo del que proceda la Cas9 utilizada, pero suele ser 5'-NGG que corresponde a la proteína Cas9 más utilizada, la SpCas9 (*S. pyogenes Cas9*).

En la Tabla 11 se recogen las regiones dianas, así como las secuencias de oligonucleótidos que se seleccionaron como sgRNAs para este experimento.

Tabla 11. Secuencias dianas seleccionada y sus oligos correspondientes. En minúscula se representan las secuencias necesarias para crear los extremos cohesivos.

	sgRNA1	sgRNA 2	sgRNA 3
Secuencia diana	CCAAAACCTCCACCTGGGG	GATATGCCGAATATTAATT	CAGAATGAAATCCCCCTAAA
Oligo UP	5'caccgCCAAAACCTCCACCTAG GGG3'	5'caccgGATATGCCGAATATT AATTT3'	5'caccgCAGAATGAAATCCCCCT AAA3'
Oligo LOW	5'aaacCCCCTAGGGTGGAGTT TTGG3'	5'aaacAAATTAATATTCGGCA TATC3'	5'aaacTTTAGGGGGATTTCATTC TG3'

12.2. CLONAJE DE LOS sgRNAs

Las secuencias de los sgRNAs deben clonarse en un vector plasmídico, con el que posteriormente transfectaremos las líneas celulares. En este caso, se eligió el plásmido pSpCas9(BB)-2A-GFP (PX458) (addgene #48138).

En primer lugar, los sgRNAs deben ligarse a un sgRNA *scaffold* presente en este plásmido, que se encuentra bajo el control del promotor U6. *Up-stream* de esta secuencia, hay una diana de restricción para la enzima BbsI que nos servirá para ligar las secuencias sgRNAs. Para ello, a las secuencias diseñadas como sgRNAs debemos incorporarles los extremos “caccg” y “aaac” para poder insertarlas correctamente en el vector (Tabla 11).

Además, el vector PX458 también tiene bajo el control del promotor CBh el gen de la hSpCas9 y de la EGFP (Figura 20), que nos valdrá para introducir en la célula el segundo componente necesario para el funcionamiento del sistema CRISPR/Cas, el gen hSpCas9, así como el gen *EGFP*, que nos servirá para monitorizar la eficacia de la transfección.

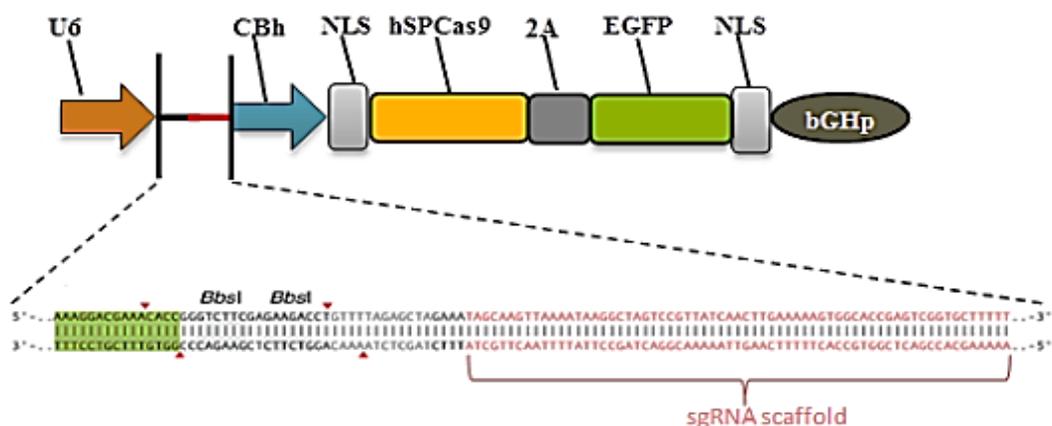


Figura 20. Esquema parcial del vector PX458.

En primer lugar, para cada sgRNA que hemos diseñado, se llevó a cabo la hibridación de los dos cebadores que dará como resultado la futura secuencia guía del sgRNA, es decir, un oligonucleótido de doble cadena con extremos cohesivos para la enzima BbsI y que será el inserto que se introducirá en el vector. Para proceder a la fosforilación y anillamiento de cada par de oligos, la enzima empleada fue la polinucleótido quinasa T4 (PNKT4), que cataliza la transferencia del grupo fosfato a los oligos. En segundo lugar, se llevó a cabo la digestión del vector con la enzima BbsI y, por último, la ligación del inserto con el vector digerido. En la Tabla 12 se especifican los reactivos necesarios para cada una de las etapas del clonaje.

Tabla 12. Reactivos necesarios en cada etapa del clonaje de los sgRNAs.

Etapa	Reactivo	Volumen
Fosforilación y anillamiento de los oligos	Oligo 1 (100 μ M)	1 μ l
	Oligo 2 (100 μ M)	1 μ l
	Buffer de ligación T4 10X (NEB)	1 μ l
	ddH ₂ O	6,5 μ l
	PNK T4 (NEB)	0,5 μ l
	Volumen total	10 μ l
Digestión	PX458	100 ng
	Oligos 1:250	2 μ l
	Buffer Tango 10X	2 μ l
	DTT 10 mM	1 μ l
	ATP 10 mM	1 μ l
	Enzima Ep1	1 μ l
	DNA ligasa T7	0,5 μ l
	ddH ₂ O	hasta 20 μ l
	Volumen total	20 μ l
Ligación	Reacción de ligación	11 μ l
	Buffer PlasmidSafe 10X	1,5 μ l
	ATP 10 mM	1,5 μ l
	Exonucleasa PlasmidSafe	1 μ l
	Volumen total	15 μ l

En la Tabla 13 se especifican las condiciones idóneas de temperatura y tiempo para cada una de las etapas:

Tabla 13. Condiciones de temperatura y tiempo para cada etapa de la clonación de las guías en el vector PX458.

	Temperatura	Tiempo	Gradiente	Ciclos
Fosforilación y anillamiento de los oligos	37°C	30 min.	-	1
	95°C	5 min.	-	1
	95°C -- 25°C	-	5°C/min.	
Digestión	37°C	5 min.	-	6
	23°C	5 min.	-	
	4°C	∞	-	
Ligación	37°C	30 min.	-	1

12.3. TRANSFORMACIÓN DE BACTERIAS COMPETENTES *E.coli DH5* MEDIANTE CHOQUE TÉRMICO

Para la transformación de las bacterias, se descongelaron en hielo un total de 150 μ l de bacterias *E.coli DH5 α* . Una vez descongeladas y bajo la llama para generar un ambiente lo más estéril posible, se añadieron 2 μ L de cada DNA plasmídico, correspondiente a cada sgRNA sobre 50 μ l de las bacterias descongeladas.

Una vez realizada la mezcla, se incubaron los viales 30 minutos en hielo y, pasado este tiempo, se sometieron las células a un choque térmico. Primero se introdujeron los viales en un baño a 42°C durante 30 segundos. Rápidamente, transcurrido este tiempo, se traspasaron de nuevo a hielo y se dejaron reposar durante 5 minutos.

Posteriormente, bajo llama, se añadieron a cada vial 250 μ L de medio 2XTY sin ampicilina previamente calentado, y se incubaron a 37°C a 100-300 rpm durante 30 minutos. Finalmente, se sembraron las bacterias con el asa de Digrafsky en placas de TYE -agar suplementadas con los antibióticos adecuados para la selección de las bacterias transformadas, en nuestro caso con ampicilina. Se dejó la placa hacia arriba para que se absorba la mezcla durante unos minutos, y finalmente se incubaron las 3 placas durante toda la noche a 37°C.

Tres de las colonias que crecieron en las placas se sembraron y crecieron en 1 mL de medio 2XTY con ampicilina. Se repitió el procedimiento para cada una de las tres placas. Esta mezcla se dejó toda la noche en agitación a 37°C. A la mañana siguiente, se centrifugó el cultivo bacteriano, durante 5 minutos a 4000 rpm. Se descartó el sobrenadante, intentando dejar el pellet bacteriano lo más seco posible y se procedió a la extracción de DNA empleando el kit de la casa comercial Danagene plasmid miniprep kit (DANAGEN-BIOTED S.L., Barcelona, España),

que se basa en el empleo de columnas que retienen el DNA, siguiendo las especificaciones del fabricante (Figura 21).

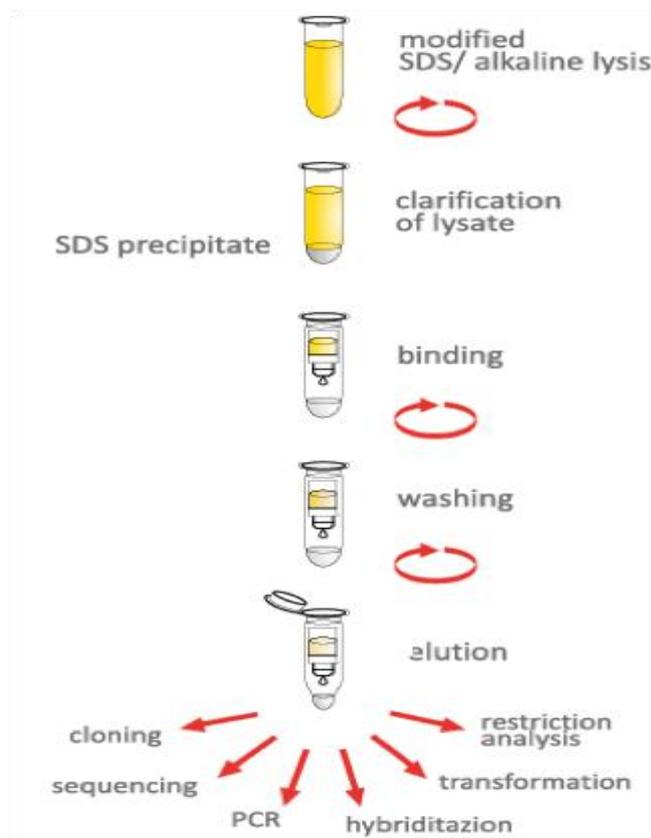


Figura 21. Esquema de los pasos a seguir para la extracción de DNA empleando columnas. Inicialmente, hay una etapa de lisis, seguida por un aclarado del lisado. Posteriormente se pasa esta solución por la columna y se produce la unión del DNA a la membrana. Una vez fijado el DNA, se lleva a cabo un par de lavados para finalmente, eluir el DNA. Modificada de DANAGENE PLASMID Miniprep protocol.

12.4. EXTRACCIÓN Y PURIFICACIÓN DE DNA PLASMÍDICO (MAXIPREP)

Una vez confirmado que las guías se habían incorporado correctamente al plásmido, se recuperaron las tres colonias de bacterias (cada una se corresponde con un sgRNA diferente) y se dejaron crecer en tres matraces diferentes con 200 mL de medio 2XTY junto con 200 μ L de ampicilina. Esta mezcla se dejó toda la noche en agitación a 37°C. A la mañana siguiente, se centrifugó el cultivo bacteriano durante 20 minutos a 4000 rpm. Se descartó el sobrenadante intentando dejar el *pellet* bacteriano lo más seco posible y, se procedió a la extracción de DNA empleando el kit GeneJET Plasmid Maxiprep Kit (Thermo Scientific™, United States) siguiendo las etapas indicadas en la Tabla 14

Tabla 14. Etapas a seguir para la extracción y purificación de DNA plasmídico (Maxiprep)

Etapa	Pasos a seguir
1	-Resuspender el pellet celular en 6 mL. de la solución de Resuspensión por vórtex o pipeteo
2	-Añadir 6 mL. de "Lisis Solution" y mezclar invirtiendo el tubo de 4 a 6 veces hasta que la solución se vuelva viscosa y de color claro. Posteriormente, incubar durante 3 min. a temperatura ambiente.
3	-Añadir 6 mL. de "Neutralization Solution" y mezclar invirtiendo el tubo de 5 a 8 veces.
4	-Añadir 0,8 mL. de "Endotoxin Binding Reagent" y mezclar invirtiendo el tubo de 5 a 8 veces. Incubar 5 min. a temperatura ambiente.
5	-Centrifugar durante 20 min. a 20.000 rpm para sedimentar todos los residuos celulares
6	-Transferir el sobrenadante a un tubo de 50mL. por decantación o pipeteo. Evitar transferir el precipitado blanco. Añadir 1 volumen de etanol al 96%. Mezclar inmediatamente invirtiendo el tubo 5-6 veces.
7	-Transferir parte de la muestra (hasta 20 mL.) a la columna. Centrifugar durante 3 min. a 2.000 x g. Retirar el sobrenadante.
8	-Repetir la etapa 7 hasta pasar todo el líquido por la columna.
9	-Añadir 8mL. de "Wash Solution I" a la columna de purificación. Centrifugar durante 2min. a 3.000 x g. Desechar el sobrenadante.
10	-Añadir 8mL. de "Wash Solution II" a la columna de purificación. Centrifugar durante 2min. a 3.000 x g. Desechar el sobrenadante.
11	-Repetir la etapa 10 una vez más.
12	-Centrifugar durante 5min. a 3.000 x g. Desechar el sobrenadante.
13	-Transferir la columna a un tubo de 50 mL. Añadir 1 mL. de "Elution Buffer" al centro de la columna de purificación. Incubar durante 2 min. a temperatura ambiente y centrifugar durante 5 min. a 2.000 x g
14	-Conservar el tubo con DNA a -20°C hasta su uso.

12.5. TRANSFECCIÓN DE LA LÍNEA CELULAR HT29 CON PX458-sgRNAs

Una vez que obtuvimos los tres plásmidos PX458, cada uno con su correspondiente sgRNA, pasamos a transfectar la línea celular HT29.

12.5.1. CONTAJE CELULAR

Para el conteo de las células se empleó la cámara de Neubauer o hematocitometro. Se trata de una gruesa placa de cristal con forma de portaobjetos, de unos 30 x 70 mm y unos 4 mm de grosor. En nuestro caso, empleamos una cámara doble, que son las más comunes. Típicamente, el rango de concentraciones que permite contar el hematocitómetro está entre 250.000 células y 2,5 millones de células por mL.

En primer lugar, se colocó un cubreobjetos sobre la cámara de Neubauer y se colocaron en posición horizontal sobre la mesa. Se tomaron 10 µl de la muestra con una micropipeta y se colocó la punta de la pipeta en el borde del cubreobjetos, en el extremo de la cámara de Neubauer. Se trata de dejar que el líquido penetre entre la cámara y el cubreobjetos desde el lateral, por capilaridad. Posteriormente, se colocó la cámara de Neubauer en la bandeja del microscopio y se comenzó a realizar el recuento de células en el primer cuadro. Existe una convención por la cual, si las células tocan el límite superior o el límite izquierdo del cuadro, deben contabilizarse, pero no se contabilizan si tocan el límite inferior o el límite derecho. Se repite el proceso para el resto de los cuadros que deseamos contar, anotando el resultado de cada uno de ellos. Cuantos más cuadros contemos, más precisión obtendremos en nuestra medida. Finalmente, se calcula la concentración:

$$\text{Concentración} = \frac{\text{Número de células} \times 10.000}{\text{Número de cuadros}}$$

12.5.2. TRANSFECCIÓN POR LIPOFECCIÓN

Inicialmente, elegimos la técnica de lipofección para la transfección de las células HT29. Para esta técnica empleamos la lipofectamina (ThermoFisher Scientific, ref. 11668019). Este reactivo causa la fusión de liposomas con la membrana plasmática de células vivas, permitiendo el paso del DNA plasmídico al interior celular. Las células se transfectaron con las siguientes combinaciones:

-Transfección con el plásmido vacío (control).

- Transfección con los plásmidos que contienen sgRNA 1 + sgRNA2.
- Transfección con los plásmidos que contienen sgRNA 1 + sgRNA3.
- Transfección con los plásmidos que contienen sgRNA 2 + sgRNA3.

Para la transfección, se emplearon placas de 6 pocillos (2 mL/pocillo). Se plaqueó 1 millón de células HT29 por pocillo para que la confluencia a las 24 horas fuese del 80%. Pasado este tiempo, se llevaron a cabo las transfecciones, todas ellas por duplicado.

Para la transfección del plásmido vacío, se añadieron 5 µg del plásmido a 150 µl de medio Opti-Mem (Life Technologies, Paisley, UK) en un tubo de 1,5 mL. En el caso de las transfecciones con los plásmidos que contenían las guías, se añadieron 2,5 µg de cada plásmido a 150 µl del mismo medio anterior en un tubo de 1,5 mL. Por otro lado, se añadieron 9 µl de lipofectamina a 150 µl de medio Opti-Mem en un tubo de 1,5 mL. Esta última mezcla se hizo por cuadruplicado (una para cada transfección). Ambas mezclas se incubaron durante 5 minutos a temperatura ambiente. A continuación, se juntaron y se llevó a cabo una incubación de 30 minutos a temperatura ambiente. Posteriormente, se añadió cada mezcla de transfección gota a gota sobre toda la superficie de su correspondiente placa de cultivo a las que previamente se había añadido medio DMEM suplementado con un 10% de FBS y un 1% de P/S (penicilina/strepto). Tras 48 horas se procede a la selección de las células que han incorporado las construcciones por la emisión de luz verde gracias a la expresión de EGFP.

12.5.3. TRANSFECCIÓN POR ELECTROPORACIÓN

Este método permite el paso del plásmido al interior celular debido a un aumento de la conductividad eléctrica y de la permeabilidad de la membrana plasmática celular causado por la aplicación externa de un campo eléctrico. Las transfecciones realizadas fueron las mismas que en el caso anterior pero, además, se aplicó el campo eléctrico a células HT29 sin transfectar con ningún plásmido para utilizarlas como control.

El primer paso consistió en mezclar en varios tubos, uno para cada condición de transfección, 10 µg de plásmido total y el volumen de buffer de electroporación necesario para tener un volumen final de 100 µl. En el caso de las transfecciones de plásmido con guía se emplearon 5 µg de cada uno de los plásmidos de la combinación y, en el caso del plásmido vacío se añadió 10 µg de este.

Para preparar 1L de buffer de electroporación:

- Hepes 4,766 g

- Dextosa 1,081 g
- Na₂HPO₄ 0,099 g
- KCl 0,370 g
- NaCl 7,73 g
- Ajustar el pH a 7,2
- Filtrar para esterilizar.

Se electroporaron 1 millón de células en cada una de las condiciones mencionadas anteriormente. Se centrifugaron las células para lograr su precipitación y, tras aspirar el medio de cultivo, se resuspendieron en el buffer de electroporación que contenía el plásmido o plásmidos de interés en cada caso. Cada combinación se pasó a una cubeta de electroporación. La electroporación se realizó con el equipo de Amaxa (Lonza), empleando el programa W017, recomendado para las células HT29. Tras la electroporación, se recogieron las células con una pipeta Pasteur y se añadieron gota a gota a pocillos de placas de 6 pocillos con 2 mL de medio DMEM con un 10% de FBS sin suplementar con antibiótico. Tras las 48 horas, con el fin de comprobar si la incorporación de los respectivos plásmidos se ha realizado con éxito, se observaron las células transformadas bajo un microscopio de fluorescencia para analizar la emisión de fluorescencia verde frente a un control negativo de células sin transfectar.

12.6. CELL SORTING MEDIANTE CITOMETRÍA DE FLUJO

Pasadas las 48 horas se llevó a cabo la separación de las células que habían incorporado el plásmido mediante citometría de flujo. La separación celular mediante citometría de flujo, también llamada *cell sorting*, consiste en la separación física de las células en base a la expresión de uno o varios parámetros analizables por citometría de flujo analítica. En este caso, la separación se hizo en función de la emisión de luz verde, ya que, como se ha indicado anteriormente, el plásmido que hemos utilizado para este experimento contiene el gen que codifica para la proteína EGFP.

Se llevó a cabo la separación célula a célula, ya que el objetivo es generar una línea celular *knockout* para *NOMO*. De este modo, para cada condición (HT29 sin plásmido, HT29 con PX458 sin guías, HT29 G1+G2, HT29 G1+G3, HT29 G2+G3) se prepararon 3 placas de 96 pocillos con 100 µl de medio de cultivo y en cada pocillo se puso una única célula que expresaba EGFP. Las células HT29 que habían sido sometidas a la electroporación sin plásmido, aunque no expresaban EGFP, fueron separadas también, para poder comparar el crecimiento a partir de

una única célula en todas las condiciones. Se dejaron crecer los clones obtenidos a partir de la edición génica con CRISPR-Cas9 durante aproximadamente un mes. Pasado este tiempo, se comprobó en el microscopio pocillo a pocillo de todas las placas, los clones que habían crecido. En cuanto fue posible se pasaron a placas de 48, 24 y 6 pocillos sucesivamente.

13. ENSAYO DE VIABILIDAD CELULAR

El crecimiento y la supervivencia celular en los clones obtenidos se evaluó mediante el ensayo de viabilidad MTT (Mosmann *et al.* 1983). Esta técnica consiste en el marcaje de células metabólicamente activas con bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolio (MTT) (Sigma-Aldrich). Se trata de un compuesto de color amarillento que, al ser reducido por la enzima succinato deshidrogenasa (SDH) celular, da lugar a un compuesto hidrofóbico de color azul oscuro, el formazán. El formazán se solubiliza con DMSO y, posteriormente, se mide la intensidad de color con un espectrofotómetro, lo que permite determinar la tasa de proliferación. La absorbancia se correlaciona con el color e, indirectamente, con la cantidad de células en cultivo metabólicamente activas.

Para llevar a cabo este ensayo se sembraron un total de 10.000 células por pocillo en placas de 24 pocillos y se las dejó crecer toda la noche. A continuación, se incubaron las placas multipocillo durante 24, 48, o 72 horas, utilizando también una placa previa (tiempo 0) como control. Una vez transcurridos los tiempos se añadió a cada pocillo 1mg/ml de MTT y se incubó durante 1 hora a 37°C. Tras esto, se eliminó el medio y se añadió DMSO para disolver los cristales de formazán depositados en cada pocillo. La absorbancia de los mismo se determinó en el lector de microplacas *UltraEvolution* (Tecan®) a 570nm.

De cada experimento se realizaron dos réplicas para confirmar los resultados.

14. ENSAYO DE MIGRACIÓN CELULAR

El ensayo de cierre de herida tiene como objetivo el estudio de la migración celular. Se basa en la observación del comportamiento de una monocapa confluyente de células a la que previamente se le ha realizado una “herida”. Las células en el borde de la herida se moverán hacia la abertura hasta establecer nuevos contactos célula-célula, cerrando así la herida. Este método permite determinar la velocidad a la que las células migran cuando se desplazan a través de una herida o *scratch*.

Para evaluar el bloqueo de *NOMO* sobre la capacidad de migración celular, se sembraron 700.000 células por pocillo en placas de 6 pocillos. El ensayo se realizó con tres replicas para cada clon. Una vez que las placas estaban confluentes se les retiró el medio de cultivo y se lavaron con tampón fosfato salino (PBS). Al retirar el PBS, con la ayuda de una punta de micropipeta, se practicaron tres *scratches* por pocillo y se añadió medio de cultivo sin ser suplementado con FBS y se agregó mitomicina C a una concentración de 0,01 mg/mL a todos los pocillos.

Con la ayuda de una cámara acoplada a un microscopio de contraste de fases (Mikon ECLIPSE TE-2000-E), se tomaron fotos de una región determinada de cada *scratch*, cada 10 minutos durante 48 horas. Durante todo el experimento, las células se mantuvieron en una atmósfera a 37°C y CO₂ al 5%.

Para calcular el porcentaje de migración se utilizó el programa ImageJ® junto con la herramienta *MRI Wound Healing Tool* que mide el area libre de células. Los resultados finales se obtuvieron de la media de las tres réplicas por cada clon.

15. MICROARRAYS GENÓMICOS CON LA PLATAFORMA CYTOSCAN™

Se realizó el estudio citogenético de las 3 líneas celulares: HT29 normal, HT29 con PX458 vacío y HT29 *knockout* para *NOMO* mediante microarrays genómico con la plataforma *Affymetrix® CytoScan™ 750k Assay* (Affymetrix, Santa Clara, California, USA).

El análisis con esta técnica otorga información de alta resolución sobre CNAs además de una gran cobertura de SNPs que permiten detectar fiablemente LOH. La matriz consiste en 2.696.550 sondas que incluyen 743,304 SNPs y 1,953,246 sondas no polimórficas. El espaciamiento medio de la sonda para los genes RefSeq es de 880 pb, encontrándose representados en el 96% de los genes. En resumen, 500 ng de DNA genómico fueron digeridos con Nsp1 durante 2 hr a 37°C (Figura 22). El DNA digerido se purificó y se ligó con Primer / adaptadores a 16°C durante 12-16 horas. Posteriormente, sobre estos productos, se generaron amplicones de DNA mediante PCR utilizando cebadores proporcionados por el fabricante (Affymetrix CA). Los productos de PCR se purificaron y se digirieron con DNaseI durante 35 min a 37°C. El DNA fragmentado fue entonces marcado con nucleótidos biotinilados a través del Desoxinucleótido Terminal Transferasa (TDT) durante 4 horas a 37°C. Finalmente, 250 µg de DNA fragmentado se hibridaron con el Chip Affymetrix chip Cytoscan HD preequilibrado a 50°C durante 18hrs. Las matrices se lavaron y se tiñeron en la GeneChip® Fluidics Station 450

(Affymetrix Inc.) y las Imágenes DAT fueron adquiridas con el GeneChip® Scanner 3000 7G (Affymetrix Inc.)

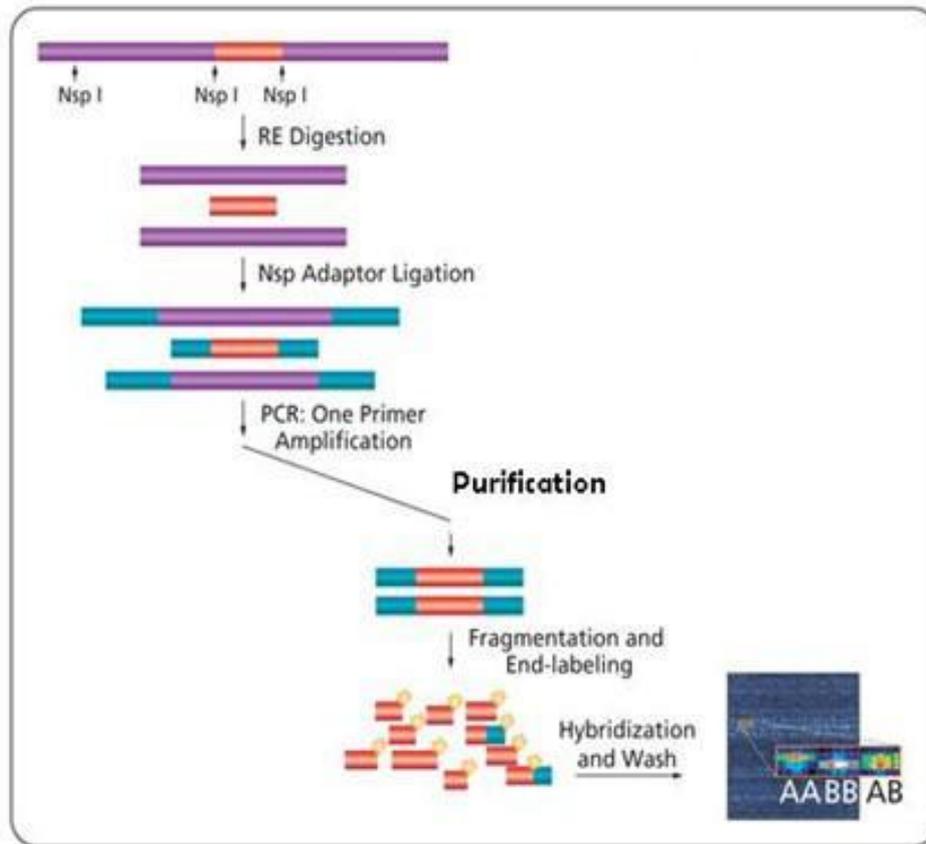


Figura 22. Esquema de trabajo de la plataforma de *microarrays* genómicos *CytoScan 750k*.

Se procesaron y analizaron los ficheros CEL *CytoScan 750 k* correspondientes a las tres muestras pareadas en la consola ChAS (*Chromosome Analysis Suite*) v3.1.0.15 de Affymetrix. Estos ficheros son archivos de intensidad de fluorescencia utilizados para el análisis. Los archivos CEL sin procesar se compararon con una referencia proporcionada por el fabricante que consta de 284 muestras de HapMap (*haplotype map*) La referencia del genoma humano utilizada fue GRCh37/Hg19 y la opción de análisis de muestra única (SSA).

El análisis elegido comienza con un preprocesamiento que se realiza mediante una doble normalización de la señal por cuartiles, obteniendo log ratios de las señales. Después, se calculan las intensidades máximas específicas para cada SNP y, por último, para analizar el número de copias (CN, *copy number*) se realiza una segmentación del genoma que nos va a permitir obtener finalmente un valor de CN para un determinado segmento; es decir, se decide si un determinado grupo de sondas pertenecen a un segmento resultando en un CN o si pertenece al conjunto próximo de sondas en otro segmento. Los log ratios se asignan a los segmentos mediante un

modelo oculto de Markov (HMM, *Hidden Markov Model*) que va a permitir detectar las regiones ganadas o perdidas. Para prevenir detectar falsos positivos, este modelo utiliza el log ratio esperado asociado a un número de copias para asignar un punto de corte al segmento teniendo en cuenta lo observado. La mediana de la diferencia entre pares absolutos (MAPD), que mide la variabilidad de la matriz en log ratios, se utilizó para evaluar la calidad de la matriz, teniendo en cuenta como muestras de buena calidad aquellas con un MAPD <0,35.

El análisis del número de copias se realizó con el software *Nexus Express 3.0* (BioDiscovery Inc., CA, EE.UU.). Este software proporciona el número de copias de cada sonda a partir de los datos brutos proporcionados por el scanner de Affymetrix, tras realizar un proceso de “smoothing” o suavización de los datos y una corrección cuadrática.

Para detectar cambios focales se utilizó el programa GISTIC 2.0 (*Genomic Identification of Significant Targets in Cancer*) que utiliza un q-valor de corte asociado con la amplitud de la alteración y la frecuencia en la que ocurre. Sólo las alteraciones que presentan un log ratio positivo se consideran ganancias; aquellas negativas son pérdidas (Mermel *et al.* 2011).

16. ANÁLISIS DE EXPRESIÓN DE PROTEÍNAS DE LA VÍA NODAL

16.1. EXTRACCIÓN DE PROTEÍNAS A PARTIR DE LÍNEAS CELULARES

La extracción de proteínas a partir de líneas celulares, se llevó a cabo a partir de las células de las placas de cultivo usando 1.5 mL del agente de disociación TrypLE 1X (Gibco), se lavaron con PBS y se centrifugaron a 1500rpm tres minutos. Los precipitados se incubaron durante 30 minutos con buffer de lisis en frío y en agitación (NaCl 140mM, EDTA 10mM, 10% glicerina, 1% Nonidet P-40, Tris 20mM pH 8.0), al que se le añadieron los siguientes inhibidores de proteasas: PMSF 1mM, pepstatina 1 μ M, aprotinina 1 μ g/mL, leupeptina 1 μ g/mL y ortovanadato sodico 1mM. Tras centrifugar los tubos durante 10 minutos a 4°C y 12000 rpm se transfirieron los sobrenadantes resultantes a un nuevo tubo Eppendorf® y se almacenaron a -80°C.

16.2. WESTERN BLOT

Mediante la técnica de Western Blot analizamos la expresión de las proteínas NOMO, NCLN, TEMEM147, ALK4, Cripto-1, Smad2, P-Smad2, Smad3, P-Smad3, Smad4 y Actr2. Para ello, las muestras de proteína total se prepararon mezclando 200 μ g de cada extracto proteico con tampón de carga (Tris 100mM pH 6.8, 20% glicerol, 4% SDS, 0.05% azul de bromofenol y 2% β -

mercaptoetanol) y se desnaturalizaron durante 5-10 minutos a 100°C en un baño seco. Tras esto, se procedió a la carga de las muestras en geles SDS-PAGE para su separación en función del tamaño mediante electroforesis vertical con minigeles discontinuos (*Sodium-Dodecylsulfate-Polyacrilamide Gel Electrophoresis*) siguiendo el protocolo de Laemmli (Laemmli *et al.* 1970). El gel separador se preparó a un porcentaje de acrilamida adecuado al tamaño de las proteínas a separar (8-12% acrilamida/bis-acrilamida 30%) (BioRad) en tampón Tris-HCl 1.5 M pH 8.8 y SDS al 10%. Encima del gel separador, se preparó un gel concentrador al 4% acrilamida/bis acrilamida 30% en tampón Tris-HCl 0.5 M pH 6.8 y SDS al 10%. Para la polimerización de los geles se utilizó APS y TEMED.

La electroforesis se realizó a 150 voltios en condiciones desnaturalizantes en tampón de electroforesis (Tris 20mM, Glicina 192mM y SDS al 1%) y como marcador de peso molecular se utilizó el marcador *PageRuler™ Prestained Protein Ladder Plus* (Thermo Fisher Scientific) en el que los tamaños que muestra en kDa son: 170, 130, 100, 70, 55, 35, 25, 15 y 10.

Una vez separadas por tamaño, las proteínas fueron transferidas a una membrana de PVDF *Immobilon-P* (Millipore Corporation) utilizando el sistema de transferencia semiseca TRANS BLOT® SD (BioRad) durante 45 minutos a 15 voltios en tampón 25mM Tris-HCl pH 8.5, 192mM Glicina y 20% metanol. Posteriormente, la membrana fue bloqueada durante 1 hora en una solución de leche desnatada en polvo al 5% en TBS-T (0.1 % Tween-20 en TBS) para saturar todos los sitios inespecíficos de unión a proteínas. A continuación, la membrana se incubó con el anticuerpo primario específico para la proteína de interés a la dilución adecuada en TBS-T con BSA (Bovine Serum Albumin, Sigma-Aldrich) al 3% toda la noche a 4°C. Tras tres lavados de la membrana con TBS-T de 5, 10 y 15 minutos respectivamente en agitación, la membrana se incubó con el anticuerpo secundario correspondiente conjugado con peroxidasa de rábano (HRP) durante 45 minutos a temperatura ambiente. Tras realizar los mismos lavados que anteriormente, se utilizó el sistema de quimioluminiscencia basado en la oxidación del luminol en presencia de peróxido de hidrogeno *Pierce™ ECL Western Blotting Substrate* (Thermo Fisher Scientific), según las especificaciones del fabricante y se expuso la membrana a películas fotosensibles de autorradiografía (FUJIFILM) para la detección de proteínas. En la tabla 15 mostramos los anticuerpos utilizados en nuestro trabajo.

Tabla 15. Anticuerpos usados en este trabajo en los experimentos de western blot.

Anticuerpo	Tipo	Origen	Conce. wb	Casa comercial	Tamaño proteína (kda)
ACTINA	PRIMARIO	RATÓN	1:10000	SigmaAldrich (A5441)	43
NOMO	PRIMARIO	CABRA	1:200	R&D Systems (AF3755)	130
NCLN	PRIMARIO	CONEJO	1:500	SigmaAldrich (SAB1501379)	62
TEMEM147	PRIMARIO	CONEJO	1.1000	ABCAM(ab97624)	25
ALK4	PRIMARIO	CONEJO	1.1000	ABCAM (ab109300)	57
CRIPTO-1	PRIMARIO	CONEJO	1:2000	Novus Biologicals(100-1598)	18
SMAD2	PRIMARIO	CONEJO	1:1000	Cell signaling (134D4)	60
P-SMAD2	PRIMARIO	CONEJO	1:500	Cell signaling (D43B4)	60
SMAD3	PRIMARIO	CONEJO	1:1000	Cell signaling (C67H9)	52
P-SMAD3	PRIMARIO	CONEJO	1:1000	Cell signaling (C25A9)	52
SMAD4	PRIMARIO	RATÓN	1:1000	Santa Cruz(sc-7966)	61
ACTR2	PRIMARIO	RATÓN	1:200	Santa Cruz(sc-390977)	58
Anti-Mouse	SECUNDARIO	CABRA	1:10000	GE Healthcare (NXA931)	
Anti-Rabbit	SECUNDARIO	CABRA	1:10000	Millipore (AP307P)	
Anti-Goat	SECUNDARIO	BURRO	1:10000	SCBT (sc-2020)	

Los fragmentos correspondientes a las proteínas de interés se cuantificaron usando el programa de análisis de imagen ImageJ.

17. ARRAYS DE EXPRESIÓN (WT PLUS DE AFFYMETRIX)

El ensayo de arrays de expresión se llevó a cabo en el servicio de Genómica del Centro de Investigación del Cáncer (Salamanca). Para ello la integridad del RNA se evaluó utilizando el bioanalizador Agilent 2100 (Agilent, Palo Alto, CA). Posteriormente se llevó a cabo el etiquetado y las hibridaciones de las muestras que se realizaron de acuerdo con los protocolos de Affymetrix. En resumen, se amplificaron y marcaron 100 ng de ARN total de cada clon utilizando el kit de reactivos *WT Plus* (Affymetrix) y, posteriormente se hibridaron en *Clariom S Human Array* (Affymetrix). El lavado y el escaneo se realizaron utilizando el sistema *GeneChip* de Affymetrix (*GeneChip Hybridization Oven 645*, *GeneChip Fluidics Station 450* y *GeneChip Scanner 7G*).

-ANÁLISIS DE DATOS (WebGestalt)

Para el análisis de los datos obtenidos mediante el ensayo de los arrays de expresión se utilizó la herramienta WebGestalt (*WEB-based Gene Set Analysis Toolkit*, <http://www.webgestalt.org>). Se trata de una herramienta web de análisis funcional que se ha ido actualizando constantemente que permite llevar a cabo análisis de enriquecimientos funcionales, pudiendo estudiar diferentes organismos, reconociendo multitud de identificadores diferentes que mapean con identificadores de genes de Entrez y dispone de 78.612 categorías funcionales donde poder clasificar los genes diferencialmente expresados (Wang *et al.* 2013).

El funcionamiento de la herramienta web es considerablemente sencillo y se puede resumir en los siguientes pasos:

- Introducir la lista de los genes diferencialmente expresados (ya sea pegando los nombres directamente o subiendo un archivo de texto que contenga dicha lista), seleccionar el organismo que se está estudiando y el tipo de identificador de los genes que se ha introducido (en nuestro caso Gene Symbol, pero acepta una gran diversidad de identificadores), los cuales se transformarán a identificadores de genes de Entrez (el programa devuelve también aquellos genes para los que no se ha encontrado identificador de genes Entrez).

- Seleccionar el tipo de análisis de enriquecimiento a realizar: GO, KEGG, Wikipathways, Pathway Commons, dianas de Factores de Transcripción, dianas de microRNAs, redes de interacción de proteínas, localización cromosómica, asociación a enfermedades, asociación a fármacos, fenotipos o PheWAS.

- Seleccionar el conjunto de datos de referencia para el análisis de enriquecimiento.

- Seleccionar el método de ajuste para testeo múltiple (BH, BY, Bonferroni, holm, hommel o ninguno) y el nivel de significancia.

18. ANÁLISIS DE LOS GENES IMPLICADOS EN LA VIA NODAL MEDIANTE SECUENCIACIÓN MASIVA

18.1. DISEÑO DEL PANEL

Para la determinación de mutaciones puntuales de los genes implicados en la vía Nodal, contactamos con la compañía Illumina Inc. (San Diego, Ca, USA) para que nos diseñaran un panel *custom* en el que se incluyeran los genes implicados en esta vía: ACTR2 (Receptor tipo II),

ACVR1B (Receptor tipo I), GPR78, LEFTY1, NCLN, NOMO1, SMAD2, SMAD3, SMAD4, TDGF1 (Crypto1) y TMEM147.

Illumina ha desarrollado distintos modelos de secuenciadores tras adquirir el primer Genome Analyzer comercializado por Solexa en 2006. Todos los secuenciadores de Illumina (HiSeq, MiSeq, NextSeq, etc.) utilizan la tecnología de secuenciación por síntesis (Ju *et al.* 2006). En la actualidad, los secuenciadores de Illumina pueden conseguir lecturas de 2 x 300 pb hasta 15 Gb y 25 millones de lecturas (ML) (modelo MiSeq) o bien lecturas de 2 x 150 pb hasta 1800 Gb y 6.000 ML (modelo HiSeqX). El equipo utilizado en este trabajo para llevar a cabo el análisis de los genes implicados en la vía Nodal fue el MiSeq. El incremento en el tamaño de lecturas es esencial para aumentar la profundidad de secuenciación, particularmente a la hora de realizar ensamblajes de novo en especies de las que no se dispone de un genoma de referencia. El uso de lecturas pareadas pertenecientes a la misma región genómica permite utilizar la posición de las mismas y su distancia entre ellas para mejorar la precisión del ensamblaje. Los secuenciadores de Illumina son capaces de generar tanto archivos de lecturas simples como de lecturas pareadas, e incluso de lecturas con direccionalidad conocida (stranded-specific). En nuestro caso, el panel se diseñó para obtener archivos de lecturas pareadas.

18.2. MEDIDA DE LA CALIDAD DEL DNA

Como la determinación de mutaciones se realizó en DNA extraído a partir de tumores fijados en bloques de parafina, lo primero que hicimos fue determinar la calidad del DNA que íbamos a incluir en este estudio. Para ello utilizamos el kit: FFPE QC kit (Illumina) siguiendo las recomendaciones del fabricante. Se realiza una qPCR que permite determinar el estado de fragmentación del DNA y, en consecuencia, el potencial de amplificación del DNA. Estos dos parámetros, tanto el estado como el potencial, se comparan con los valores resultantes para el *ACD1* (DNA control) y se calcula el ΔCq para cada muestra. Dependiendo del valor obtenido para el ΔCq se pondrá la cantidad óptima de DNA como se indica en la Tabla 16.

Tabla 16. Cantidad de DNA necesaria dependiendo de la calidad.

Tipo DNA	Tamaño amplicones	Calidad DNA	ΔCq	Cantidad (ng)
FFPE DNA	150 – 175	ALTA	-1 a 1	10
		MEDIA	>1.0 a 2.5	20/50
		BAJA	2.5 a 4.0	50/100

Todas las muestras que no cumplieron estos criterios de calidad se excluyeron del estudio.

18.3. PREPARACIÓN DE LA LIBRERÍA

Para la amplificación de todos los genes que se incluyeron en el panel se siguió el esquema que se representa en la figura 23.

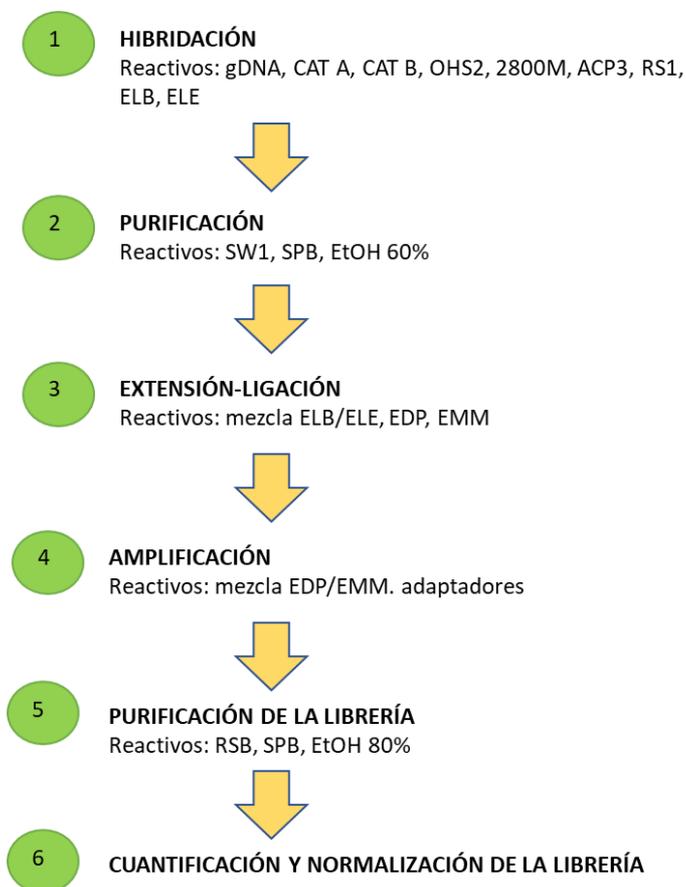


Figura 23. Esquema y reactivos de las diferentes etapas a seguir en la preparación de la librería.

18.3.1. HIBRIDACIÓN

Esta etapa consiste en hibridar todas las sondas diseñadas para la amplificación de los genes que comprenden el panel con las regiones específicas de interés. En nuestro caso, como se trata de identificar variantes somáticas, se aconseja realizar para cada muestra dos réplicas (*pool A* y *pool B*) con el fin de aumentar la especificidad. Para ello se utilizaron placas de PCR de 96 pocillos donde se realizó la mezcla de hibridación: ADN (1:5), FPA (Oligo Pool A) o FPB (Oligo Pool B) (1:10) y OHS3 (Oligo Hybridization for Sequencing 3) siguiendo las especificaciones del fabricante. Para la hibridación se siguieron las etapas que se indican en la tabla 17.

Tabla 17. Condiciones adecuadas para llevar a cabo la hibridación de los oligos con las regiones específicas de interés.

Etapa	Condiciones
1	95°C durante 3 min.
2	Desde los 90°C, disminuir 0.5°C, calentar durante 30 seg., rampa de 0,1°C por seg.
3	Repetir la etapa 2 59x
4	Desde los 60°C, disminuir 0.5°C, calentar durante 1 min., rampa de 0,1°C por seg.
5	Repetir la etapa 4 19x
6	Desde los 50°C, disminuir 1°C, calentar durante 2 min., rampa de 0,1°C por seg.
7	Repetir la etapa 6 9x
8	Desde los 40°C, calentar durante 10 min., rampa de 0,1°C por seg.

18.3.2. PURIFICACIÓN

Durante esta etapa se utilizaron bolas magnéticas (SPB) para eliminar los oligos que no se unieron al DNA. Tres pasos de lavado con SW1 (*Wash Solution 1*) y un lavado con EtOH al 60% aseguraron la eliminación completa de los oligos no unidos. Para la purificación de las muestras se siguieron las especificaciones del fabricante.

18.3.3. EXTENSIÓN Y LIGACIÓN

El resultado de esta etapa es la formación de productos de DNA que contienen las regiones de interés flanqueadas por secuencias requeridas para la amplificación. Para ello una DNA polimerasa extiende el oligo upstream a través de la región del DNA diana para que una DNA ligasa lo ligue al extremo 5' del oligo downstream. Así se consigue la formación de un producto que contiene las regiones de interés flanqueadas por secuencias necesarias para la amplificación posterior.

El proceso se realizó siguiendo las especificaciones del fabricante añadiendo 22µl de reactivo ELE/ELB (*Extension-Ligation Enzyme/Extension-Ligation Buffer*) a cada muestra. Posteriormente, las muestras se sometieron a las condiciones indicadas en la tabla 18.

Tabla 18. Condiciones para llevar a cabo la extensión y ligación.

Etapa	Condiciones
1	37°C durante 45 min.
2	70°C durante 20 min.
3	Enfriar hasta 4°C.

18.3.4. AMPLIFICACIÓN DE LA LIBRERÍA

Durante esta etapa se amplificaron los productos resultantes de la extensión-ligación y, además, se añadió un index 1 (i7) y un index 2 (i5) a cada muestra para poder ser identificada en el análisis. Para ello, se realizó la mezcla de index i5 e i7 en una placa de PCR de 96 pocillos (Figura 24) en la cual se añadió una mezcla de EDP (PCR Master Mix 2) y TDP1 (Truseq DNA Polymerase 1) siguiendo las instrucciones del fabricante. Posteriormente se llevó a cabo la PCR en un el termociclador siguiendo las etapas que se indican en la Tabla 19.

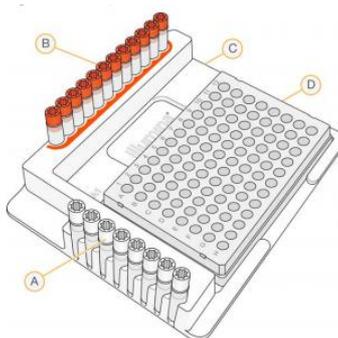


Figura 24. Montaje para la asignación de los index a las muestras. A.-Index i5. B.-Index i7. C.-Soporte. D.-Placa para llevar a cabo la reacción. Los index 5 van del 1 al 8 y los index 7 van del 1 al 12 de tal manera que cada i7 se asigna a una columna de la placa y cada i5 a una fila. TruSeq Custom Amplicon Low Input Kit. Reference Guide. Illumina.

Tabla 19. Condiciones para llevar a cabo la amplificación de librerías.

Etapa	Condiciones
1	95°C durante 3 min.
2	98°C durante 20 seg.
3	67°C durante 20 seg.
4	72°C durante 40 seg.
5	Repetir las etapas 2, 3 y 4 por 29 ciclos
6	72°C durante 1 min.
7	Enfriar hasta 10°C

18.3.5. PURIFICACIÓN DE LA LIBRERÍA

En esta etapa de nuevo se purificaron los productos de las PCRs siguiendo los mismos pasos que se describieron en el apartado 17.3.2. Posteriormente se cuantificó en el *Qubit*[®] la cantidad de DNA que se obtuvo de cada muestra.

18.4. SECUENCIACIÓN EN EL EQUIPO MiSeq (ILLUMINA)

Una vez cuantificadas y normalizadas todas las muestras, se procedió a la secuenciación de estas en un equipo MiSeq (Illumina) siguiendo las especificaciones del fabricante. Para la secuenciación de estas muestras se utilizó un cartucho V3 (2x151 ciclos).

18.5. ANÁLISIS DE LAS VARIANTES

Para analizar los resultados obtenidos en la secuenciación se utilizaron los programas de análisis:

-VariantStudio Software Illumina™ 3.0 (Illumina, San Diego, CA): este programa proporciona un listado con todas las variantes encontradas en cada muestra. Aporta varios datos de cada variante encontrada: lecturas del amplicón donde se encuentra la variante, frecuencia alélica, datos de calidad, tipo de variante, nomenclatura, predicción de la patogenicidad con los programas Sfit y Polyphen, rs (en caso de que exista), estudios de la frecuencia de la variante en diferentes tipos de poblaciones, etc. Este programa proporciona comandos para anotar variantes, filtrar los resultados utilizando varias opciones de filtrado, clasificar las variantes según su impacto biológico y exportar los resultados a un informe.

-DNA Nexus: este programa proporciona al igual que el anterior, un listado con todas las variantes encontradas en cada muestra. Básicamente facilita la misma información que el VariantStudio pero, los filtros aplicados en las lecturas obtenidas para los amplicones de cada programa son diferentes.

-Integrative Genomics Viewer (IGV 2.4): este programa permite la visualización de la secuencia de nucleótidos de cada amplicón secuenciado. Aporta el número de lecturas de cada amplicón así como los cambios detectados. En nuestro caso, se usó para la confirmación de las variantes dadas por los dos programas anteriores.

18.5.1. PROGRAMAS PARA EL ESTUDIO *IN SÍLICO* DE MUTACIONES NO DESCRITAS

Los estudios *in silico* permiten estudiar de manera predictiva el grado de patogenicidad de una mutación. Es importante su realización siempre que se encuentre una mutación no descrita anteriormente y sirve como estudio complementario a otros análisis. Los resultados de los estudios *in silico* nunca poseen un carácter determinante pero sí orientativo. Para el estudio de mutaciones missense utilizamos tres programas basados en la conservación en la evolución del aminoácido o secuencia de aminoácidos a los que afecta:

-POLYPHEN (POLYMORPHISM PHENOTYPING)

Este programa predice el posible impacto de una sustitución aminoacídica en la estructura y función de una proteína humana, empleando consideraciones físicas y comparativas.

Para realizar la predicción se lleva a cabo un alineamiento de secuencias pertenecientes a las proteínas homólogas en las especies que se encuentren descritas en BLAST. El resultado del alineamiento múltiple es empleado por el software PSIC (Position-Specific Independent Counts) para calcular la matriz del perfil. Los elementos de la matriz son ratios (valores del perfil) logarítmicos que indican la probabilidad de que en una posición determinada se encuentre un aminoácido, frente a que este aminoácido se localice en cualquier posición (frecuencia de fondo). El programa PolyPhen analiza los valores absolutos de la diferencia entre los valores de perfil de las dos variantes alélicas en la posición polimórfica. Un valor muy alto de esta diferencia podría indicar que la sustitución estudiada aparece poco o nada en la familia de proteínas. PolyPhen también muestra el número de secuencias alineadas en la posición de búsqueda. Este número indica la credibilidad de los cálculos del valor del perfil.

Los resultados del estudio pueden determinar que el cambio es:

- Probablemente patogénico: existe una gran probabilidad de que afecte a la estructura o a la función de la proteína.
- Posiblemente patogénico: se cree que podría afectar a la función o estructura de la proteína.
- Benigno: es muy probable que carezca de efecto fenotípico.
- Desconocido: en algunos casos, la carencia de datos no permite que el programa pueda hacer una predicción.

PolyPhen también muestra un número de secuencias alineadas en la posición de búsqueda para el análisis de la conservación interespecie ordenadas de mayor a menor homología en la secuencia de la proteína dada.

-SIFT (SHORTING INTOLERANT FROM TOLERANT)

El programa SIFT es una herramienta que clasifica las sustituciones de aminoácidos en una proteína y predice si estos cambios provocarán un efecto en el fenotipo de la proteína.

Se basa en la premisa de que los aminoácidos importantes de una proteína están conservados en la evolución, por lo que los cambios en los mismos deben afectar a la funcionalidad de la proteína. Con una secuencia proteica dada, SIFT escoge proteínas relacionadas y forma un alineamiento múltiple de éstas, y basándose en los aminoácidos presentes en cada posición del alineamiento, realiza una predicción de las sustituciones que afectarán a la proteína. Las sustituciones en una posición conservada en el alineamiento serán consideradas “no toleradas” para la mayoría de los cambios, mientras que las posiciones que no están conservadas en el alineamiento tolerarán mejor los cambios de aminoácido³⁸⁰.

Se introduce la secuencia a analizar en formato FASTA y el programa primero busca secuencias similares a la introducida por nosotros, después escoge las secuencias estrechamente relacionadas que puedan tener una función similar y luego obtiene el alineamiento de las secuencias escogidas. Por último, calcula las probabilidades normalizadas para todas las posibles sustituciones del alineamiento. Las sustituciones con una puntuación menos de 0.05 serán clasificadas como deletéreas y las puntuaciones mayores o iguales a 0.05, “tolerables” o neutrales.

La precisión en las predicciones de SIFT es del 60-80%, obteniéndose un 30% aproximadamente de falsos positivos. No obstante, la precisión de las predicciones depende del alineamiento que se obtenga en cada caso.

-CANCER GENOME INTERPRETER

Cancer Genome Interpreter (Tamborero et al., 2018): es una herramienta para la interpretación de variantes. Utiliza dos herramientas de predicción In Silico: OncodriveMUT y CADD. La primera, OncodriveMUT (Tamborero et al, 2016) clasifica las variantes según sus capacidades tumorigénicas en el desarrollo del tumor. La segunda, CADD, es una herramienta de anotación funcional que integra datos de herramientas existentes de una manera innovadora posibilitando un enfoque integral en la anotación de variantes (Kircher et al., 2014). CADD otorga una puntuación para valorar la causalidad de una mutación siendo > 30 muy alto, 25-30 alto, 20-25 medio-alto, 15-20 medio-bajo y <10 bajo.

18.5.2. ANÁLISIS DE SITIOS REGULADORES DE *SPLICING*

ESE Finder (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>)

Para saber si las variantes en la secuencia afectan a los sitios consenso de unión de las proteínas SR (*Serine/Arginine-rich proteins*) y se altera la unión de éstas a los sitios ESE (*Exon Splicing Enhancer* o sitios exónicos de potenciación del procesamiento) utilizamos el programa de predicción ESE Finder (Cartegni *et al.* 2003; Smith *et al.* 2006). Para ello, se introduce la secuencia normal y la secuencia con el cambio. El programa valora la presencia de sitios ESE tanto en la secuencia normal como en la mutada que puede provocar el cambio de nucleótido. A nivel general, consideran significativos los cambios exónicos que hacen que desaparezca algún sitio ESE. El programa estudia el cambio de los sitios consenso de unión de estas proteínas SF2/ASF, SF2/ASF (IgM-BRCA1), SC35, SRp40 y SRp55. Considera los siguientes valores umbral para cada una de ellas: SF/ASF 1,956; SF/ASF (IgM-BRCA1) 1,867; SC35 2,386; SRp40 2,67; SRp55 2,676.

Como valor umbral se considera al valor mínimo para que una secuencia determinada se considere significativa para la unión de alguna de las proteínas de procesamiento.

19. BASES DE DATOS

Las diferentes bases de datos con las que hemos trabajado para el desarrollo de este trabajo se detallan a continuación.

-PUBMED “U.S. National Library of Medicine National Institutes of Health”

www.ncbi.nlm.nih.gov/pubmed

Es un motor de búsqueda de libre acceso a bases de datos bibliográficas, entre ellas principalmente MEDLINE. Esta base de datos es un recurso creado por la “National Library of Medicine” de Estados Unidos, y recoge referencias bibliográficas de artículos publicados en más de 4.500 revistas médicas desde 1966 hasta la fecha.

-UCSC

genome.ucsc.edu

La base de datos UCSC Genome Browser fue creada en el año 2000 por dos investigadores de la Universidad de California Santa Cruz y posteriormente entró a formar parte del proyecto Genoma Humano. Con 13 años de antigüedad, contiene secuencias de 58

organismos y ofrece un amplio rango de posibilidades debido a las múltiples herramientas que ofrece³⁷⁷.

-COSMIC “Catalogue of Somatic Mutations in Cancer”

www.sanger.ac.uk/genetics/CGP/cosmic

Esta base de datos es uno de los recursos disponibles on-line que ofrece el Instituto Sanger para proveer información sobre mutaciones somáticas y detalles relacionados con cáncer en humanos.

-ClinVar-NCBI-NIH

ncbi.nlm.nih.gov/clinvar

ClinVar es un archivo público de acceso gratuito de información sobre la relación entre las variantes humanas y los fenotipos. Esta base de datos aporta información acerca de la patogenicidad de las variantes encontradas. Además, también recoge referencias bibliográficas de artículos publicados acerca de estas variantes.

RESULTADOS



1. ESTUDIO DE LA REGIÓN 16p13.12-p13.1

1.1. ESTUDIO DE LA REGIÓN 16p13.12-p13.11 EN PACIENTES CON CCR

En este trabajo, partimos del estudio previo de nuestro grupo (Arriba *et al.* 2015) (anexo 11), en el que se realizó el análisis del perfil cromosómico de 146 tumores de colon mediante la técnica de CGH arrays (aCGH), para comprobar si existían diferencias moleculares entre el grupo de tumores diagnosticados en el paciente joven y el grupo de tumores diagnosticados en el paciente anciano.

Se analizó con más profundidad la región 16p13.12-p13.11 de los 146 tumores que se habían incluido en el estudio inicial. Dentro de este grupo de tumores, 60 pertenecían a pacientes jóvenes y 86 a tumores procedentes de adultos. El análisis de estos datos reveló una deleción recurrente en esta región cromosómica que se podía presentar aislada, como único cambio en la región, o bien, acompañada de otros cambios (Figura 25).

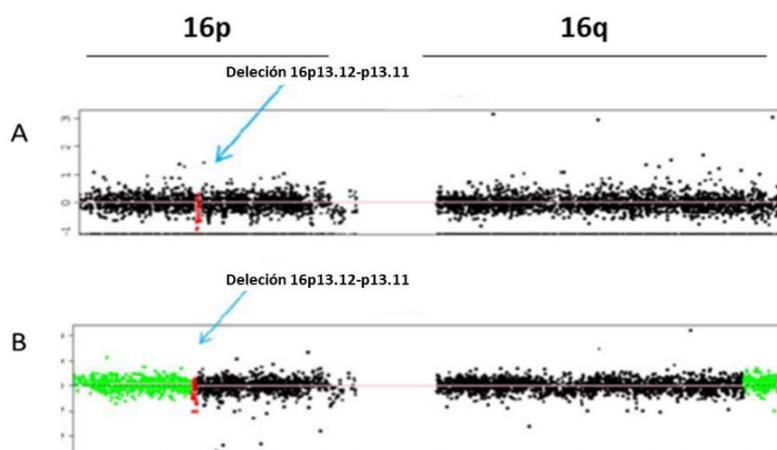


Figura 25. Perfil cromosómico del cromosoma 16 de dos tumores incluidos en el estudio. A) Este tumor solo presenta la deleción de la región 16p13.12-p13.11. B) En este caso, el tumor analizado presenta la deleción de la región 16p13.12-p13.11 junto con otras alteraciones cromosómicas. Las relaciones logarítmicas normalizadas de las intensidades de la sonda se representan como puntos de diferente color. Las ganancias de material genómico se muestran en verde, las pérdidas se muestran en rojo y los números de copia normales se muestran en negro. Se señala con flecha azul la deleción recurrente encontrada (Perea *et al.* 2017) (Anexo 11).

La deleción estaba presente en 34 muestras, lo que supone el 23,3% de los tumores incluidos en el estudio. Llamó la atención que la deleción era más frecuente en los EO CRC que en los LO CRC. En el grupo de los jóvenes se encontró en 20 de los 60 tumores analizados (33%) mientras que en el grupo de los ancianos la presentaban 14 de los 86 tumores (16,3%).

1.2. HIBRIDACIÓN *IN SITU* POR FLUORESCENCIA DE LA REGIÓN 16p13.12-p13.11

El siguiente objetivo fue confirmar esta alteración cromosómica observada en la región 16p13.12-p13.11. Para ello, se llevó a cabo un análisis de hibridación *in situ* por fluorescencia (FISH) en un total de 5 muestras que presentaban la deleción por la técnica de aCGH. En todos los casos, el análisis mediante FISH confirmó los resultados obtenidos por aCGH. Es decir, en las 5 muestras analizadas, había pérdida de material genético entre las bases 14738223 y 15353060 con un tamaño de 614838 pb (http://www.ensembl.org/Homo_Sapiens/Gene/Sequence) (Figura 26) (Perea *et al.* 2017) (Anexo 11).

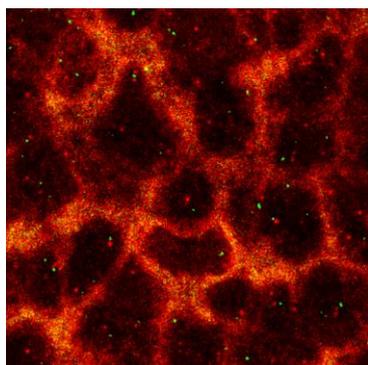


Figura 26. Confirmación mediante FISH de la deleción de la región 16p13.12-p13.11. En la técnica FISH, la hibridación se llevó a cabo con la sonda 354N7 (16q) (color rojo) y con la sonda 2504F3 (16p) (color verde) (Perea *et al.* 2017) (Anexo 11).

Definida la región alterada, se llevó a cabo la búsqueda de los genes localizados en ella a través de la base genome.ucsc.edu. Esta zona alberga un total de 9 genes: *BFAR*, *PLA2G10*, *NPIPA2*, *ABCC6P2*, *NOMO1*, *PDXDC1*, *NTAN*, *RRN3* y *PKD1P6*, cada uno de estos genes está implicado en diferentes funciones celulares, como ya se describieron en la introducción de este trabajo. El único gene implicado en cáncer hasta el momento era *NOMO1*. La proteína que codifica este gen forma un complejo proteico, junto con otras dos proteínas, que inhiben la actividad de la vía Nodal. Esta vía juega un importante papel en carcinogénesis (Gong *et al.* 2014). Por ello. Empezamos a estudiar el gen *NOMO1* en estos tumores.

1.3. COMPARACIÓN DE LAS CARACTERÍSTICAS CLÍNICO-PATOLÓGICAS, MOLECULARES Y FAMILIARES ENTRE DIFERENTES SUBGRUPOS

1.3.1. GRUPO GLOBAL DE PACIENTES CON PÉRDIDA DE LA REGIÓN 16p13.12-p13.11 vs. GRUPO GLOBAL DE PACIENTES SIN PÉRDIDA DE LA REGIÓN 16p13.12-p13.11

En primer lugar, se compararon las características clínicas, clínico-patológicas y familiares del grupo de pacientes que presentaban la deleción en la región 16p13.12-p13.11 con el grupo de pacientes que no presentaban la deleción sin tener en cuenta ningún criterio de edad (Tabla 20) (Perea *et al.* 2017) (Anexo 11).

Tabla 20. Características clínicas, patológicas y familiares de todos los casos de cáncer de colon (grupo global) sin tener en cuenta la edad (Perea et al. 2017) (Anexo 11). NS: no significativo.

	Pacientes con la delección en el 16p n (%)	Pacientes sin la delección en el 16p n (%)	p (x ²)
Nº de pacientes	34 (23,3)	112 (76,7)	
Sexo:			
hombre	14 (41,2)	68 (60,7)	
mujer	20 (58,8)	44 (39,3)	0,05
Localización:			
colon derecho	12 (35,3)	34 (30,4)	
colon izquierdo	16 (47,1)	31 (27,7)	0.024
recto	6 (17,6)	47 (42)	
Diferenciación tumoral:			
pobre	2/30 (6,7)	5/99 (5,1)	NS
Producción moco	6/30(20)	23/99 (23,2)	NS
Células "anillo de sello"	0	4/99 (4)	NS
Estadio TNM:			
I	6(17,6)	13 (11,6)	
II	18 (52,9)	48 (42,9)	
III	4 (11,8)	29 (25,9)	NS
IV	6 (17,6)	22 (19,6)	
Pólipos asociados	25 (73,5)	66 (58,9)	NS
Media número de pólipos	2,3 (2,8)	2,8 (6,3)	NS
Tipo:			
Adenomatoso	11 (44)	39 (59)	
Hiperplásico	3 (22)	9 (14)	NS
Mezcla	11 (44)	18 (27)	
CCRs sincrónicos o metacrónicos	4 (11,8)	19 (17)	NS
Recurrencia	2 (7,7)	15 (16,7)	NS
Mortalidad	6 (17,6)	36 (32,1)	NS
Supervivencia libre de enfermedad	55,8 (46,4)	32,9 (31,9)	0,02
Supervivencia total	61,3 (43,7)	38,7 (30,3)	0,15
MIS	4 (11,8)	12 (10,7)	NS
Mutaciones en genes MMR	3 (8,8)	5 (4,5)	NS
CIMP-ALTO	9 (26,5)	22 (19,6)	NS
CNA	175	110	0,01
Historia familiar de cáncer			
Familias positivas Amsterdam II	4 (11,8)	7 (6,3)	
Neoplasia con agregación de Lynch	13 (38,2)	27 (24,1)	NS
Neoplasia sin agregación de Lynch	8 (23,5)	22 (19,6)	
Casos esporádicos	17 (50)	75 (67)	

Considerando el grupo total de pacientes que presentaban la delección en el brazo corto del cromosoma 16 frente a los que no la presentaban, sin agruparlos según la edad de aparición del tumor, se puede observar que la alteración es más frecuente en mujeres que en hombres (58,8% vs. 41,2%). Además, en el grupo de tumores con la alteración, se observa una baja tasa de incidencia en el recto (17,6%) y apenas hay diferencias entre la localización del tumor en el colon derecho o en el izquierdo (35,3% vs. 47,1%). Sin embargo, en el grupo que no presenta la alteración predominan los tumores localizados en el recto (42%) frente a las otras localizaciones (derecho 30,4%, izquierdo 27,7%).

Por otra parte, no se encontraron diferencias significativas entre los dos grupos respecto a la diferenciación tumoral o a la producción de moco. Para ambos, el mayor porcentaje de tumores fueron diagnosticados en etapas no muy avanzadas de la enfermedad, concretamente en estadio II (52,9% con alteración y 42,9% sin alteración). La media de pólipos encontrados por tumor fue prácticamente la misma tanto para el grupo que presentaba la delección como para el grupo que no la presentaba, predominando los pólipos adenomatosos frente a los hiperplásicos (44% vs. 22% grupo con la alteración, 59% vs. 14% grupo sin la alteración).

Además, el grupo de pacientes con la delección presentaba mejor pronóstico que el grupo sin la delección. Las curvas de Kaplan-Meier que se muestran en la figura 27 confirman el mejor pronóstico para el grupo con la delección. De hecho, la supervivencia libre de enfermedad en este grupo era del 46,4% y la del grupo sin la alteración apenas sobrepasaba el 32% ($p=0,02$).

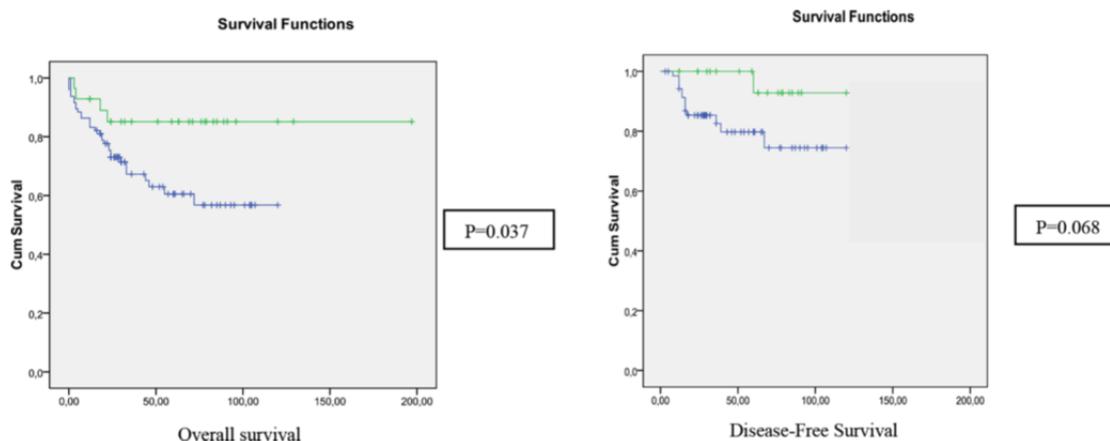


Figura 27. Diagrama de Kaplan-Meier de la supervivencia global y la supervivencia libre de enfermedad de acuerdo con el estado de la región 16p. En color verde se representa el grupo de pacientes con la alteración y en azul aquellos que no la presentan. En ambos casos, la supervivencia del grupo de pacientes con la delección cromosómica es superior a la de los pacientes que no presentan esta delección (Perea *et al.* 2017) (Anexo 11).

Al comparar la incidencia de tumores MIS no se observaron diferencias estadísticamente significativas entre ambos grupos (11,8% de tumores con la delección mostraron MIS vs. 10,7% de tumores sin la delección presentaron MIS). Tampoco se observaron

diferencias con respecto a los tumores que presentaban mutaciones en los genes MMR en ambos grupos (8,8% vs. 4,5%) ni con respecto a los tumores que presentaban CIMP alto (26,5% vs.19,6%). Sin embargo, en el grupo con la deleción en el brazo corto del cromosoma 16, había mayor número de CNA que en el grupo que no portaba la deleción (175vs.110 respectivamente, $p=0,01$).

Finalmente, los datos muestran que no se encontraron diferencias estadísticamente significativas entre ambos grupos al comparar la historia familiar de cada paciente ni tampoco en el número de tumores esporádicos encontrados en cada grupo (50% de los tumores esporádicos presentaban la alteración vs. 67% de los tumores esporádicos que no presentaban dicha deleción).

1.3.2. PACIENTES JÓVENES CON PÉRDIDA DE LA REGIÓN 16p13.12-p13.11 VS. PACIENTES JÓVENES SIN PÉRDIDA DE LA REGIÓN 16p13.12-p13.11

Posteriormente, se compararon las mismas características clínicas, clínico-patológicas, moleculares y familiares entre el grupo de pacientes que presentaban la deleción en el brazo corto del cromosoma 16 y el grupo de pacientes que no la presentaban, pero, teniendo en cuenta únicamente a los pacientes diagnosticados a una edad igual o inferior a los 45 años (Tabla 21) (Perea *et al.* 2017) (Anexo 11).

Tabla 21. Características clínicas, patológicas y familiares de los pacientes diagnosticados a una edad igual o inferior a los 45 años, con y sin la alteración en el brazo corto del cromosoma 16 (Perea et al. 2017) (Anexo 11). NS: no significativo.

	Pacientes jóvenes con la deleción en el 16p n (%)	Pacientes jóvenes sin la deleción en el 16p n (%)	p (x ²)
Nº de pacientes	20 (33,3)	40 (66,7)	
Sexo:			
hombre	9 (45)	27 (67,5)	
mujer	11(55)	13 (32,5)	NS
Localización:			
colon derecho	7 (35)	6 (15)	
colon izquierdo	13 (65)	15 (37,5)	0,001
recto	0	19 (47,5)	
Diferenciación tumoral:			
pobre	2/16 (12,5)	1/33 (3)	NS
Producción moco	5/16 (31,3)	9/33 (27,3)	NS
Células "anillo de sello"	0	2/33 (6,1)	NS
Estadio TNM:			
I	4 (20)	11 (27.5)	
II	11 (55)	12 (30)	
III	1 (5)	9 (22.5)	
IV	4 (20)	8 (20)	NS
Pólipos asociados	13 (65)	22 (55)	NS
Media número de pólipos	2.2 (2,6)	2,8 (8)	NS
Tipo:			
Adenomatoso	5 (38,5)	9 (40,9)	
Hiperplásico	2 (15,4)	5 (22,7)	NS
Mezcla	6 (46,1)	8 (36,4)	
CCRs sincrónicos o metacrónicos	3 (15)	2 (5)	NS
Recurrencia	2 (12,5)	5 (15,6)	NS
Mortalidad	3 (15)	11 (27,5)	NS
Supervivencia libre de enfermedad	65,4 (47)	51,7 (36,5)	0,2
Supervivencia total	72 (42,8)	61,1 (29,8)	0,3
MIS	3 (15)	6 (15)	NS
Mutaciones en genes MMR	3 (15)	4 (10)	NS
CIMP-ALTO	6 (30)	6 (15)	NS
CNA	160	63	0,004
Historia familiar de cáncer			
Familias positivas Amsterdam II	4 (20)	7 (17,5)	
Neoplasia con agregación de Lynch	10 (50)	19 (47,5)	NS
Neoplasia sin agregación de Lynch	8 (40)	14 (36)	
Casos esporádicos	32 (46,4)	17 (42,5)	

En el caso del grupo de pacientes jóvenes y a diferencia de lo que sucede cuando se hace la comparación global, al comparar las características clínico-patológicas del grupo de tumores que presentaban la delección en la región cromosómica 16p13.12-p13.11 (33,3%) con el grupo de tumores que no mostraban la delección (66,7%), no se encontraron diferencias respecto al sexo. Dentro del grupo con la alteración el 45% de los tumores fueron diagnosticados en hombres y el 55% en mujeres mientras que, en el grupo de tumores sin la alteración, el 67,5% de ellos fueron diagnosticados en varones y el 32,5% en mujeres.

Sin embargo, al igual que ocurría en el grupo global, se encontraron diferencias estadísticamente significativas respecto a la localización del tumor. En el grupo con la delección, llamó la atención que la mayoría de los tumores (65%) se localizaban en el colon izquierdo mientras que la tasa de tumores en el recto era nula. Por el contrario, en el grupo sin la alteración, predominaba la tasa de tumores en el recto (47,5%) frente a la localización en el colon derecho o izquierdo (15% y 37,5% respectivamente).

Tampoco se encontraron diferencias entre ambos grupos respecto a la diferenciación tumoral y a la producción de moco. En este caso, en el grupo que presentaba la delección, la mayoría de los tumores fueron diagnosticados en las primeras etapas de la enfermedad (estadio II), sin embargo, en el grupo de tumores que no tenían la delección, el reparto en los diferentes estadios en el momento del diagnóstico fue más homogéneo.

En el caso del grupo de pacientes jóvenes, el grupo de pacientes con la delección también presentaba mejor pronóstico que el grupo sin la delección. La supervivencia libre de enfermedad en el grupo con la delección era del 47% y la del grupo sin la alteración apenas sobrepasaba el 36%. La supervivencia total era del 43% para el grupo con la delección y del 30% para el grupo que no la presentaba.

Al igual que lo que sucedía al hacer la comparación global, al comparar la incidencia de tumores MIS no se observaron diferencias estadísticamente significativas entre ambos grupos (ambos grupos presentaron el 15% de tumores con MIS). Tampoco se observaron diferencias con respecto a los tumores que presentaban mutaciones en los genes MMR en ambos grupos (15% de tumores presentaban mutaciones dentro del grupo de pacientes con la delección vs. 10% de tumores presentaban mutaciones dentro del grupo de pacientes sin la delección) ni con respecto a los tumores que presentaban CIMP alto (30% vs.15%). Sin embargo, en el grupo con la delección en el brazo corto del cromosoma 16, había mayor número CNA que en el grupo que no portaba la delección (160 vs. 63 respectivamente, $p=0,004$).

El porcentaje de tumores asociados a una historia familiar de CCR fue similar en ambos subgrupos (20% tumores con la deleción vs. 17,5% tumores sin la deleción) al igual que el porcentaje de tumores esporádicos (46,4 vs. 42,5 respectivamente).

1.3.3. ANÁLISIS DE LA LONGITUD TELOMÉRICA EN LOS TUMORES DE LOS DIFERENTES SUBGRUPOS DE PACIENTES

La longitud de los telómeros se ha relacionado con diversos estados patológicos incluida la tumorigénesis. Sin estar claro si la modificación de la longitud telomérica es un evento que colabora con la aparición del tumor o si es tan solo una consecuencia del proceso, ha sido vinculada a diversos tipos de cánceres humanos (Wang *et al.* 2002; Martinez-Delgado *et al.* 2012). Por ello, en este trabajo nos propusimos estudiar la longitud telomérica en los tumores previamente analizados mediante la técnica de aCGH y ver si existen diferencias entre los subgrupos que se establecieron en función de si presentaban o no la deleción en el brazo corto del cromosoma 16 y en función de la edad en el momento del diagnóstico.

De las 146 muestras iniciales, solo pudimos incluir en el estudio de la longitud telomérica 91 ya que del resto no se disponía de la cantidad suficiente de DNA que requiere la técnica. De los 91 tumores analizados, 46 se correspondían con tumores diagnosticados en pacientes ancianos: 35 no presentaban la deleción y 11 de ellos si presentaban la alteración de la región 16p13.12-p13.11. El resto, los otros 45, se correspondían con tumores diagnosticados en individuos jóvenes, 28 de ellos no mostraban la alteración cromosómica y 17 si la presentaban.

Los datos de la media de los tres valores de C_T obtenidos para las curvas de amplificación tanto de *TEL1* (telómero) como de *36B4* (control endógeno) para cada tumor se muestran en el anexo 1. Además, en esta tabla, también se especifica el estado de la región 16p13.12-p13.1 para cada muestra incluida en el estudio.

En la Figura 28 se comparan los valores obtenidos para la media de las longitudes teloméricas de los tumores pertenecientes a cada subgrupo establecido según el estado de la región 16p y de la edad en el momento del diagnóstico.

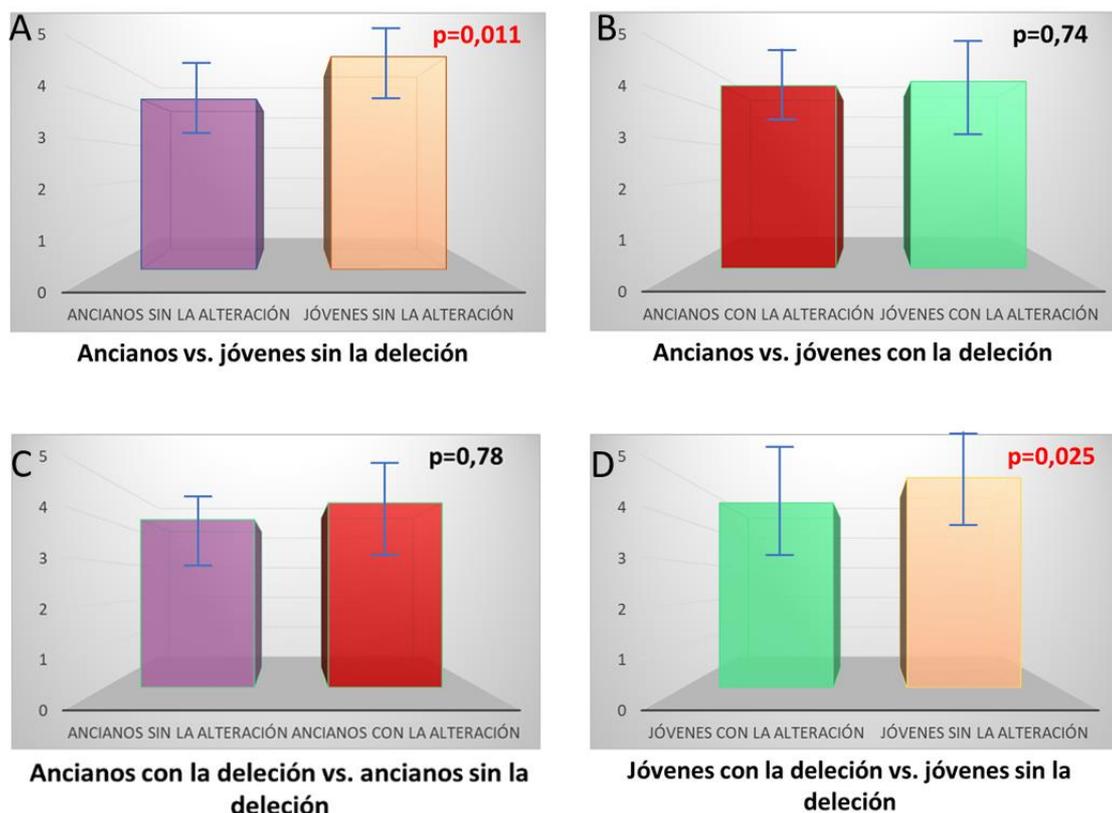


Figura 28. Representación gráfica de la longitud telomérica media de los tumores según la edad de diagnóstico y el estado de la región 16p13.12-p13.1. A-Representación gráfica de la longitud telomérica media de los tumores diagnosticados en ancianos vs. jóvenes sin la alteración. B- Representación gráfica de la longitud telomérica media de los tumores diagnosticados en ancianos vs. jóvenes con la alteración. C- Representación gráfica de la longitud telomérica media de los tumores diagnosticados en ancianos con la alteración vs. ancianos sin la alteración. D- Representación gráfica de la longitud telomérica media de los tumores diagnosticados en jóvenes con la alteración vs. jóvenes sin la alteración.

Como se puede observar en la Figura 28A, al comparar la media de la longitud telomérica de los tumores pertenecientes al grupo de **ancianos sin la alteración** frente al grupo de tumores correspondientes a **jóvenes sin la alteración** en la región 16p, la longitud media de los telómeros del grupo de ancianos fue significativamente menor que la correspondiente al grupo de jóvenes ($p=0,011$) (Figura 28A).

Sin embargo, al comparar la media de la longitud telomérica de ambos grupos incluyendo solo los que presentaban la alteración, se observa que no hay diferencias estadísticamente significativas entre los tumores pertenecientes al grupo de los ancianos y los tumores de aparición precoz; es decir, en este caso, la longitud telomérica es similar en ancianos que en jóvenes (Figura 28B).

Posteriormente, se comparó la media de la longitud telomérica de todos los tumores diagnosticados en pacientes con edad igual o superior a los 70 años, con y sin la alteración en la

región 16p13.12-p13.11. No se observaron resultados estadísticamente significativos entre los dos subgrupos (Figura 28C).

Finalmente, se comparó la media de la longitud telomérica del grupo de tumores diagnosticados en pacientes con edad igual o inferior a los 45 años con y sin la delección en la región 16p13.12-p13.11. En este caso, la media de la longitud telomérica del grupo de tumores que presentaban la alteración en el brazo corto del cromosoma 16 era significativamente inferior a la media de la longitud telomérica del grupo de tumores que no mostraban la alteración cromosómica (Figura 28D).

1.4. ESTUDIO DEL GEN *NOMO1* EN LA REGIÓN 16p13.12-p13.11

1.4.1. EXISTENCIA DE UN ÚNICO GEN *NOMO*

Inicialmente, debido al importante papel que desempeña la vía Nodal en carcinogénesis, y, puesto que el gen *NOMO1* forma un complejo proteico que actúa como antagonista de dicha vía, se estudió el estado de este gen en muestras que presentaban la delección de la región 16p13.12-p13.11 mediante la técnica de aCGH.

En el análisis de los resultados obtenidos en la q-PCR, se observó que, en la muestra utilizada como control, la cantidad de DNA amplificada para el fragmento de *NOMO1* era igual que la cantidad de DNA amplificada para el fragmento del gen endógeno *LEMD3*. Sin embargo, el resultado que se esperaba era que la cantidad de DNA amplificada para *NOMO1* fuera el triple que la cantidad amplificada para el control endógeno ya que el fragmento del gen *NOMO1* que se amplificaba guardaba una homología del 100% en la secuencia de nucleótidos para los tres genes *NOMO1*, *NOMO2* y *NOMO3* (Figura 29A) y, el control endógeno *LEMD3*, es un gen de copia única.

Con el fin de confirmar la existencia de un solo gen *NOMO* se amplificó otro fragmento de este gen correspondiente a parte del exón 1. Este fragmento también guarda una homología del 100% en la región codificante de los tres genes (Figura 29B). Los resultados muestran que, en los dos fragmentos analizados para *NOMO*, la cantidad de DNA amplificada era la misma que para el gen de copia única *LEMD3* (Figura 30). Lo cual sugería la existencia de un gen de copia única.

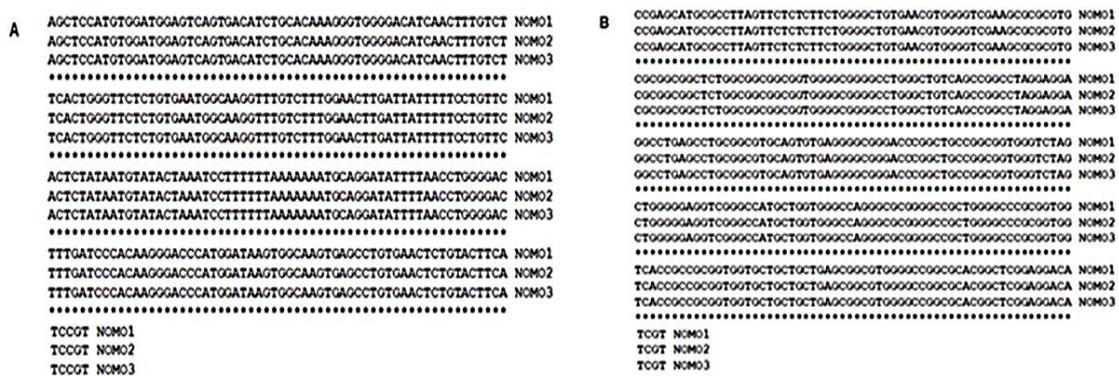


Figura 29. Alineamiento de la secuencia de nucleótidos amplificada para dos fragmentos diferentes de los genes *NOMO1*, *NOMO2* y *NOMO3*. A) Amplificación de parte del exón y del intrón 4. B) Amplificación de un fragmento del exón 1. El alineamiento muestra una homología del 100%.

Como este gen se localiza dentro de una región de duplicaciones, quisimos verificar la validez de la técnica amplificando diferentes controles:

- Amplificamos parte del exón 2 del gen *RBFOX1*, localizado dentro del brazo corto del cromosoma 16 pero fuera de la región de duplicaciones dónde se encuentra *NOMO*. La cantidad de DNA amplificada para este fragmento del gen *RBFOX1* es igual que la cantidad de DNA amplificada para los dos fragmentos de *NOMO* y para el gen de copia única *LEMD3* (Figura 30).

- Como control de región genómica duplicada, amplificamos el exón 13 del gen *PKD1* que tiene un pseudogen, por tanto, no es un gen de copia única ya que este exón, guarda una homología del 100% con la secuencia de nucleótidos de su pseudogen y, como consecuencia, el genoma tiene cuatro copias de este fragmento. Este gen se localiza dentro de la región de duplicaciones donde se encuentra *NOMO*. A su vez, también amplificamos parte del exón 30 de *PKD1* que no tiene pseudogen y es de copia única. Los resultados obtenidos en este caso mostraron que la cantidad de DNA amplificada para el exón 30 del gen *PKD1* era la misma que la cantidad de DNA amplificada para los genes de copia única *NOMO*, *LEMD3* y *RBFOX1*, sin embargo, la cantidad amplificada para el exón 13 de este gen era prácticamente el doble que para los genes de copia única debido a la existencia del pseudogen (Figura 30).

- Finalmente, como control de gen presente en un solo alelo, amplificamos el exón 5 del gen *STS*. Este gen está localizado en el cromosoma X por lo que en varones solo existe una única copia en el cromosoma X. En este caso, amplificamos DNA procedente de un varón, por lo que la cantidad de DNA amplificada es la mitad respecto a la cantidad de DNA amplificada en los genes de copia única (Figura 30).

Las barras de normalización correspondientes a la cantidad de DNA amplificado para cada fragmento se muestran en la Figura 30.

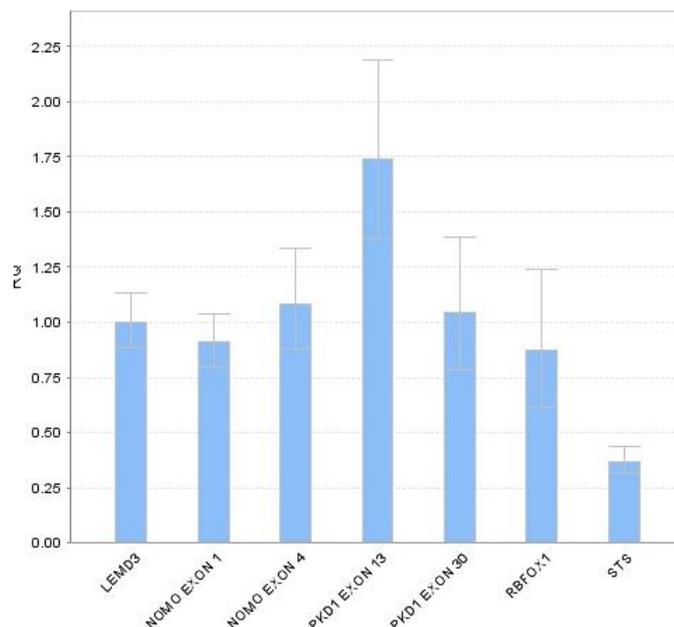


Figura 30. Representación gráfica de las barras de normalización correspondientes a la cantidad de DNA amplificada para cada fragmento de diferentes genes incluidos en el estudio. La cantidad amplificada de DNA es la misma para los tres genes de copia única (*LEMD3*, *RBFOX1* y para el exón 30 de *PKD1*) que para los dos fragmentos diferentes amplificados para el gen *NOMO1*. La cantidad de DNA amplificada para el gen *STS* es la mitad que para los genes de copia única y, la cantidad de DNA para el fragmento correspondiente al exón 13 de *PKD1*, al contener un pseudogen, es el doble a la de los genes de copia única.

Si se compara la región codificante de los tres genes, *NOMO1*, *NOMO2* y *NOMO3*, solo se diferencian en 20 nucleótidos. De estos 20 cambios, 13 codifican el mismo aminoácido, es decir, son cambios silenciosos y 7 dan lugar a cambio de aminoácido (mutaciones de cambio de sentido (mutaciones *missense*) (Tabla 22). Tan solo una de estas mutaciones *missense* (c.26C>T/p.(L9P)) no está incluida en la base de datos dSNP (<https://www.ncbi.nlm.nih.gov/snp>). Además, también se recogimos datos de las frecuencias alélicas de estas variantes en diferentes poblaciones (<http://exac.broadinstitute.org>). El análisis de esta información muestra que el alelo *wild type* del gen *NOMO1* tiene una frecuencia próxima al 100% en la mayoría de las poblaciones. Solo existe una variante *missense* en la que la frecuencia poblacional para el alelo *wild type* del gen *NOMO1* es inferior al 80%, pero, aun así, es el alelo más frecuente (c.1374 A>C; A:0,67 - C:0,33).

Tabla 22. Frecuencia alélica de los nucleótidos que difieren en la secuencia nucleotídica de los tres genes (*NOMO1*, *NOMO2* y *NOMO3*).

GVSc	NOMO			Rs	Población ID	Sujetos	Frec. ale	
	1	2	3				>	<
c.1468 A>G p.(N490D)	A	G	G	062412	AMR	694	A=0.943	G=0.056
					AFR	1322	A=0.976	G=0.023
					EAS	1008	A=0.995	G=0.004
					SAS	978	A=0.919	G=0.080
					EUR	1006	A=0.886	G=0.113
c.1210 A>G p.(I404V)	A	G	G	062412	ExAc	121400	A=0.999	G=0.001
c.1738 A>G p.(M580V)	A	G	G	1510032	ExAc	119700	A=0.989	G=0.010
c.1477 A>G p.(M493V)	A	G	G	1418607	ExAc	121206	A=0.999	G=0.001
c.1238 A>G p.(Q413R)	A	G	A	453341				
c.1374 A>C p.(K458N)	A	A	C	379603	CEPH	184	A=0.670	C=0.330
c.26AC>T p.(L9P)	C	T	C	No rs				
c.1185G>A p.(P395P)	G	G	A	1785532	ExAc	121402	G=0.999	A=0.001
c.3318C>T p.(D1106D)	C	T	T	2001756	ExAc	121406	C=0.999	T=0.001
c.696C>G p.(N232N)	C	G	C	2019541				
c.3216A>G p.(T1072T)	A	G	G	7563367	ExAc	121310	A=0.999	G=0.001
c.2694C>T p.(S898S)	C	T	T	7715801	ExAc	121402	C=0.999	T=0.001
c.2187C>T p.(G729G)	C	T	T	7796274	ExAc	121396	C=0.999	T=0.001
c.1260C>G p.(P420P)	C	G	C	370942	AMR	694	C=1.000	T=0.000
					AFR	1322	C=1.000	T=0.000
					EAS	1008	C=1.000	T=0.000
					SAS	978	C=1.000	T=0.000
					EUR	1006	C=0,999	T=0.001
c.2211C>T p.(P737P)	C	C	T	1132626	ExAc	814	T=0.814	C=0.185
c.156T>G p.(S52S)	T	G	T	No rs				
c.2388T>C p.(H796H)	T	C	C	No rs				
c.2586G>A p.(A862A)	G	A	A	No rs				
c.3583C>G p.(R1195R)	C	G	G	No rs				
c.3666T>Gp.(T1222T)	T	G	G	No rs				

1.4.2. ESTUDIO DEL GEN *NOMO* EN CCR

Inicialmente, se seleccionaron 5 muestras que presentaban la delección de la región 16p13.12-13.11 mediante la técnica de aCGH para analizar, mediante qPCR, el estado del gen *NOMO* (en todos los casos a los que nos referiremos a partir de este punto, analizamos el fragmento correspondiente a parte del exón y del intrón 4). Los resultados de la q-PCR mostraron que, de todas ellas, una presentaba pérdida del gen en uno de los alelos (delección heterocigota), tres presentaban pérdida en los dos alelos (delección homocigota) y tan solo una de ellas no tenía pérdida del gen *NOMO* (Figura 31).

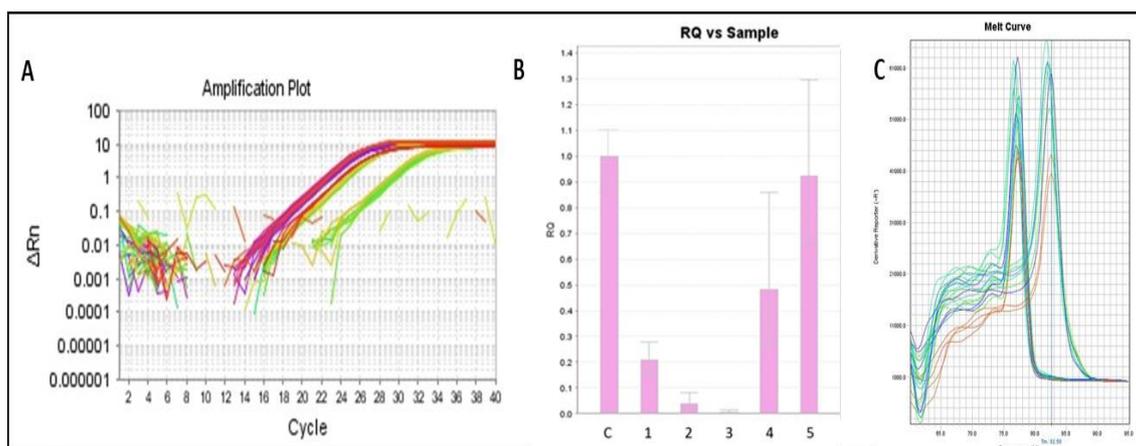


Figura 31. q-PCR correspondiente al análisis de un fragmento de *NOMO* en los 5 tumores. A) Curvas de amplificación del gen *NOMO* y del gen utilizado como control endógeno *LEMD3*. B) Diagrama correspondiente a las barras de normalización de la cantidad de DNA amplificado del fragmento de *NOMO* para cada uno de los 5 tumores: las muestras 1, 2 y 3 presentan delección homocigota de *NOMO*, la muestra 4 delección heterocigota y, en la 5 amplifica la misma cantidad de DNA que la muestra utilizada como control (C) por tanto, no presenta pérdida del gen en ninguno de los alelos. C) Melt curves correspondientes a los dos fragmentos amplificados *NOMO* y *LEMD3*.

Seguidamente, analizamos el estado del gen *NOMO* en todos los tumores que previamente habían sido incluidos en el análisis de aCGH. De los 146 tumores iniciales solo se pudieron estudiar 51 puesto que de alguno de ellos no se disponía de la cantidad suficiente de muestra requerida por la técnica.

Los 20 tumores de pacientes jóvenes que presentaban delección de la región 16p13.12-p13.11 por la técnica de aCGH, mostraban delección homocigota del gen (Figura 32). (Los valores normalizados de la cantidad de DNA amplificado para cada tumor se recogen en el anexo 2).

Del grupo de los 14 tumores diagnosticados en pacientes con edades avanzadas que presentaban la delección, se pudo analizar el estado del gen *NOMO* en 9 de ellos. De todos ellos, solo dos presentaban delección homocigota del gen, cinco tenían pérdida en uno de los alelos y dos tumores, no mostraban ninguna anomalía (Figura 32). (Los valores normalizados de la cantidad de DNA amplificado para cada tumor se recogen en el anexo 2).

A continuación, se analizó el estado del gen *NOMO* en los tumores que no presentaban la delección en la región 16p por la técnica aCGH. De los 14 tumores analizados correspondientes al grupo de pacientes jóvenes, todos ellos presentaban delección homocigota de *NOMO*. De los 8 tumores que se analizaron pertenecientes al grupo de pacientes ancianos sin la delección, ninguno de ellos presentaba anomalías en el número de copias del gen *NOMO* (Figura 32). (Los valores normalizados de la cantidad de DNA amplificado para cada tumor se recogen en el anexo 2).

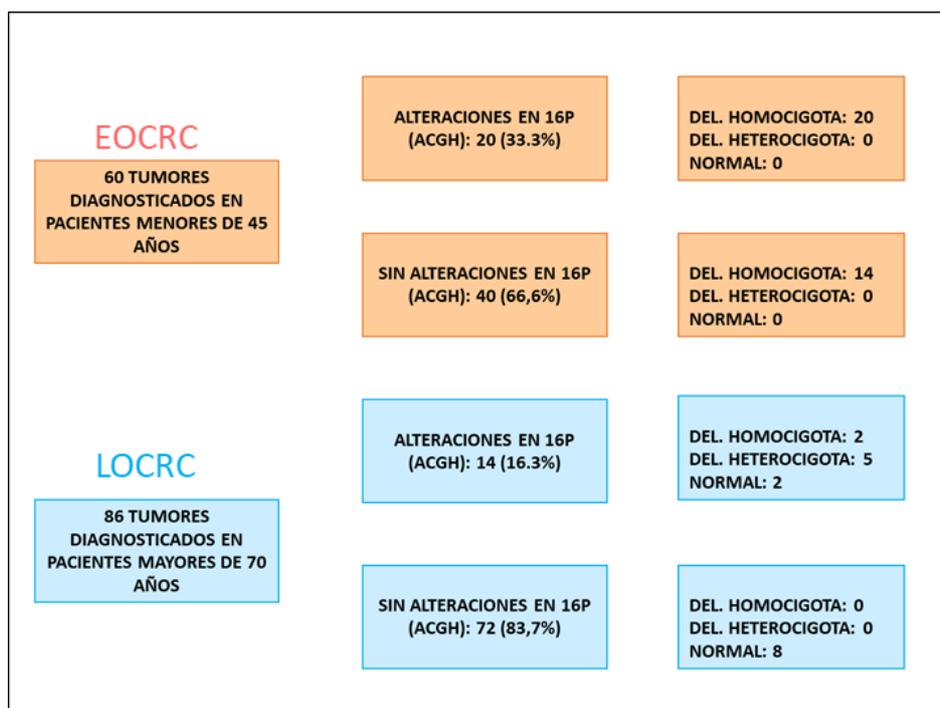


Figura 32. Estado del gen *NOMO* en los tumores EOCRC y LOCRC dependiendo del estado de la región 16p13.12-p13.11 (Perea *et al.* 2017) (Anexo 11).

En resumen, el 100% de los tumores diagnosticados en el grupo de los jóvenes presentan delección homocigota del gen *NOMO*, mientras que los tumores LOCRC solo 2 de los 17 estudiados (11,7%) tenían delección homocigota del gen, 5 (29,4%) mostraron pérdida en uno de los alelos y el resto (58,9%) no presentó ninguna anomalía en cuanto al número de copias en dicho gen (Perea *et al.* 2017) (Anexo 11).

Posteriormente, con el fin de confirmar la alta tasa de tumores con pérdida de *NOMO* detectados en esta serie preliminar, se aumentó el número de muestras a un total de 95 casos de CCR de inicio precoz. De todos ellos, 68 (71,5%) presentan delección homocigota del gen, 13 (13,7%) tienen delección heterocigota y 14 (14,8%) no presentan ninguna alteración en el número de copias para este gen (Tabla 23). (Los valores normalizados de la cantidad de DNA amplificado para cada tumor se recogen en el anexo 3).

También se estudiaron 50 muestras más de casos de tumores diagnosticados a edades avanzadas. De todas ellas, solo se detectó una (2%) con pérdida homocigota de *NOMO*, 4 (8%) presentaban delección heterocigota y la mayoría de los tumores (90%) no presentaban alteración en el número de copias de este gen. Los valores normalizados de la cantidad de DNA amplificado para cada tumor se recogen en el anexo 4. En conjunto, de los 67 tumores pertenecientes al grupo LOCRC que se han estudiado, 3 (4,5%) presentaron delección homocigota, 9 (13,4%)

delección heterocigota y 55 (82,1%) no mostraron ninguna anomalía en el número de copias del gen *NOMO* (Tabla 23) (Perea et al.2017) (Anexo 11).

Además, también quisimos ver el estado del gen *NOMO* en tumores procedentes de pacientes con edades intermedias, es decir, edades comprendidas entre los 46 y los 69 años, siendo la edad más frecuente del grupo los 64 años y la media 65,7 años. Analizamos 67 tumores, de los cuales 4 de ellos (6%) presentaron delección homocigota, 8 (12%) delección heterocigota y 55 (82%) mostraban el gen *NOMO* germinal (Tabla 23). (Los valores normalizados de la cantidad de DNA amplificado para cada tumor se recogen en el anexo 5) (Perea et al.2017) (Anexo 11).

Tabla 23. Resumen del estado del gen *NOMO* en todos los tumores analizados en función de la edad de diagnóstico.

	Delección homocigota <i>NOMO</i>	Delección heterocigota <i>NOMO</i>	<i>NOMO</i> sin alteraciones en el nº de copias
GRUPO JÓVEN n(%)	68 (71,5)	13 (13,7)	14 (14,8)
GRUPO INTERMEDIO n(%)	4 (6)	8(12)	55 (82)
GRUPO MAYOR n(%)	3 (4,5)	9 (13,5)	55 (82)

Posteriormente, clasificamos los tumores pertenecientes al grupo EOCRC en dos grupos: aquellos que tenían inestabilidad de microsatélite (MIS) y aquellos que eran estables (MSS). Los tumores MSS-EOCRC presentaban una alta tasa de delección homocigota de *NOMO* (54 de 59; 91,5%) mientras que, en los tumores MIS-EOCRC, la delección solo estaba presente en 7 de los 16 analizados (43,74%). La pérdida heterocigota del gen era poco frecuente en tumores MSS-EOCRC (2 de 59; 3,3%), al contrario que los MIS-EOCRC (5 de 16; 31,2%). En resumen, la mayoría de los tumores MSS-EOCRC (92%) presentaban delección homocigota de *NOMO* mientras que en el grupo MIS-EOCRC la tasa de tumores con delección homocigota disminuyó hasta un 44% (Tabla 24) (Perea et al.2017) (Anexo 11).

Tabla 24. Estado de *NOMO* en tumores EOCRC en función de si los tumores son MSS o MIS. Dentro del grupo inicial se incluyeron los 34 tumores de la serie preliminar previamente estudiados por la técnica aCGH. El grupo añadido comprende los 41 tumores añadidos posteriormente con el fin de confirmar los resultados preliminares. El grupo global hace referencia a la suma de los dos anteriores.

		Deleción homocigota	Deleción heterocigota	<i>NOMO</i> sin alteraciones
		<i>NOMO</i>	<i>NOMO</i>	en el nº de copias
GRUPO INICIAL	MSS n(%)	31 (100)	0	0
	MIS n(%)	3 (100)	0	0
GRUPO AÑADIDO	MSS n(%)	23 (82)	2 (7)	3 (11)
	MIS n(%)	4(31)	5(38)	4 (31)
GRUPO GLOBAL	MSS n(%)	54 (92)	2 (3)	3 (5)
	MIS n(%)	7 (44)	5 (31)	4 (25)

1.4.3. LA DELECIÓN DE *NOMO* ES SOMÁTICA

Para estudiar si la pérdida de *NOMO* en uno de los alelos era a nivel germinal y la otra pérdida a nivel somático, se realizó el estudio de este gen en 13 muestras procedentes de sangre periférica de pacientes a los que previamente se les había realizado el estudio en el tumor y presentaban deleción homocigota del gen. Todos estos pacientes pertenecían al grupo EOCRC. Los resultados muestran que el gen *NOMO* está germinal en todos los casos (Figura 33). Es decir, la deleción de este gen es somática. (Los valores normalizados de la cantidad de DNA amplificado para cada muestra de sangre periférica se recogen en el anexo 6) (Perea et al.2017) (Anexo 11).

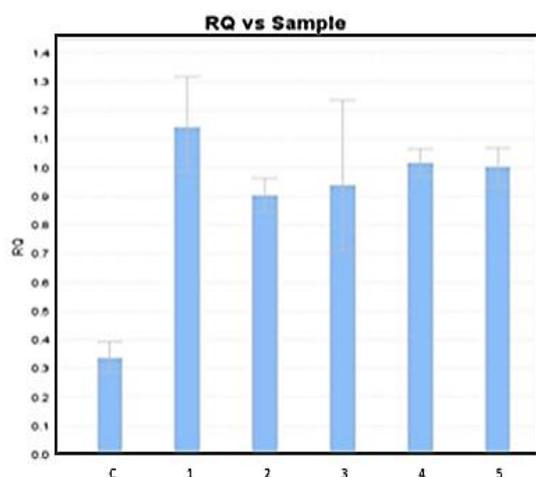


Figura 33. Diagrama correspondiente a las barras de normalización de la cantidad de DNA amplificado del fragmento de *NOMO* para 5 de las muestras de sangre periférica incluidas en el estudio. La muestra C corresponde a uno de los tumores con pérdida de *NOMO*.

1.4.4. ESTADO DEL GEN *NOMO* EN PÓLIPOS ADENOMATOSOS

Analizamos el número de copias del gen *NOMO* en pólipos adenomatosos para ver si la pérdida del gen se producía en los primeros estadios de la carcinogénesis colorrectal. Hemos estudiado un total de 16 pólipos adenomatosos procedentes de pacientes diagnosticados antes de los 45 años. De todos ellos, 7 (43,75%) presentan delección heterocigota del gen y en el resto de los pólipos analizados, 9 (56,25%) el gen está germinal. (Los valores normalizados de la cantidad de DNA amplificado para cada pólipo se recogen en el anexo 7) (Perea et al.2017) (Anexo 11).

1.4.5. ESTADO DEL GEN *NOMO* EN LA REGIÓN PERITUMORAL

También analizamos el estado de *NOMO* en la región peritumoral, es decir en la región sana más próxima al tumor. Se consiguió obtener 5 muestras de región peritumoral procedentes de pacientes a los que previamente se les había realizado el estudio de *NOMO* en el tumor y todos ellos presentaban delección homocigota. De las 5 muestras, solo se pudo analizar el número de copias del gen en tres de ellas ya que de las otras dos no se consiguió extraer DNA con la calidad que requiere la técnica, por lo que no se pudo amplificar. Los resultados de la qPCR de las tres muestras que se analizaron determinan que en todas ellas había pérdida homocigota del gen (Figura 34). (Los valores normalizados de la cantidad de DNA amplificado para cada muestra se recogen en el anexo 8).

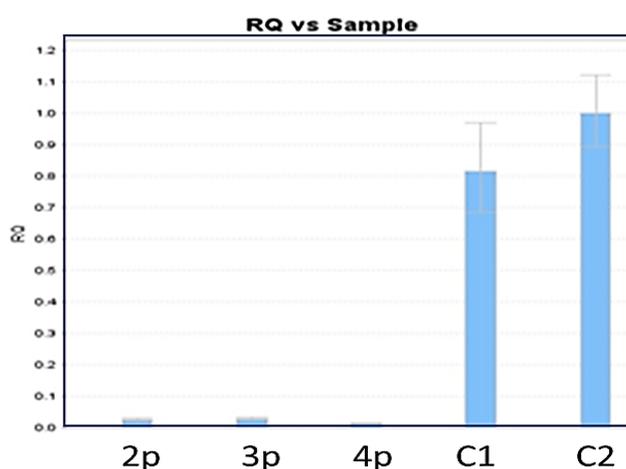


Figura 34. Diagrama correspondiente a las barras de normalización de la cantidad de DNA amplificada del fragmento de *NOMO* para tres de las muestras de DNA extraído de región peritumoral. Las muestras 1P, 2P y 3P se corresponden con las tres muestras de región peritumoral analizadas y las muestras Control 1 y Control 2 son DNA control.

1.4.6. ESTUDIO DEL GEN *NOMO* EN OTROS TUMORES

Con el fin de determinar si la pérdida de *NOMO* era exclusiva del CCR, sobre todo del CCR diagnosticado en el adulto joven, ampliamos nuestro estudio extendiéndolo a otros tipos de tumores.

-Cáncer de mama: se analizó el estado del gen *NOMO* en 22 tumores de mama, todos ellos diagnosticados en pacientes con edades iguales o inferiores a los 45 años. De los casos estudiados, uno de ellos (4,5%) presentaba delección homocigota del gen y tres (13,6%) tenían pérdida heterocigota del gen (Tabla 25). El resto, 18 tumores (82%) presentaban el gen germinal.

-Cáncer de ovario: se estudiaron 9 tumores de ovario diagnosticados en mujeres menores de 45 años. Únicamente uno de estos tumores (11,1%) presentaba delección heterocigota del gen. En el resto, el gen *NOMO* estaba germinal (Tabla 25).

-Cáncer de endometrio: estudiamos el gen *NOMO* en un total de 10 tumores de endometrio, todos ellos procedentes de pacientes diagnosticados a edades iguales o inferiores a los 45 años. Ninguno de ellos presentaba cambios en el número de copias en este gen (Tabla 25).

-Glioblastomas: analizamos el estado del gen *NOMO* en un total de 15 glioblastomas. En este caso, seleccionamos muestras de pacientes diagnosticados a edades iguales o inferiores a los 50 años puesto que si consideráramos como edad de corte los 45 años, el número de tumores se reducía a una cifra muy pequeña. Los resultados obtenidos muestran que la cantidad de DNA amplificada para todos los glioblastomas era la misma que para el gen utilizado como control endógeno excepto para uno de ellos que presentaba delección heterocigota (6,7%) (Tabla 25).

-Tumores de cabeza y cuello: finalmente, también estudiamos el estado de *NOMO* en 13 muestras procedentes del grupo de tumores de cabeza y cuello, en pacientes con edades inferiores a los 50 años. La mayoría de ellos se localizaban en la laringe. De todos ellos, dos (15,3%) mostraban delección heterocigota del gen y ninguno tenía pérdida completa de *NOMO* (Tabla 25). El resto no presentó ninguna anomalía.

(Los valores normalizados de la cantidad de DNA amplificado para cada muestra se recogen en el anexo 9).

Tabla 25. Estado del gen *NOMO* en diferentes tipos de tumores.

	<i>NOMO</i> delección homocigota n(%)	<i>NOMO</i> delección heterocigota n(%)	<i>NOMO</i> sin alteraciones en el nº de copias n(5)
Cáncer de mama	1 (4,5)	3 (13,5)	18 (81,8)
Cáncer de ovario	0	1 (11)	8 (89)
Cáncer de endometrio	0	0	10 (100)
Glioblastomas	0	1 (6,6)	14 (93,4)
Tumores de cabeza y cuello	0	2 (15,3)	11 (84,7)

2. ESTUDIO DEL PAPEL DE *NOMO* EN LA VÍA NODAL

Para estudiar el efecto que tiene la inactivación del gen *NOMO* en la expresión de las proteínas integrantes de la vía Nodal, comprobamos mediante q-PCR el estado del gen en las dos líneas celulares de cáncer de colon de las que disponíamos: HT29 y HCT116 (Figura 35). La cantidad de DNA amplificada en las dos líneas celulares del fragmento del gen *NOMO*, fue la misma que la cantidad amplificada para la muestra control. Ninguna de las dos líneas celulares presentaba delección del gen.

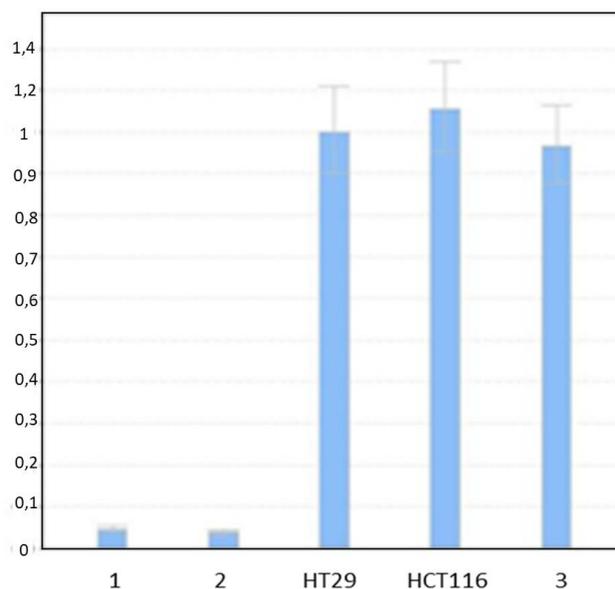


Figura 35. Barras de normalización de la cantidad de DNA amplificada del fragmento de *NOMO* para las líneas celulares HT29 y HCT116. Las muestras 1 y 2 se corresponden con controles para delección homocigota del gen y la muestra 3 se corresponde con un control donde no hay anomalías en el gen estudiado.

Dado que ninguna de las dos líneas celulares tenía delección del gen, se decidió inactivar *NOMO* en la línea HT29, por tratarse de una línea celular procedente de un paciente con CCR menor de 45 años, mediante la técnica de edición génica CRISPR/Cas9.

2.1. GENERACIÓN DE UNA LÍNEA CELULAR *KNOCKOUT* PARA *NOMO* MEDIANTE LA TÉCNICA DE EDICIÓN GÉNICA CRISPR/CAS9

2.1.1. SCREENING DE COLONIAS POR SECUENCIACIÓN DEL PLÁSMIDO PX458

El plásmido seleccionado para el clonaje de los diferentes sgRNAs fue el plásmido PX458 el cual contiene el promotor U6 (Figura 36). En primer lugar, los sgRNAs se ligaron a un sgRNA *scaffold* presente en este plásmido que se encuentra bajo el control del promotor U6. *Up-stream* de esta secuencia, hay una diana de restricción para la enzima *BbsI* que nos sirvió para integrar las secuencias de los sgRNAs en el plásmido.

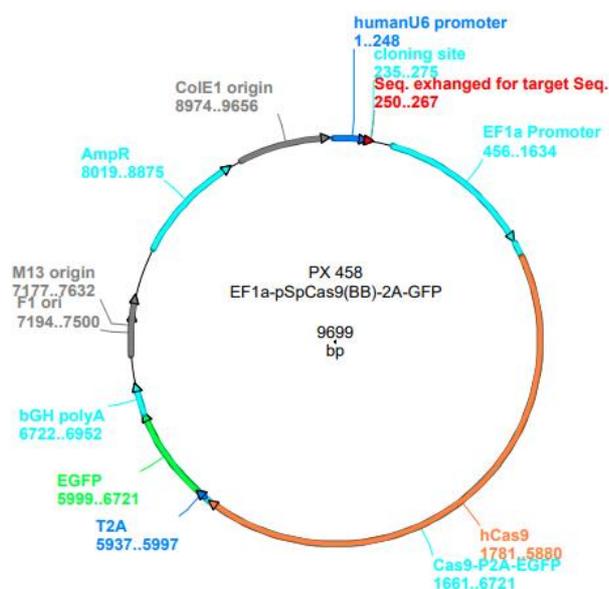


Figura 36. Esquema del plásmido pSpCas9(BB)-2A-GFP (PX458). El sgRNA se inserta en el "cloning site" precedido por el promotor U6 (www.addgene.org).

Transformadas las bacterias DH5 α con los tres plásmidos diferentes, se dejaron crecer las colonias durante toda la noche y, al día siguiente, se sembraron tres colonias de cada construcción. Con el fin de poder seleccionar las colonias de *E. coli* DH5 α que contienen el plásmido con el sgRNA correspondiente, se extrajo DNA de cada una de ellas y se secuenció el fragmento correspondiente al final de la región del promotor U6 seguido por él. De las tres

colonias que se seleccionaron para cada vector, todas ellas insertaron correctamente el sgRNA al plásmido (Figura 37).

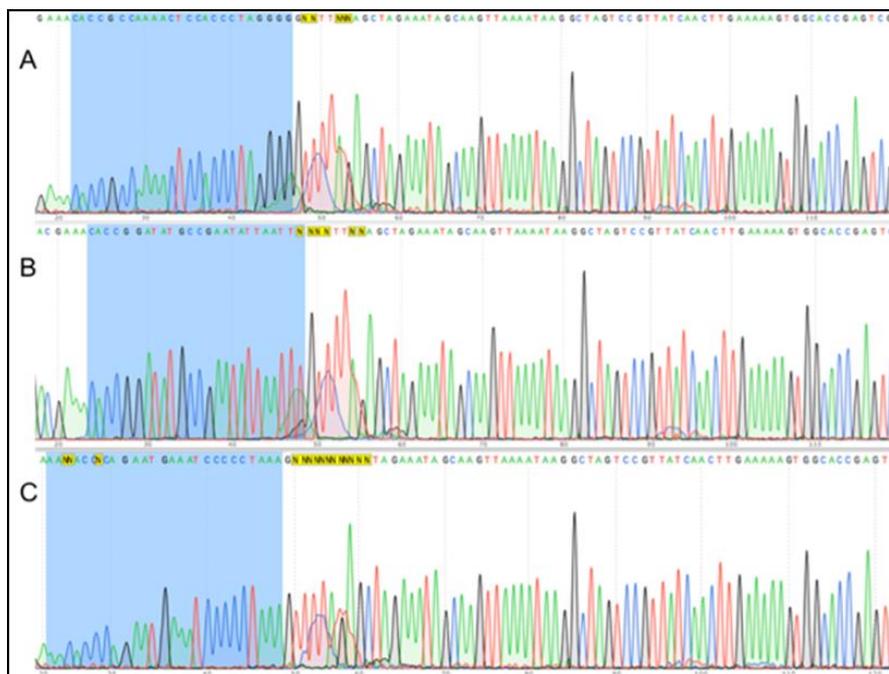


Figura 37. Secuencias de nucleótidos de cada sgRNA insertado en el plásmido PX458. Toda la secuencia es idéntica en los tres casos excepto en los nucleótidos correspondientes a los sgRNAs donde cada uno tiene el que ha insertado (marcado en azul). A) En azul está marcada la secuencia de nucleótidos correspondiente al sgRNA1 (guía 1). El resto de la secuencia se corresponde con un fragmento de la secuencia del plásmido PX458. B) En azul está marcada la secuencia de nucleótidos correspondiente al sgRNA2 (guía 2). El resto de la secuencia se corresponde con un fragmento del plásmido PX458. C) En azul está marcada la secuencia de nucleótidos correspondiente al sgRNA3 (guía 3). El resto de la secuencia se corresponde con un fragmento del plásmido PX458.

2.1.2. TRANSFECCIÓN DE LA LÍNEA CELULAR HT29

Confirmada la presencia de colonias transfectadas para cada construcción, se procedió a la extracción del DNA de cada uno de los tres cultivos bacterianos.

Inicialmente, la transfección se llevó a cabo mediante lipofección utilizando una lipofectamina. Este reactivo causa la fusión de liposomas con la membrana plasmática de células vivas, permitiendo el paso del DNA al interior celular. Seleccionamos 4 placas con una confluencia de células de aproximadamente un 80% y en cada una de ellas se llevó a cabo la transfección con las siguientes combinaciones:

- PLACA 1:** plásmido PX458 sin ningún sgRNA (PX458 vacío).
- PLACA 2:** plásmido con sgRNA 1 + plásmido con sgRNA 2 (G1+G2).
- PLACA 3:** plásmido con sgRNA 1 + plásmido con sgRNA 3 (G1+G3).
- PLACA 4:** plásmido con sgRNA 2 + plásmido con sgRNA 3 (G2+G3).

Transcurridas 48 horas tras la transfección, el número de células que emitían luz verde y que, por tanto, tenían el plásmido fue muy pequeño, por lo que decidimos utilizar otra técnica que resultara más eficaz para la transfección en esta línea celular.

Para ello, se recurrió a transfectar mediante electroporación. Este método permite el paso del plásmido al interior celular debido a un aumento de la conductividad eléctrica y de la permeabilidad de la membrana plasmática celular causado por la aplicación externa de un campo eléctrico. Se transfectaron las células con la misma combinación de sgRNA indicada anteriormente, pero en este caso, también sometimos a células HT29 sin transfectar al choque eléctrico para usarlas posteriormente como control. Transcurridas las 48 horas tras la transfección, observamos que teníamos un importante número de células que emitían luz verde para cada combinación (Figura 38).

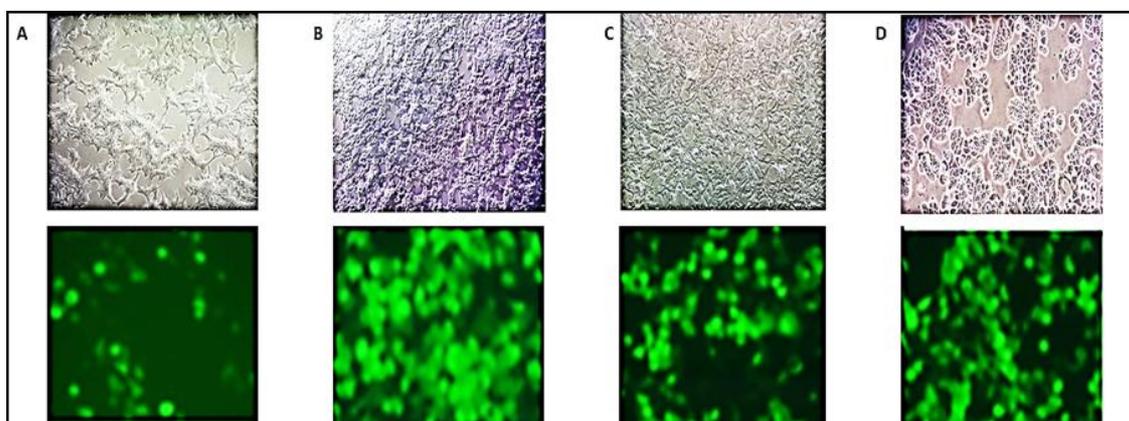


Figura 38. Imágenes en contraste de fases y fluorescencia de células HT29 transfectadas con PX458-sgRNA. Para todos los casos, la imagen superior corresponde a la imagen de las células con luz blanca y la imagen inferior se corresponden con las células vistas con luz verde. A) Células HT29 transfectadas con el plásmido PX458 vacío. B) Células HT29 transfectadas con combinación G1+G2. C) Células HT29 transfectadas con combinación G1+G3. D) Células HT29 transfectadas con combinación G2+G3.

Dado que teníamos un elevado número de células transfectadas, se llevó a cabo la separación de las que habían incorporado la construcción de interés mediante citometría de flujo. Esta técnica consiste en la separación física de las células en base a la emisión de luz verde, ya que el plásmido empleado contiene el gen que codifica la proteína verde fluorescente *EGFP*. De cada placa de células transfectadas con las diferentes combinaciones de vectores obtuvimos cantidades variables de células positivas (Tabla 26). Además, también se separaron células transfectadas con el plásmido vacío (PX458) y células sin transfectar (HT29 normal) que se usaron como controles.

Tabla 26. Informe aportado por el Servicio de Separación Celular del Centro de Investigación del Cáncer de Salamanca. En dicho informe se indica el número de células positivas separadas mediante *single cell* y el número de células positivas separadas en un *pool* de células en eppendorf.

Muestra	Población separada	Número de células totales en eppendorf	Células separadas mediante <i>Single cell</i>
G1+G2	EGFP	783	2 placas de 96 con medio de cultivo
G1+G3	EGFP	915	2 placas de 96 con medio de cultivo
G2+G3	EGFP	6489	2 placas de 96 con medio de cultivo
PX458 VACIO	EGFP	No muestra	2 placas de 96 con medio de cultivo
Células sin transfectar (HT29 Normal)	Células vivas	220.000	2 placas de 96 con medio de cultivo)

Se mantuvieron las 10 placas durante aproximadamente tres semanas comprobando que de todas las células que se separaron mediante *single cell* (960 células), solo crecieron 5 clones. De estos 5 clones, dos se correspondían con la combinación G1 + G2 (finalmente solo creció uno), uno con G2 + G3, otro al plásmido PX458 vacío y otro a las células HT29 sin transfectar. Estos datos indican la baja efectividad de la técnica para la línea celular que estamos utilizando la cual apenas supera el 0,5%. Pasamos cada uno de estos clones a placas de 24 pocillos, después a placas de 12, de 6 y, finalmente, a placas individuales.

2.1.3. GENERACIÓN DE UNA LÍNEA CELULAR *KNOCKOUT* PARA *NOMO*

Confirmación por Sanger

Tras extraer DNA de las células correspondientes a los dos clones con la combinación de guías, al clon HT29 con el plásmido PX458 vacío y al clon HT29 sin transfectar, se amplificó mediante PCR el fragmento correspondiente al exón 3 y parte del intrón 3 donde se encontraban los tres sgRNA. Los productos de PCR, una vez purificados se secuenciaron mediante el método de Sanger para confirmar que, en todos los casos, el sistema CRISP/Cas9 había funcionado correctamente (Figura 38). Al mismo tiempo, también se confirmó que tanto en el clon transfectado con el plásmido PX458 vacío como en el clon HT29 sin transfectar el fragmento analizado del gen *NOMO* permanecía germinal, sin estar afectada su secuencia nucleotídica (Figura 39).

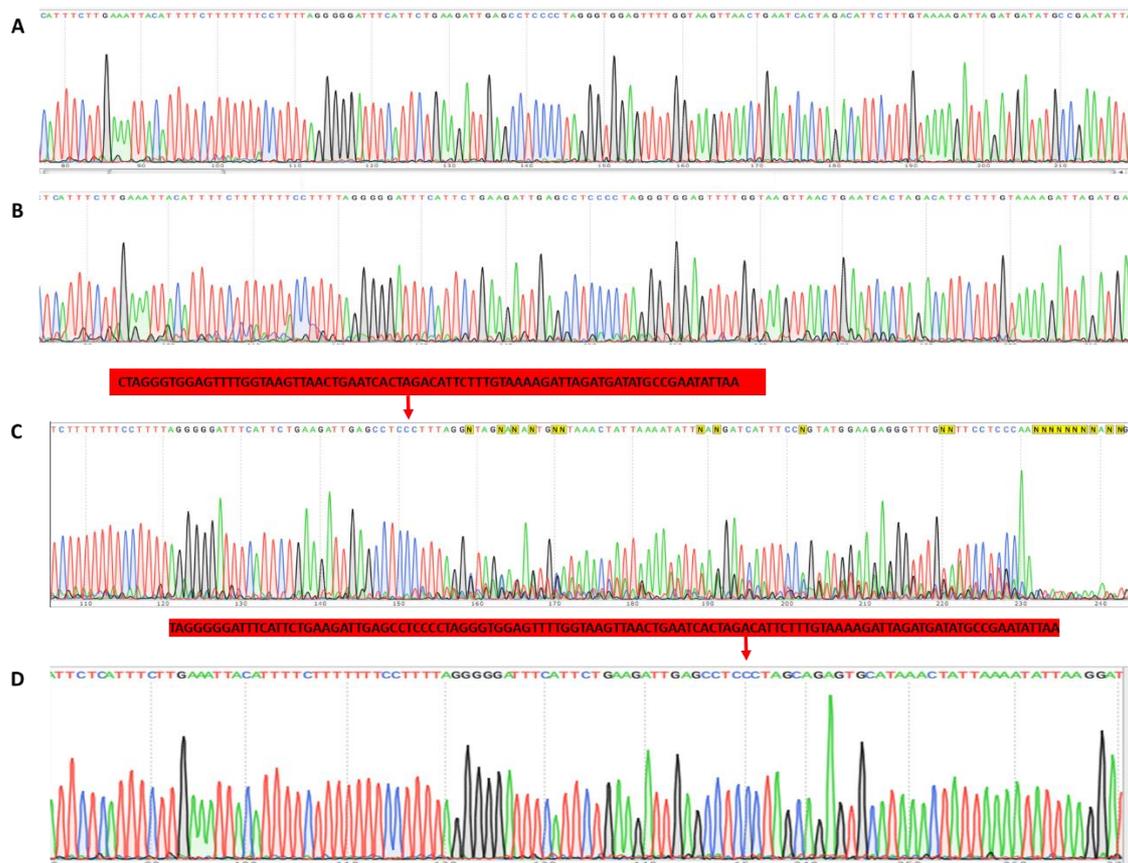


Figura 39. Fragmento de la secuencia de nucleótidos correspondiente a parte del exón 3 y del intrón 3 del gen *NOMO*. A) Secuencia correspondiente al clon HT29 sin transfectar donde la secuencia está sin ningún cambio con respecto a la secuencia de referencia. B) Secuencia correspondiente al clon HT29 transfectado con el plásmido PX458 vacío donde tampoco hay ninguna alteración. C) Secuencia correspondiente al clon G1+G2 (clon1) donde se pierde todo el fragmento de nucleótidos comprendidos entre el sgRNA1 y el sgRNA2 marcado en rojo. D) Secuencia correspondiente al clon G2+G2 (clon 2) donde se pierde todo el fragmento de nucleótidos comprendidos entre el sgRNA2 y el sgRNA3 marcado en rojo.

Confirmación por qPCR

También confirmamos mediante la técnica de PCR cuantitativa a tiempo real que para el fragmento del gen *NOMO* estudiado, en los dos clones *knockout* para *NOMO* no había amplificación de DNA (Figura 39). En este caso, los fragmentos de DNA que se amplificaron se correspondían con la secuencia de DNA que se perdió para cada clon. Es decir, en este caso se utilizará para cada amplificación los oligos correspondientes a los sgRNA que se habían diseñado para el experimento.

Los resultados muestran que no hay amplificación de DNA para el fragmento del exón 3 del gen *NOMO* en los dos clones *knockout* para *NOMO*. De esta manera, se confirmamos que los dos clones obtenidos de la combinación de guías tienen *NOMO* inactivado (Figura 40).

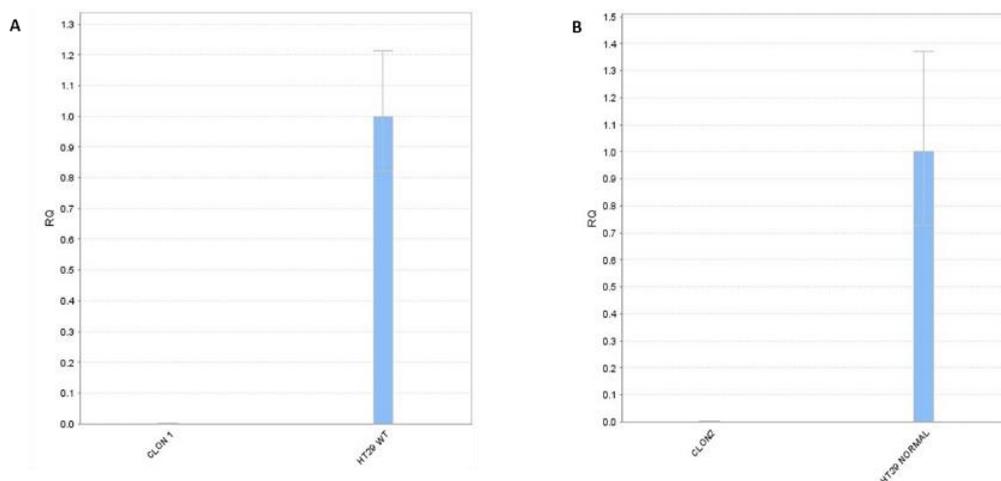


Figura 40. Barras de normalización de la cantidad de DNA amplificada del fragmento de NOMO para los clones obtenidos mediante combinación de guías. A) Clon correspondiente a la combinación sgRNA1+sgRNA2 B) Clon correspondiente a la combinación sgRNA2+sgRNA3. En ningún caso existe amplificación del fragmento de NOMO en los clones.

Confirmación por Western Blot

Por último, también analizamos mediante western blot la expresión de la proteína NOMO. Como se puede ver en la Figura 41, el resultado del Western Blot muestra que en el clon HT29 sin transfectar y en el que se transfectó con el plásmido PX458 vacío había expresión de NOMO mientras que en los clones editados genéticamente no había expresión de NOMO.

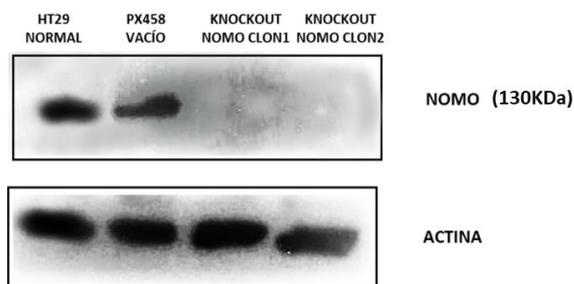


Figura 41. Expresión de la proteína NOMO en el clon HT29 sin transfectar (HT29 NORMAL), en el HT29 transfectado con el plásmido vacío (PX458 VACÍO) y en los dos clones *knockout* para NOMO (KNOCKOUT NOMO). La expresión de las proteínas está calculada en forma de ratio proteína de interés/ β -actina

A través de todas estas técnicas, confirmamos la generación de dos líneas celulares *knockout* para el gen *NOMO* utilizando el sistema de edición génica CRISPR/Cas9 basada en la reparación mediante la unión de extremos no homólogos (NHEJ).

2.2. CARACTERIZACIÓN FENOTÍPICA DE LOS CLONES *KNOCKOUT* PARA *NOMO* OBTENIDOS

Confirmada la generación de dos líneas *knockout* para el gen *NOMO*, analizamos las diferencias fenotípicas que existían entre estos dos clones y los dos utilizados como control (PX458 vacío y HT29 sin transfectar).

2.2.1. ESTUDIO DE VIABILIDAD CELULAR

Se analizó el efecto que la inactivación de *NOMO* ejerce sobre la proliferación celular mediante un ensayo de viabilidad celular colorimétrico basado en la reducción del MTT a los tiempos de 0h, 24h, 48h y 72h (Figura 42). Este ensayo se hizo por duplicado analizando en cada caso tres réplicas de cada clon.

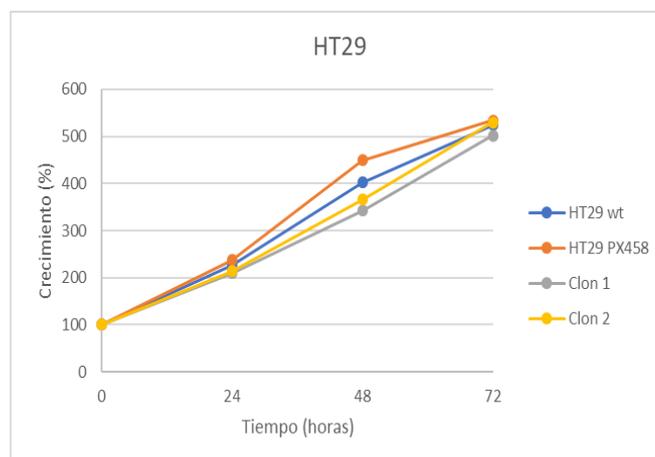


Figura 42. Efecto del del bloque del gen *NOMO* sobre la proliferación celular.

Los resultados obtenidos muestran que el bloque de *NOMO* no afecta a la proliferación celular ya que no se observan diferencias en el crecimiento entre los clones *knockout* para *NOMO* y los dos clones utilizados como controles.

2.2.2. FORMACIÓN DE ESFERAS

Sin embargo, si se apreciaron diferencias en el crecimiento de los clones. Como se puede observar en la Figura 43, los dos clones *knockout* para *NOMO* crecían de manera diferente a los dos clones usados como controles. Mientras que en los clones control las células crecían en monocapa, se aprecia como en los clones *knockout* para *NOMO* las células tendían a agregarse formando *clumps* o esferas. Es decir, se sembraban la misma cantidad de células resuspendidas en 4 placas cada una correspondiente a un clon y, a las 24 horas ya se podían apreciar estas agregaciones en los clones *knockout* para *NOMO*.

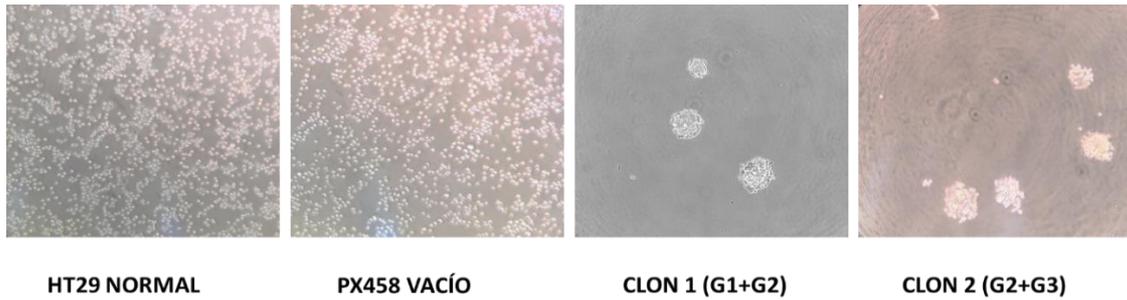


Figura 43. Imagen del crecimiento de las células de los diferentes clones obtenidos. Las células correspondientes al clon1 (G1+G2) y al clon 2 (G2+G3) forman *clumps* o esferas.

Es decir, el bloqueo de *NOMO* conduce a la formación de esferas en esta línea celular. Por ello, también quisimos estudiar la capacidad migratoria de estas células.

2.2.3. ESTUDIO DE LA CAPACIDAD MIGRATORIA CELULAR

Para ello se llevó a cabo un ensayo de migración celular también conocido como cierre de herida. El estudio se hizo con tres réplicas para cada clon y se tomaron fotos cada 10 minutos del estado de cada herida (Figura 44).

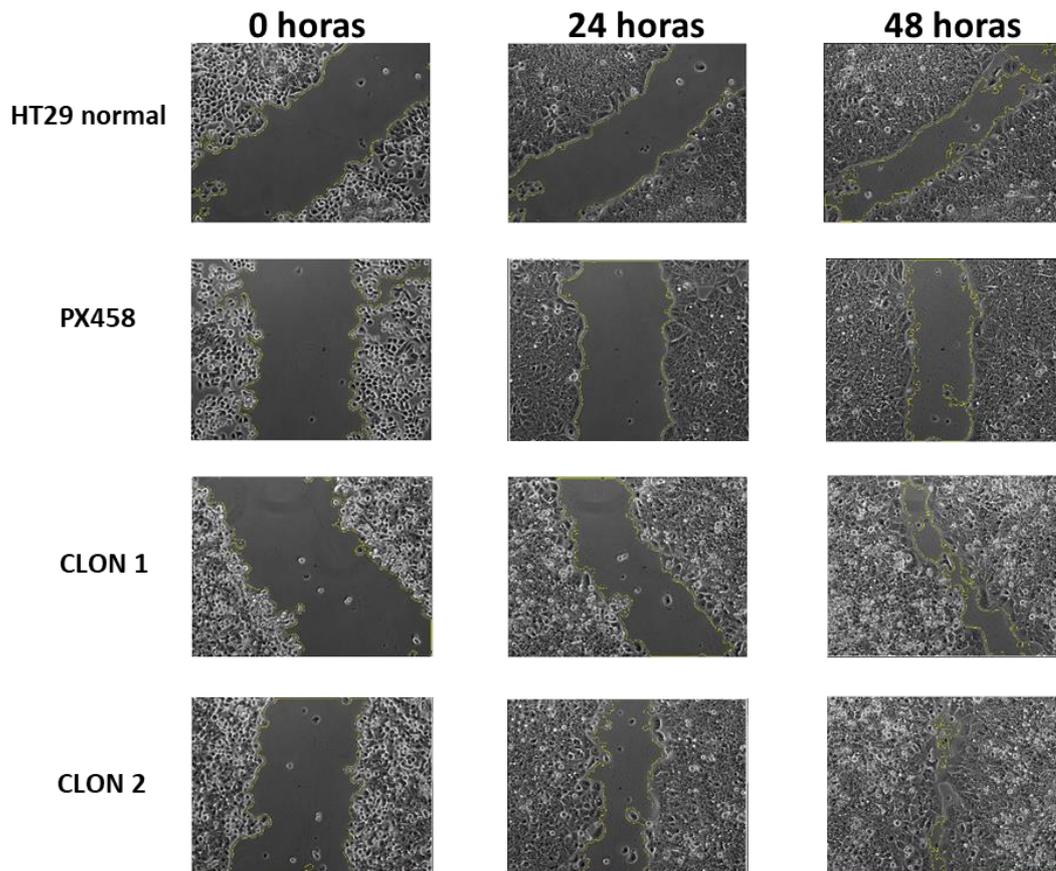


Figura 44. Fotos tomadas a las 0h, 24h y 48h de la migración celular para los dos clones *knockout* para *NOMO* (clon 1 y clon 2) y los dos clones control (HT29 normal y PX458) correspondientes a una de las réplicas. Se observa como a las 48 horas, las heridas en los dos clones *knockout* para *NOMO* están prácticamente cerradas.

Se calculó el porcentaje de área ocupada para cada réplica a las 24h y 48h (Figura 45(A)) y se obtuvo el valor promedio de todos los datos (Figura 45 (B)). Los resultados muestran que el mayor porcentaje de área ocupada transcurridas las 48 horas lo presentaban los dos clones *knockout* para *NOMO* con un promedio de 82,6% para el clon1 correspondiente a los sgRNA1+sgRNA2 y de 88,59% para el clon 2 correspondiente a los sgRNA2+sgRNA3.

Es decir, el bloqueo de *NOMO* conduce a una mayor migración celular (Figura 45(C)).

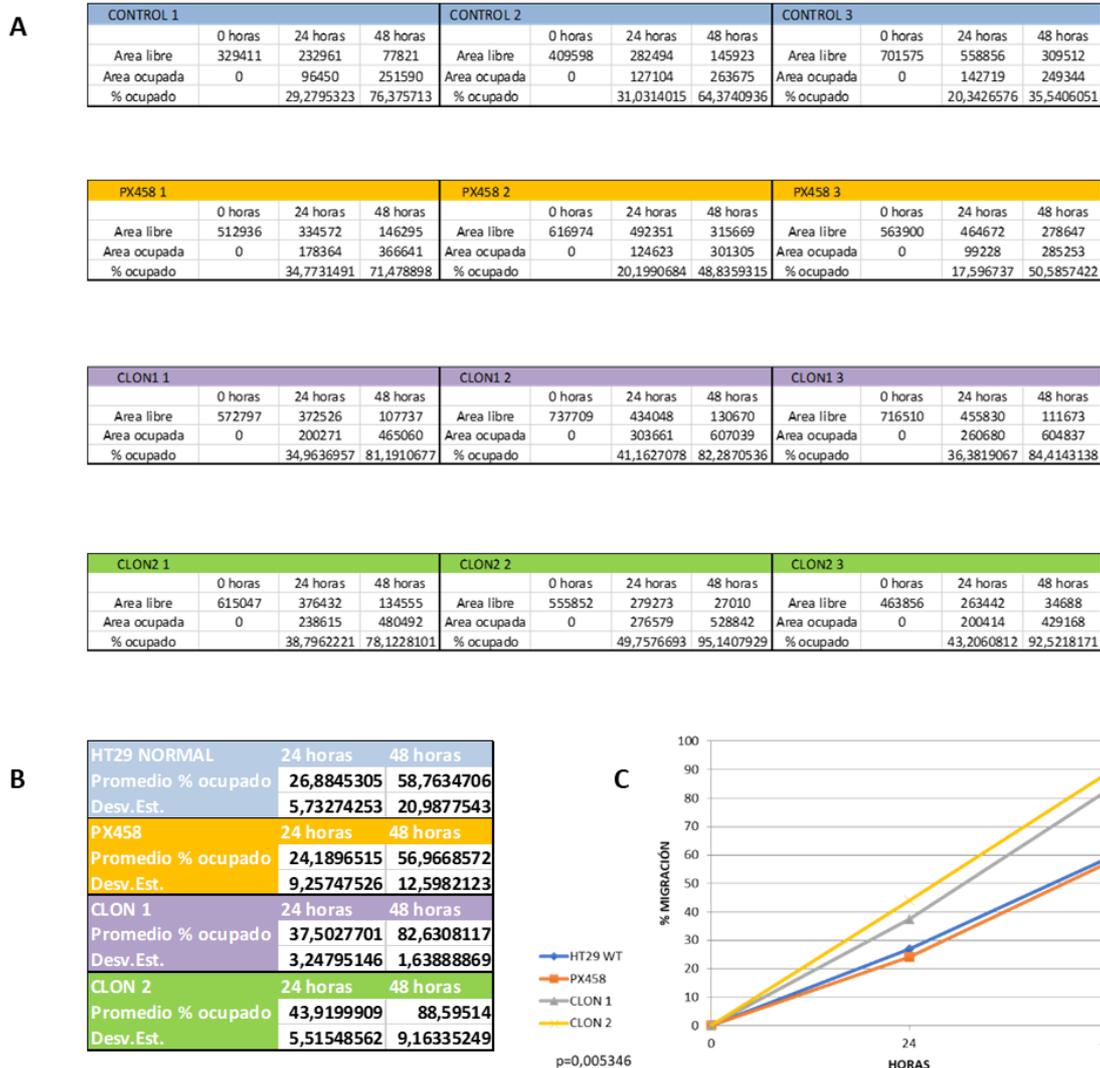


Figura 45. Análisis de los datos de migración. A) Cálculo del porcentaje del área ocupada para cada réplica de cada clon a las 24h y 48h. B) Promedio de los valores de porcentaje de área ocupada para cada clon junto con las desviaciones estándar. C) Representación del porcentaje de migración para cada uno de los 4 clones a las 0h, 24h y 48h.

2.3. ESTUDIO DE ALTERACIONES *OFF-TARGET* EN LA LÍNEA *KNOCKOUT* PARA *NOMO* MEDIANTE ACGH

La revolución del CRISPR-Cas9 nos ha permitido eliminar, reparar y reemplazar partes enteras del genoma con una rapidez y una precisión nunca vistas. Pero, esta tecnología puede introducir cientos de mutaciones no deseadas en el material genético intervenido. Desde el punto de vista teórico CRISPR-Cas9 es una herramienta tremendamente precisa, pero, desde el punto de vista práctico, al usar esta técnica, se han encontrado multitud de mutaciones inesperadas en otros lugares del genoma.

Por ello, nos propusimos ver si existían diferencias a nivel cromosómico entre los clones obtenidos mediante CRISPR-Cas9 y los otros dos clones utilizados como control (HT29 sin transfectar y HT29 transfectada con el plásmido PX458 vacío). Al utilizar esta técnica es fundamental que tengamos en cuenta todos los riesgos potenciales provocados por todas las mutaciones indeseadas causadas por CRISPR, incluyendo mutaciones en las regiones no codificantes y en zonas *off target* (Sharma *et al.* 2018).

Para este estudio utilizamos la plataforma de *microarrays* genómicos CYTOSCAN™, que permite el estudio de alteraciones genómicas en la muestra con una mayor resolución que las técnicas convencionales utilizadas anteriormente como los aCGH.

Se generaron gráficos que señalaban las localizaciones donde se encontraban los cambios. Mediante estas representaciones se identificaron algunas regiones similarmente alteradas en las tres líneas, así como otras diferentes en cada una de ellas (Figura 46).

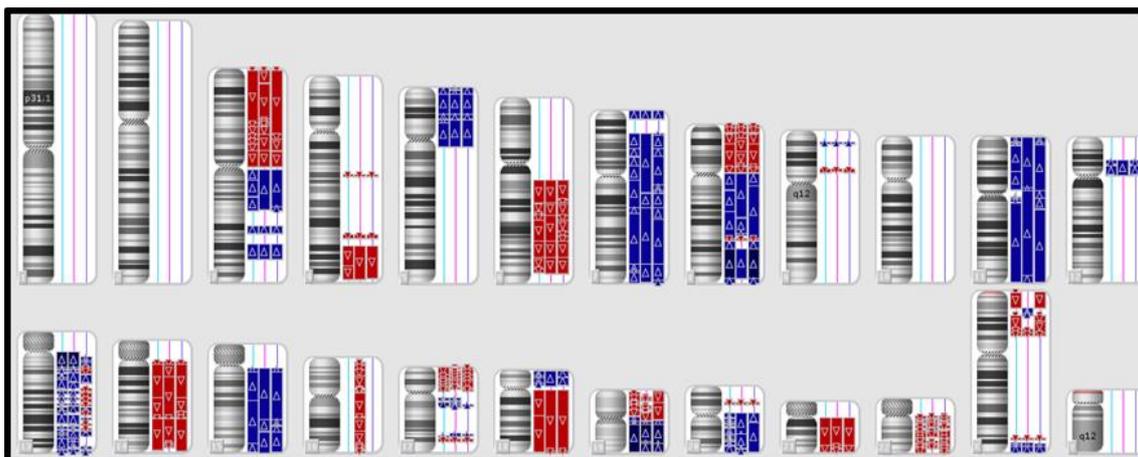


Figura 46. Representación esquemática de las alteraciones en el número de copias detectadas para los tres clones analizados en los diferentes cromosomas. La línea azul claro se corresponde con las alteraciones encontradas en el clon control HT29 normal, la rosa con el clon *knockout* para *NOMO* y la línea azul oscura se corresponde con el clon con el plásmido vacío (PX458). En rojo se representan las pérdidas de material genético y en azul las ganancias en las diferentes regiones cromosómicas.

En las tres muestras analizadas se observó un elevado número de ganancias y pérdidas de material genético en zonas en las que se localizaban genes importantes implicados en tumorigénesis. En la línea sin transfectar (HT29 normal) se identificaron un total de 111 regiones con ganancias de material genético y 83 regiones con pérdidas, en el clon con el plásmido PX458 vacío, 76 regiones con ganancia de material genético y 85 regiones con pérdida. Por último, en el clon *knockout* para *NOMO* se detectaron un total de 81 regiones con ganancia y 95 con pérdida.

Tanto el clon *knockout* para *NOMO* como los dos clones control (PX458 vacío y HT29 normal) compartían una serie de alteraciones, entre ellas: pérdida de material genético en el brazo corto del cromosoma 3, concretamente en la región 3p26-3p12.1, en diferentes regiones del brazo largo del cromosoma 4, 4q22.1, 4q31.22, 4q32.1-4q35.2; en la región 6q14.1-6q26 del cromosoma 6; en el cromosoma 8 hay pérdida de material genético en las regiones 8p23.2, 8p23.1-8p11.1; en la región 9p13 y; también en la región 14q11.2-q32.3. Además, en los tres clones se observó que había pérdida de material genético en dos regiones diferentes del cromosoma 17, concretamente las regiones 17p13.3-17p11.2 y 17q23.3; una región del 18, 18q11.2-q23; otra región del cromosoma 19, 19p13.3-q11; en el 20 también presentaron pérdida de material genético en la región 20p12.1; en diferentes zonas del cromosoma 22, 22q11.1, 22q11.21, 22q11.23, 22q12.2-q13.2 y; también mostraron pérdidas en el cromosoma X, concretamente en las regiones Xp21.1-Xp11.4, Xq27.2-q27.3.

Además de estas deleciones comunes a las tres muestras, también presentaban zonas comunes con ganancias de material genético, entre ellas: tres zonas diferentes del cromosoma 3, 3q11.1-3q22.1, 3q24-3q25.2 y 3q26.1-3q26.33; una región del cromosoma 5, 5p15.33-5q11.22; dos del cromosoma 7, 7p22.3-p22.1 y 7p21.1-q36.2; en el cromosoma 8 también había ganancia de material genético para las tres líneas en las regiones 8p11.1-8q22.3 y 8q23.3-8q24.3 y; en el cromosoma 9 en la región 9p23. Siguiendo con el análisis del perfil cromosómico encontramos más regiones comunes a las tres líneas celulares con ganancias en 11p15.5-11q25; 12p12.1-p11.1; 4 zonas diferentes en el cromosoma 13, 13q13.3-13q14.2, 13q21.1, 13q31.3-q34 y 13q34; la región 15q11.2-q26.3; tres del cromosoma 17, 17p11, 17q11.2-q12 y 17q24.2; las regiones 18p11.32 18p11.21; del cromosoma 19, 19q11-q13.43; la región 20p11.21-q13.33 y; finalmente dentro del cromosoma X también se detectó ganancia de material genético en la región Xq27.3-q28.

También encontramos otras específicas para cada muestra. Es decir, la tecnología de edición génica mediante el sistema CRISPR/Cas9, además de introducir en el genoma de las células las alteraciones de interés, introduce otras muchas no deseadas. Así, por ejemplo, se vio

que el clon modificado genéticamente con la tecnología CRISPR/Cas9 no solo incorporaba la delección de la región 16p13.11 donde se encuentra el gen *NOMO*, sino que presenta otra serie de alteraciones que no existen el clon HT29 sin transfectar. Entre estas diferencias destacan: en la región 8q23.1-8q23.3 y 8q24.3 no hay ganancia de material genético como ocurre en las otras dos líneas. En este clon, se produce una pérdida en heterocigosis prácticamente completa del cromosoma 16 lo cual implica numerosos genes que codifican proteínas con importantes funciones celulares. Además, en el clon HT29 *knockout* para *NOMO* se han encontrado pérdidas de material genético en la región 22q11.23 y ganancias en dos regiones del cromosoma X, Xp22.2-p12.2 y Xp22.11 que, en el clon HT29 sin transfectar no se detectaron.

Es decir, al utilizar esta herramienta de edición génica hay que tener en cuenta todos los cambios *off target* introducidos porque pueden afectar a los posteriores resultados. Los genes que codifican las proteínas que incluimos en nuestro trabajo no se encuentran en estas regiones con una carga genética diferente a los controles por los efectos *off target*.

2.4. ESTUDIO DE LA EXPRESIÓN DE LAS PROTEÍNAS DEL COMPLEJO NCLN/NOMO/TEMEM147

Una vez generada la línea celular *knockout* para *NOMO*, en la que confirmamos con diferentes técnicas la inactivación del gen, estudiamos la expresión de las proteínas con las que forma el complejo proteico.

La proteína *NOMO* forma un complejo proteico de membrana localizado en el RE junto con las proteínas *NCLN* y *TEMEM147*. Apenas se tienen datos sobre la formación de este complejo. Existe un estudio donde se destaca el papel regulador de la proteína *NCLN* la cual, no solo controla los niveles de expresión de *NOMO*, sino también los de *TEMEM147*, lo que corrobora la idea de que la *NCLN* representa el factor limitante en la formación del complejo. Se ha propuesto que este complejo proteico se forma de manera secuencial, inicialmente *NOMO* se une a la proteína *NCLN* a través de su dominio luminal y, posteriormente se une *TEMEM147* a través del dominio transmembrana (Dettmer et al. 2010).

Mediante western blot, analizamos la expresión de las proteínas *NCLN* y *TEMEM147* en los tres clones del apartado anterior: HT29 normal, HT29 transfectada con el plásmido vacío (PX458 vacío) y el clon *knockout* para *NOMO* (Figura 47).

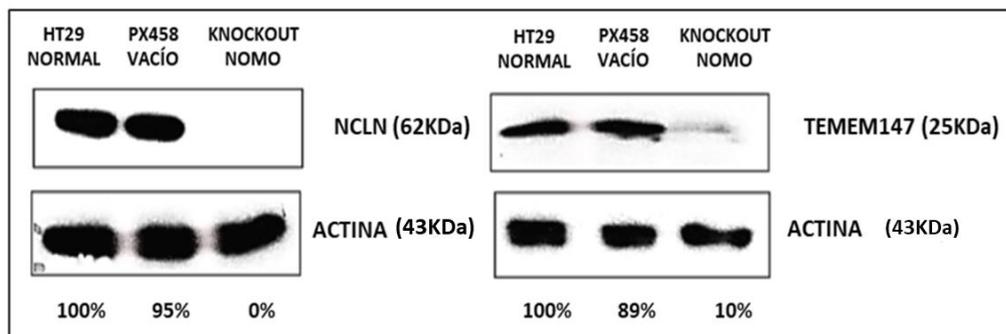


Figura 47. Expresión de las proteínas NCLN y TEMEM147 en los tres clones: HT29 normal (control), HT29 transfectada con el plásmido PX458 vacío (control) y la línea *knockout* para *NOMO*. La expresión de las proteínas está calculada en forma de ratio proteína de interés/ β -actina.

Los resultados muestran que, en los clones utilizados como control en este trabajo, hay la misma expresión de las proteínas NCLN y de TEMEM147 en los dos mientras que, en el clon con *NOMO* inactivado no existe expresión de NCLN y la expresión de TEMEM147 es prácticamente nula.

2.5. ESTUDIO DE LA EXPRESIÓN DE LAS PROTEÍNAS INTEGRANTES DE LA VÍA NODAL

Hay estudios que confirman el papel antagonista que desempeña el complejo proteico NCLN/*NOMO*/TEMEM147 en la vía de señalización Nodal, la cual es fundamental durante las primeras etapas del desarrollo embrionario. Sin embargo, la forma en la que inhibe la actividad de esta vía aún no está establecida (Dettmer *et al.* 2010). Por ello, en este trabajo, quisimos estudiar el efecto que tenía la inhibición de *NOMO* en la expresión de las proteínas integrantes de esta vía de señalización.

En primer lugar, analizamos la expresión de las proteínas que actúan como receptores de la vía. Los ligandos nodales señalan a través del receptor serina-treonina quinasa tipo I, ALK4 (ActRIB / Acvr1b), junto con el receptor tipo II, ActRII (ActRIIA; Acvr2a) o ActRIIB (Acvr2b). Analizamos la expresión de estas dos proteínas en los tres clones con los que estamos trabajando en este estudio (Figura 48).

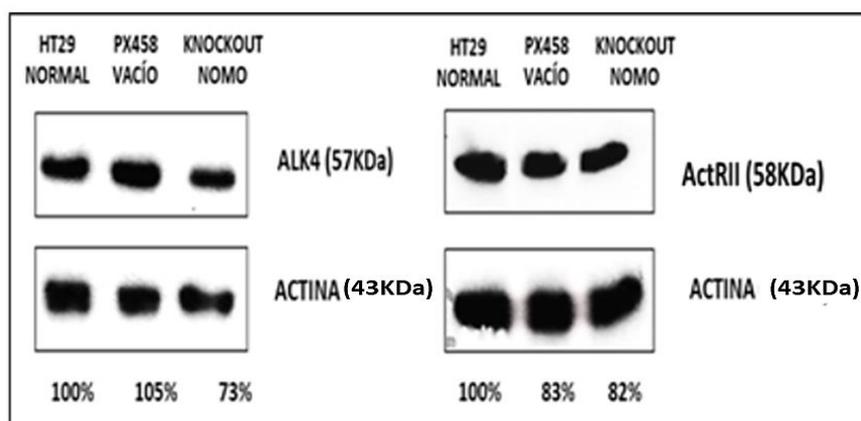


Figura 48. Expresión de las proteínas ALK4 y ActRII en los tres clones: HT29 normal (control), HT29 transfectada con el plásmido PX458 vacío (control) y el clon knockout para *NOMO*. La expresión de las proteínas está calculada en forma de ratio proteína de interés/ β -actina.

Los resultados del western blot muestran que no hay alteraciones en la expresión proteica de ActRII ya que es similar en los tres clones que se están analizando, sin embargo, en el caso de ALK4 si se puede apreciar una ligera reducción de la expresión en el clon con *NOMO* inactivado.

Posteriormente, continuamos analizando la expresión de las proteínas que actúan aguas abajo en la vía de señalización Nodal, para ver si la inactivación de *NOMO* y esta ligera reducción en la expresión del receptor tipo I, ALK4, implicaba alguna otra alteración en la expresión del resto de las proteínas que forman parte de esta vía.

Analizamos la expresión de la proteína CRIPTO en los tres clones. Esta proteína actúa como correceptor esencial para la vía de señalización Nodal que confiere especificidad para el receptor tipo I ALK4 a través de interacciones proteicas. Estos correceptores son de especial importancia ya que los ligandos nodales carecen de actividad de señalización en ausencia de ellos (Yan *et al.* 2002; Yeo y Whitman, 2001). La proteína soluble CRIPTO, puede activar las vías RAS / RAF / MAPK y PI3K / AKT de manera independiente a la vía Nodal en células mamarias epiteliales (Bianco *et al.* 2002; Bianco *et al.* 2003).

Los resultados del western blot se representan en la Figura 49.

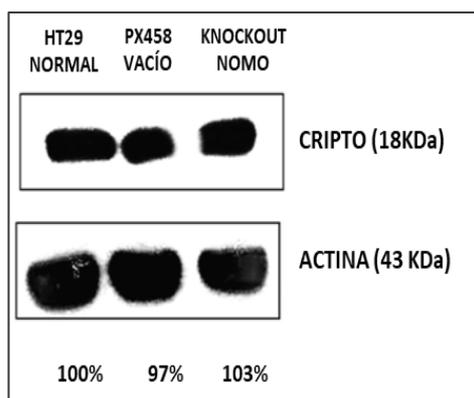


Figura 49. Expresión de la proteína CRIPTO en los tres clones: HT29 normal (control), HT29 transfectada con el plásmido PX458 vacío (control) y el clon knockout para *NOMO*. La expresión de las proteínas está calculada en forma de ratio proteína de interés/ β -actina.

Los resultados obtenidos revelaron que la expresión de la proteína CRIPTO no se ve modificada cuando se inactiva *NOMO*. En los tres clones el nivel de expresión de la proteína CRIPTO era similar.

Aguas abajo de los receptores ALK4 y ActRII, la señalización de la vía Nodal se lleva a cabo a través de la rama de las proteínas SMADs concretamente por SMAD2 y / o SMAD3 junto con el mediador común SMAD4 (Massague *et al.* 2005). Cuando SMAD2 y SMAD3 se fosforilan, interactúan con el mediador SMAD4 y como consecuencia se forma el complejo transcripcional en el núcleo.

Dado el importante papel que desempeñan en esta vía, también analizamos la expresión de las proteínas SMADs cuando se inactiva *NOMO*. Analizamos la expresión de SMAD2, SMAD3, SMAD4 y, también la expresión de las proteínas en su estado activado con el objetivo de ver si la ausencia de *NOMO* impedía la fosforilación de alguna de ellas y, como consecuencia, no tenía lugar la formación del complejo transcripcional (Figura 50).

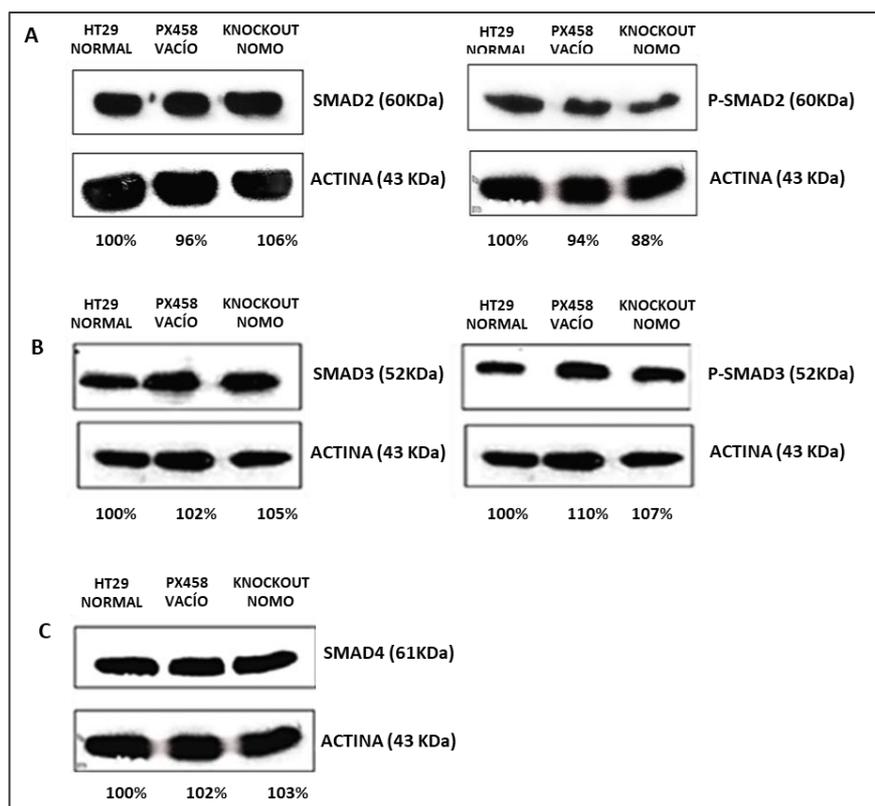


Figura 50. Expresión de las proteínas SMADs: SMAD2, SMAD2-P, SMAD3, SMAD3-P y SMAD4 en los tres clones: HT29 normal (control), HT29 transfectada con el plásmido PX458 vacío (control) y el clon *knockout* para *NOMO*. La expresión de las proteínas está calculada en forma de ratio proteína de interés/ β -actina. A) Expresión de las proteínas SMAD2 y SMAD2-P. B) Expresión de las proteínas SMAD3 y SMAD3-P. C) Expresión de SMAD4.

El estudio de la expresión de las proteínas SMADs mostró resultados homogéneos en todos los clones analizados. Si comparamos la expresión de estas proteínas en los dos clones que utilizamos como control con respecto al clon *knockout* para *NOMO*, se puede observar que no existen diferencias en los niveles de expresión en ninguno de ellos.

Es decir, la inactivación de *NOMO* no influye en la expresión de las proteínas integrantes de la vía Nodal, sin embargo, sí influye en la expresión de las proteínas con las que forma el complejo proteico NCLN/*NOMO*/TEMEM147. La NCLN en ausencia de *NOMO* no se expresa y TEMEM147 reduce tanto la expresión que prácticamente también se puede considerar nula cuando se inactiva *NOMO*.

3. ESTUDIO DEL TRANSCRIPTOMA EN AUSENCIA DE *NOMO*

Como los resultados obtenidos en este trabajo indicaron que el bloqueo de *NOMO* no altera la actividad de la vía Nodal, quisimos profundizar más en el papel que juega esta proteína en el CCR. Por ello, se llevó a cabo un análisis del transcriptoma en los cuatro clones obtenidos (2 *knockout* para *NOMO* y 2 controles). El análisis de los resultados mostró que al bloquear

NOMO se desregulan un total de 1.294 genes. De entre todos ellos, 504 genes estaban sobrexpresados (fold change>2) y 790 genes estaban infraexpresados (fold change<0,5). En la tabla 27 se muestran los genes que presentan diferencias estadísticamente significativas en la expresión en los clones *knockout* para *NOMO* respecto a la expresión de los clones control.

Tabla 27. Genes con diferencias estadísticamente significativas en su expresión al inactivar *NOMO*. Estudio en dos clones *knockout* para *NOMO* y dos clones utilizados como control.

Gen	Descripción	Clon 1	Clon 2	HT29 normal	PX458 vacío	Fold Change
<i>HERC6</i>	HECT and RLD domain containing E3 ubiquitin protein ligase family member 6	6,758	7,053	10,099	10,464	0,096
<i>CXCL11</i>	chemokine (C-X-C motif) ligand 11	3,243	3,282	7,072	7,652	0,058
<i>PLAC8</i>	placenta specific 8	12,007	12,175	8,841	9,072	8,781
<i>AKR1B10</i>	aldo-keto reductase family 1, member B10 (aldose reductase)	10,487	10,528	7,645	7,633	7,308
<i>DEPTOR</i>	DEP domain containing MTOR-interacting protein	3,888	4,588	9,772	9,261	0,025
<i>EXT1</i>	exostosin glycosyltransferase 1	9,319	9,385	12,551	12,665	0,104
<i>NOMO1</i>	NODAL modulator 1	8,670	7,312	14,487	14,388	0,011
<i>NOMO2</i> , <i>NOMO3</i>	NODAL modulator 2; NODAL modulator 3	8,012	6,764	13,268	13,433	0,016
<i>COL9A3</i>	collagen, type IX, alpha 3	7,795	7,335	3,575	3,519	16,201
<i>SAMD9</i>	sterile alpha motif domain containing 9	8,849	9,303	12,540	13,057	0,057
<i>PON3</i>	paraoxonase 3	5,540	5,548	7,310	7,313	0,293
<i>JPH1</i>	junctophilin 1	3,480	3,603	6,120	6,457	0,149
<i>DERL1</i>	derlin 1	11,314	11,456	13,649	13,685	0,205
<i>FOLR1</i>	folate receptor 1 (adult)	2,710	3,377	6,589	6,691	0,082
<i>MMP7</i>	matrix metalloproteinase 7	8,236	7,941	11,459	11,051	0,111
<i>ELF5</i>	E74-like factor 5 (ets domain transcription factor)	6,966	6,572	3,630	3,417	9,482

Posteriormente, se llevó a cabo un análisis preliminar con la herramienta WebGestalt en el que se obtuvieron las vías de señalización desreguladas al introducir en la herramienta web los 1.294 genes alterados que resultan tras bloquear *NOMO*. Los resultados que se obtuvieron muestran que las rutas más afectadas ante el bloqueo de *NOMO* fueron rutas relacionadas con la infección por virus y rutas que intervienen en procesos del sistema inmunitario (Figura 51).

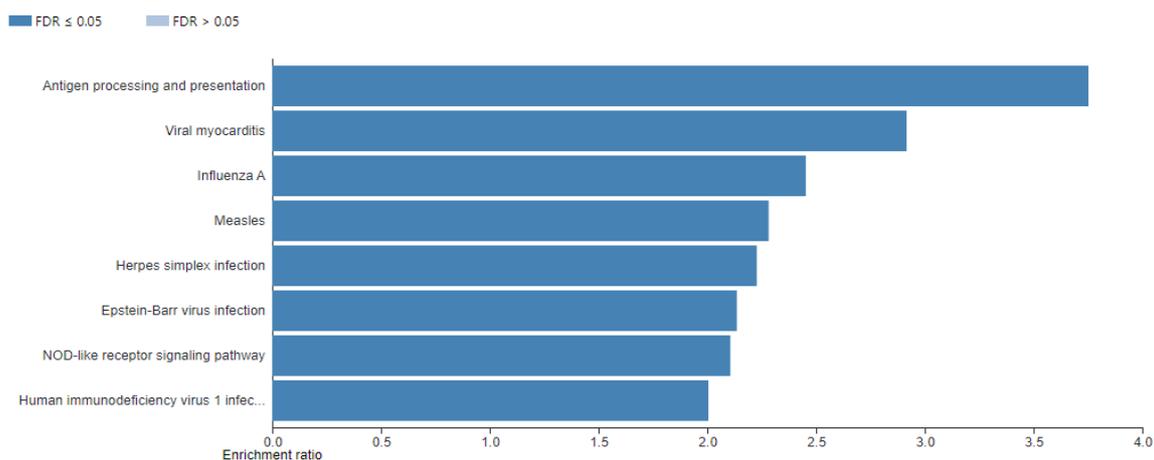


Figura 51. Vías de señalización alteradas considerando todos los genes desregulados ante el bloqueo de *NOMO*.

En este primer análisis nos llamó la atención la desregulación de la vía NOD la cual está implicada en la enfermedad inflamatoria intestinal de Crohn, que se asocia con un aumento de riesgo a desarrollar CCR. Los datos obtenidos mostraron que la expresión de los genes implicados en esta vía de señalización en los clones *knockout* para *NOMO* era inferior respecto a los clones utilizados como control (Anexo 10). Por ello, entre nuestros pasos futuros, está analizar el papel de *NOMO* en esta vía de señalización, así como encontrar otras posibles vías alteradas teniendo en cuenta por un lado genes sobreexpresados y por otro, genes infraexpresados al inactivar *NOMO*.

4. ESTUDIO GENÉTICO DE LOS GENES IMPLICADOS EN LA VÍA DE SEÑALIZACIÓN NODAL Y DEL COMPLEJO PROTEICO *NOMO/NCL/TEMEM147* MEDIANTE SECUENCIACIÓN MASIVA (NGS)

4.1. ASPECTOS GENERALES

Para llevar a cabo el estudio de los genes implicados en la vía de señalización Nodal y del complejo proteico que actúa como inhibidor de la vía, se diseñó a través de la compañía Illumina Inc. (San Diego, Ca, USA), un panel específico con todos los genes implicados en esta vía. El diseño de este panel no garantizaba una cobertura del 100% en todas las regiones amplificadas de los genes que se incluyeron en el estudio debido a diferentes obstáculos encontrados en el diseño de las sondas que se detallan en la Tabla 28.

Tabla 28. Genes incluidos en el panel diseñado para el estudio genético de los genes implicados en la vía Nodal. Junto a los genes que componen el panel se detalla si se encuentra en una región de repeticiones y/o si tiene una alta porción de contenido en CG lo cual, ambas situaciones, afectan al rendimiento de la secuenciación.

Gen	Observaciones
<i>ALK4 (ACVR1B)</i>	Hay una región de repetición que afectaría al rendimiento en la secuenciación de algunos amplicones del gen. También parece haber un alto contenido de GC.
<i>ActR11</i>	Hay una región de repetición que afectaría al rendimiento en la secuenciación de algunos amplicones del gen. No parece haber un alto contenido de GC.
<i>SMAD3</i>	No hay región de repeticiones y el contenido en CG no parece ser muy alto.
<i>NCNL</i>	No hay región de repeticiones, pero el contenido en CG es bastante alto por lo que puede afectar al rendimiento de la secuenciación.
<i>SMAD4</i>	No hay región de repeticiones y el contenido en CG no parece ser muy alto.
<i>NOMO1</i>	Hay una región de repetición que afectaría al rendimiento en la secuenciación de algunos amplicones del gen. No parece haber un alto contenido de GC.
<i>SMAD2</i>	No hay región de repeticiones y el contenido en CG no parece ser muy alto.
<i>TEMEM147</i>	Hay una región de repetición que afectaría al rendimiento en la secuenciación de algunos amplicones del gen. También parece haber un alto contenido de GC.
<i>LEFTY</i>	No hay región de repeticiones y el contenido en CG no parece ser muy alto.
<i>GPR78</i>	No hay región de repeticiones y el contenido en CG no parece ser muy alto.
<i>TDGF1</i>	No hay región de repeticiones y el contenido en CG no parece ser muy alto.

Los avances de los últimos años en las tecnologías de ultrasecuenciación o secuenciación de alto rendimiento han hecho posible el estudio de numerosos genes simultáneamente. A pesar de que en muchos amplicones no se garantizaba una cobertura superior al 75% debido a que algunos de estos genes se localizan en regiones de repeticiones o al alto contenido en CG, el uso de NGS en este proyecto tiene ventajas respecto a la utilización de técnicas tradicionales como la secuenciación automática por Sanger, puesto que se trata de una técnica más sensible y barata, proporcionando resultados robustos y de manera mucho más rápida.

Para el estudio de mutaciones en los genes implicados en la vía Nodal, se seleccionaron un total de 50 muestras procedentes de pacientes diagnosticados de CCR a los que previamente se les había analizado el número de copias de *NOMO* mediante q-PCR.

En todas estas muestras de DNA procedentes de tumor embebido en bloque de parafina, lo primero que se hizo fue analizar la calidad de cada una de ellas. De las 50 muestras analizadas, 4 de ellas no cumplían los requisitos de calidad exigidos por la técnica por lo que fueron excluidas del estudio. El resto de las muestras se secuenciaron en el equipo MiSeq (Illumina) y, finalmente se analizaron mediante los softwares de análisis: VariantStudio

Software_Illumina™ 3.0 (Illumina, San Diego, CA), DNA_Nexus e Integrative Genomics Viewer (IGV 2.4), todas las variantes obtenidas.

Para el análisis de todos los archivos generados en la secuenciación, se tuvieron en cuenta cada uno de los filtros recogidos por los propios programas de análisis utilizados (únicamente se contabilizaron las variantes que pasaron todos los filtros de calidad) y, además, solo consideramos aquellas variables que se presentaron con una frecuencia alélica superior al 10%.

-Eficiencia de la técnica

De las 46 muestras incluidas en el estudio, 36 se secuenciaron de manera correcta, es decir, en la mayoría de los amplicones secuenciados obtuvimos una cobertura mínima por base de 1000x. En las 10 restantes, no se alcanzó el número de lecturas mínimo requerido por la técnica y, por tanto, fueron excluidas del estudio. Las características de las muestras que finalmente se secuenciaron con éxito se detallan en la Tabla 29.

Tabla 29. Muestras secuenciadas de manera correcta en el estudio de mutaciones en genes implicados en la vía Nodal.

	NOMO sin alteraciones en el nº de copias (n)	NOMO Heterocigoto (n)	NOMO Homocigoto (n)
MAYORES DE 70 AÑOS	3	4	1
MENORES DE 45 AÑOS	8	10	8
PÓLIPOS EN MENORES DE 45 AÑOS	2	-	-

4.2. ANÁLISIS DE LOS DATOS

El análisis reportó unos datos de mínima cobertura por base (1000x) y de cobertura media por amplicón (5000X). Se comprobó que las mutaciones encontradas no estaban contenidas en las regiones de cobertura insuficiente. Las variantes localizadas en estas regiones no se tuvieron en cuenta en el análisis. Además, tan solo se dieron válidas mutaciones que aparecían en ambas hebras.

Este estudio ha permitido detectar tanto mutaciones patogénicas, como variantes de significado clínico desconocido y variantes neutrales que aparecen con una frecuencia alélica superior al 10% en los tumores analizados. Los resultados de la secuenciación con el panel *custom* diseñado por Illumina, mostraron que, de los 36 tumores analizados, todos ellos (100%) presentaban cambios en alguno de los genes que componen el panel. Solo 6 (15%) de ellos presentaban únicamente variantes neutrales, el resto de las muestras (85%), además de tener variantes neutrales, tenían o bien mutaciones patogénicas o variantes de significado clínico desconocido en alguno de los genes que componen el panel. En total, en este estudio se han

identificado 286 cambios localizados en los diferentes genes que se incluyen en el panel. De todos ellos, 150 se corresponden con cambios neutrales, 112 con variantes de significado clínico desconocido de las cuales: 45 estaban descritas por los programas de predicción como variantes benignas, 23 variantes descritas como variantes benignas por uno de los programas y como patogénicas por otro, por lo que las consideramos como variantes de significado clínico desconocido y, 44 descritas como patogénicas por los programas. Finalmente, 24 mutaciones patogénicas de las cuales 17 se correspondían con mutaciones *nonsense* que generan un codón de parada prematuro y, por tanto, dan lugar a una proteína truncada, 3 son mutaciones missenses descritas por otros grupos como mutaciones patogénicas y, las 4 restantes son mutaciones intrónicas localizadas en el primero o último nucleótido intrónico adyacente al exón.

4. 1.1. VARIANTES NEUTRALES

En este trabajo hemos encontrado un total de 150 variantes neutrales: 6 (4%) en el gen *LEFTY*, 12 (8%) en *ActR11*, 5 (3,5%) en *ALK4*, 15 (10%) en *GPR78*, 37 (24,6%) en *NOMO*, 9 (6,3%) en *SMAD2*, 18 (12%) en *SMAD3*, 8 (5,6%) en *SMAD4*, 32 (21%) en *NCLN* y, 8 (5%) en *TDGF1* (Tabla 30).

Hemos considerado variantes neutrales a todos aquellos cambios silenciosos o variantes sinónimas, definidas como aquellas en las que la sustitución de un nucleótido por otro en un codón determinado no genera un cambio de aminoácido. En nuestro estudio hemos encontrado un total de 59 variantes silenciosas diferentes repartidas en las regiones codificantes de los genes incluidos en el panel (Tabla 30).

Por otra parte, en este trabajo también se han considerado como variantes neutrales todas las variantes intrónicas localizadas en regiones alejadas de las zonas de *splicing*. Las variantes intrónicas son cambios de nucleótidos en zonas no codificantes del gen. Estas alteraciones, de manera general, se pueden localizar en regiones alejadas de los exones, las cuales no suelen afectar al procesamiento del gen o en las regiones que flanquean a los exones que sí pueden afectar al procesamiento del RNA y dar lugar al truncamiento de la proteína. El diseño empleado no nos permite analizar zonas intrónicas alejadas del exón, pero sí las más próximas. En este trabajo, se han identificado un total de 73 variantes intrónicas de las cuales algunas se localizan en regiones bastante alejadas de los exones mientras que otras se localizan en regiones más próximas y, por tanto, el riesgo de afectar a la región de *splicing* y con ello al procesamiento del RNA aumenta considerablemente (Tabla 30). Estas variantes localizadas en regiones que pueden afectar al *splicing* son: c.2757+3A>G y c.3028-3C>T ambas se encuentran en el gen *NOMO*. No existen estudios funcionales de ninguna de estas dos mutaciones para

indicar que pudieran afectar a la región de splicing, sin embargo, las hemos incluido dentro de este grupo de cambios neutrales porque en la base de datos dSNP (<https://www.ncbi.nlm.nih.gov/snp>) los estudios poblacionales realizados apuntaban una frecuencia superior al 1% del alelo mutado en varios subgrupos de poblaciones.

Además, también se identificaron cambios localizados en los extremos UTR de algunos de los genes analizados. Concretamente contabilizamos un total de 5 cambios en el extremo 5`UTR y otros 5 en el extremo 3` (Tabla 30). Las alteraciones en los extremos UTR de los genes, han sido consideradas de manera clásica como variantes neutrales al no localizarse dentro de la secuencia codificante o en las secuencias consenso implicadas en la maquinaria de maduración del ARN.

Por último, también se han considerado variantes neutrales 8 variantes *missense* cuya frecuencia poblacional en la base de datos dbSNP era superior al 1%, por lo que pueden considerarse polimorfismos. Las mutaciones *missense* se describen como mutaciones puntuales en las que un único cambio de nucleótido da como resultado un codón que codifica un aminoácido diferente. Algunas de estas mutaciones pueden hacer que la proteína resultante no funcione bien porque el codón codificante se cambia a otro que da lugar a un aminoácido diferente. Pero, por otra parte, no todas las mutaciones *missense* conducen a cambios apreciables de la proteína. Un aminoácido puede ser reemplazado por un aminoácido de propiedades químicas muy similares, en cuyo caso, la proteína todavía puede funcionar de manera normal, en este caso, se considera un polimorfismo o cambio neutral. No sólo esto, la sustitución en una zona de la proteína que no sea importante para la función de ésta tampoco puede tener repercusión, aunque el aminoácido que se sustituya sea de propiedades diferentes.

Tabla 30. Listado de variantes consideradas como variantes neutrales identificadas en los diferentes genes que componen el panel de genes implicado en la vía Nodal. En la tabla se registran todas las variantes identificadas junto con el rs (las que están descritas en la base de datos dbSNP) y la frecuencia de aparición en los tumores analizados.

Gen	Variante	rs	Frec. tumor (%)	Variante	rs	Frec. tumor (%)
LEFTY	c.-68C>T		1 (0,66)	c.497+13C>G	74578461	4 (2,6)
	c.-10G>A	35310234	1 (0,66)	c.498-43T>A		1 (0,66)
	c.375C>T (p.=)	12753531	2 (1,3)	c.543C>T(p.=)		2 (1,3)
ActRII	c.-19G>A		1 (0,66)	c.463+41delC	113527021	1 (0,66)
	c.48+28G>A		1 (0,66)	c.600+10A>C	268871	10 (6,7)
	c.90C>T(p.=)		1 (0,66)	c.600+26A>C		2 (1,3)
	c.336C>A(p.=)		2 (1,3)	c.600+76G>A	756439882	2 (1,3)
	c.391-46G>A	753901904	2 (1,3)	c.896+12G>A	268862	2 (1,3)
	c.463+32T>G	75758611	1 (0,66)	c.*19G>A	268875	10 (6,7)
ALK4	c.331+12C>T	2242106	4 (2,6)	c.1173C>T(p.=)		2 (1,3)
	c.580+131T>C		1 (0,66)	c.1384+87C>T		1 (0,66)
	c.955+49T>C		1 (0,66)			
GPR78	c.33C>T(p.=)	17844776	8 (5,3)	c.789G>A(p.=)	61746124	2 (1,3)
	c.147T>G(p.=)	17844777	12 (7,92)	c.852C>T(p.=)		2 (1,3)
	c.153C>T(p.=)		2 (1,3)	c.948C>T(p.=)		1 (0,66)
	c.576G>A(p.=)		1 (0,66)	c.988G>A/ p.(Ala330Thr)		1 (0,66)
	c.603A>C/ p.(Arg201Ser)	17844778	17 (11,22)	c.1025G>A/ p.Arg342His	11941659	1 (0,66)
	c.782+23C>A	3115387	1 (0,66)	c.*65C>T	9685931	4 (2,6)
	c.783-96C>T		8 (5,3)	c.*76C>A	9683448	4 (2,6)
	c.783-101C>T	79784260	2 (1,3)			
NOMO	c.165+53C>T		3 (2)	c.2274-16A>G		12 (7,9)
	c.255+56G>A		1 (0,66)	c.2357-96C>A		1 (0,66)
	c.816C>T(p.=)		2 (1,3)	c.2379C>T(p.=)	201279272	2 (1,3)
	c.964-32C>T	746000290	4 (2,6)	c.2644-22G>T		3 (2)
	c.972C>T(p.=)		1 (0,66)	c.2733C>T(p.=)	200650450	2 (1,3)
	c.1257C>T(p.=)		1 (0,66)	c.2736C>A(p.=)	393246	10 (6,7)
	c.1456C>T/	368046747	1 (0,66)	c.2757+3A>G	375487	5 (3,3)
	p.(Leu486Phe)	1062412	1 (0,66)	c.2758-18G>A		1 (0,66)
	c.1468A>G/	200314993	5 (3,3)	c.2862C>A(p.=)		2 (1,3)
	p.(Asn490Asp)		1 (0,66)	c.2211T>A(p.=)		2 (1,3)
	c.1395+26A>G	77211612	1 (0,66)	c.2273+34T>G		1 (0,66)
	c.1491C>T(p.=)		1 (0,66)	c.3027+27G>A		2 (1,3)
	c.1537+50A>G		1 (0,66)	c.3028-3C>T		1 (0,66)
	c.1670-32C>T	375667439	3 (2)	c.3222+100GA		5 (3,3)
	c.1677C>T(p.=)		5 (3,3)	c.3222+126TA	145176199	3 (2)
	c.1776G>A(p.=)	778568131	3 (2)	c.3591C>T(p.=)	751973033	1 (0,66)
	c.1806+45T>A	11863419	5 (3,3)	c.*54G>A	571281596	1 (0,66)
	c.1958+27G>A		2 (1,3)	c.*68C>T		1 (0,66)
	c.2055-34C>T		2 (1,3)			
	c.2070T>C(p.=)					
	c.2166G>A(p.=)					

SMAD2	c.30A>G(p.=)	746254264	1 (0,66)	c.627A>C(p.=)	4 (2,6)
	c.207C>T(p.=)		1 (0,66)	c.630T>C(p.=)	4 (2,6)
	c.327-6C>T		1 (0,66)	c.696A>C(p.=)	762990520 2 (1,3)
	c.327-39delT	34501809	4 (2,6)	c.1135+74C>T	1 (0,66)
	c.621A>C(p.=)		4 (2,6)		
SMAD3	c.-15G>A	1061427	2 (1,3)	c.573C>A(p.=)	1 (0,66)
	c.207-37C>T		1 (0,66)	c.608-5C>T	777329823 3 (2)
	c.207-93C>T		2 (1,3)	c.632C>T/ p.(Pro211Leu)	772510704 1 (0,66)
	c.252T>C(p.=)		1 (0,66)	c.658+18C>T	2 (1,3)
	c.309A>G(p.=)	1065080	17 (11,2)	c.801G>A(p.=)	764958567 2 (1,3)
	c.400+59G>C		13 (8,6)	c.871+43C>T	140143645 1 (0,66)
	c.401-44T>A		4 (2,6)	c.963C>T(p.=)	1 (0,66)
	c.532+128G>A	2289259	4 (2,6)	c.969C>A(p.=)	1 (0,66)
	c.532+158C>A		2 (1,6)	c.1023G>A(p.=)	2 (1,3)
SMAD4	c.259C>A(p.=)		1 (0,66)	c.904+44_904+45dupTT	386387676 5 (3,3)
	c.425-78C>A		1 (0,66)	c.905-15T>A	1 (0,66)
	c.668-52C>T		2 (1,3)	c.975C>T(p.=)	1 (0,66)
	c.904+29A>T	543382977	1 (0,66)	c.1551C>T(p.=)	2 (1,3)
NCLN	c.-45G>A		1 (0,66)	c.889+73G>A	2 (1,3)
	c.37G>C(p.=)		1 (0,66)	c.800+71A>G	311625 7 (4,6)
	c.57G>T(p.=)	11551095	2 (1,3)	c.801-104G>A	2 (1,3)
	c.213G>A(p.=)		1 (0,66)	c.1029+30A>G	8106438 4 (2,6)
	c.219C>T(p.=)		1 (0,66)	c.1128C>T(p.=)	1 (0,66)
	c.57G>T(p.=)	11551095	2 (1,3)	c.1297-36G>T	757610211 1 (0,66)
	c.184+77G>A		1 (0,66)	c.1336-25A>G	482758 5 (3,3)
	c.213G>A(p.=)		1 (0,66)	c.1419C>T(p.=)	1 (0,66)
	c.219C>T(p.=)		1 (0,66)	c.1476G>A(p.=)	2 (1,3)
	c.420C>T(p.=)	757637380	1 (0,66)	c.1518C>T(p.=)	1 (0,66)
	c.520+58A>G	499877	3 (2)	c.1554-51T>C	312068 6 (4)
	c.615+24C>T		1 (0,66)	c.1554-41G>C	312069 7 (4,6)
	c.615+32C>T		1 (0,66)	c.1563G>A(p.=)	76473246 1 (0,66)
	c.696+33A>T	776111068	3 (2)	c.1587T>C(p.=)	1 (0,66)
	c.697-38C>T	373531721	1 (0,66)	c.1632+28C>T	1 (0,66)
	c.714G>A(p.=)	575977828	2 (1,3)	c.1652A>G/ p.(Lys551Arg)	2288949 1 (0,66)
TDGF1	c.35+60G>A	55955974	1 (0,66)	c.224-106A>T	9819070 12 (7,94)
	c.65T>C/ p.(Val22Ala)	11130097	9 (5,94)	c.306C>T(p.=)	1 (0,66)
	c.168G>A(p.=)	2293024	1 (0,66)	c.456T>C(p.=)	112928637 1 (0,66)
	c.223+43T>A		3 (2)		

Frec. tumor.: frecuencia de la variante en los 36 tumores analizados en este estudio.

4. 1.2. VARIANTES DE SIGNIFICADO CLÍNICO DESCONOCIDO

Todas las mutaciones de significado clínico desconocido encontradas en este trabajo son mutaciones *missense* o variantes intrónicas que, por su cercanía a la región codificante, pueden afectar al *splicing* del RNA.

De todas las variantes *missenses* identificadas en este trabajo, únicamente tres de ellas (c.1229T>G/ p.Val410Gly (*SMAD3*), c.1268G>A/p.Ser423Asn (*SMAD3*) y c.1081C>T/p.Arg361Cys (*SMAD4*)) han sido previamente descritas por otros grupos como variantes que afectan al funcionamiento de la proteína y, por ello, se catalogan como mutaciones patogénicas (Schepers *et al.* 2018; Houlston *et al.* 1998). El resto de las variantes, las hemos clasificado en función de lo establecido por los programas de predicción utilizados en: posiblemente benignas (los programas de predicción utilizados la catalogan como neutral), variantes de significado clínico desconocido (alguno de los programas de predicción utilizados la catalogan como neutral y otros como patogénica) y, posiblemente patogénicas (los programas la clasifican como patogénica).

-Variantes probablemente neutrales

En este estudio se identificaron un total de 45 variantes *missenses* clasificadas por los programas de predicción como neutrales o benignas (Tabla 31). De todas ellas, 36 (80%) no estaban recogidas en ninguna base de datos y, por tanto, no hay ningún tipo de estudio que facilite la clasificación de ellas. El resto de estas variantes, 9 (20%), si estaban recogidas en la base de datos dbSNP. Los datos obtenidos de las frecuencias alélicas de los estudios poblacionales concluyen que, en ninguna variante, se supera el 1% de frecuencia en la población control y por ello, al igual que las 36 variantes que no están descritas, sería necesario realizar estudios funcionales para su clasificación, mientras tanto se deben de considerar como variantes de significados clínico desconocido.

De las 45 variantes clasificadas como probablemente neutrales: 2 (4,5%) en el gen *LEFTY*, 2 (4,5%) en *ActR11*, 4 (8,8%) en *ALK4*, 5 (11,1%) en *GPR78*, 19 (42,2%) en *NOMO*, 2 (4,5%) en *SMAD2*, 1 (2,25%) en *SMAD3*, 4 (8,8%) en *SMAD4*, 5 (11,1%) en *NCLN* y, 1 (2,25%) en *TDGF1* (Tabla 31).

Tabla 31. Listado de variantes descritas como neutrales por los programas de predicción Sift, Polyphen y DACC. En la tabla se describe cada una de las variantes identificadas junto con el número del rs (las que están descritas en la base de datos dbSNP) y la frecuencia de aparición en los tumores analizados. También se indica la predicción propuesta por cada programa y la frecuencia alélica global de aparición en población ExAC (<http://exac.broadinstitute.org>) en las variantes que aparecen recogidas en esta base de datos. Frec. tumor.: frecuencia de la variante en los 36 tumores analizados en este estudio. Frec. poblacional: frecuencia de la variante en población control.

Gen	Variante	rs	Frec. tu. % (n)	Sift	Polyphen	CADD Score	Frec. poblacional (%)
LEFTY	c.59C>T/ p.(Ala20Val)		2.7 (1)	tolerated (0.09)	benign (0.007)	6.916	No datos
	c.144G>A/ p.(Met48Ile)		2.7 (1)	tolerated (0.07)	benign (0.004)	9.791	No datos
ActR11	c.806C>T/ p.(Ala269Val)		2.7 (1)	tolerated (0.4)	benign (0.03)	13,4	No datos
	c.1134G>A/ p.(Met378Ile)	752116809	2.7 (1)	tolerated (1)	benign (0.003)	12.16	0
ALK4	c.1234C>T/ p.(Pro412Ser)		2.7 (1)	tolerated (0.41)	benign (0.396)	22.6	No datos
	c.1297A>G/ p.(Met433Val)	751359032	2.7 (1)	tolerated (0.44)	benign (0.006)	12.91	0
	c.1438T>A/ p.(Ser480Thr)		2.7 (1)	tolerated (0.16)	benign (0.06)	12.0	No datos
	c.1508G>A/ p.(Ser503Asn)		2.7 (1)	tolerated (0.06)	benign (0.017)	12.9	No datos
GPR78	c.7C>T/ p.(Pro3Ser)		2.7 (1)	tolerated (0.49)	benign (0.003)	0.292	No datos
	c.643G>A/ p.(Ala215Thr)		2.7 (1)	tolerated (0.07)	benign (0.069)	2.86	No datos
	c.733A>T/ p.(Ile245Phe)		2.7 (1)	tolerated (0.11)	benign (0.219)	1,156	No datos
	c.777G>A/ p.(Met259Ile)		2.7 (1)	tolerated (1)	benign (0)	0.006	No datos
	c.997G>A/ p.(Val333Met)	572850332	2.7 (1)	tolerated (1)	benign (0)	0.004	No datos
NOMO	c.31G>A/ p.(Gly11Arg)		2.7 (1)	tolerated (1)	benign (0)	7.273	No datos
	c.82C>T/ p.(Pro28Ser)		5,5 (2)	tolerated (0.64)	benign (0)	26.8	No datos
	c.935C>T/ p.(Thr312Ile)		2.7 (1)	tolerated (0.13)	benign (0.049)	20.8	No datos
	c.967G>A/ p.(Val323Met)	374612573	2.7 (1)	tolerated (0)	benign (0.375)	15.06	0,02
	c.1165G>A/ p.(Val389Ile)		2.7 (1)	tolerated (0.65)	benign (0.004)	0.048	No datos
	c.1181C>T/ p.(Ala394Val)		2.7 (1)	tolerated (0.07)	benign (0.122)	22.9	No datos
	c.1267G>A/ p.(Val423Ile)	777398188	2.7 (1)	tolerated (0.91)	benign (0.002)	6.37	0
	c.1490C>T/ p.(Ala497Val)	62038475	2.7 (1)	tolerated (0.28)	benign (0.003)	21.7	No datos
	c.1510T>A/ p.(Ser504Thr)		2.7 (1)	tolerated (0.23)	benign (0.113)	14.3	No datos
	c.2178T>A/ p.(Asn726Lys)		20 (7)	tolerated (1)	benign (0)	0.02	0,16
	c.2260T>G/ p.(Ser754Ala)	768758230	2.7 (1)	tolerated (0.06)	benign (0.078)	10.79	No datos
	c.2411A>C/ p.(Glu804Ala)		5,5 (2)	tolerated (0.27)	benign (0.01)	12.35	No datos
	c.2549C>T/ p.(Thr850Met)		5,5 (2)	tolerated (0.09)	benign (0.416)	23.4	0.00001
	c.2562G>T/ p.(Gln854His)		2.7 (1)	tolerated (0.17)	benign (0.019)	17.18	No datos
	c.2650G>A/ p.(Ala884Thr)		2.7 (1)	tolerated (1)	benign (0.08)	13.39	No datos
	c.3563C>T/ p.(Thr1188Ile)		2.7 (1)	tolerated (0.41)	benign (0.049)	6.56	No datos
	c.3569G>A/ p.(Arg1190Gln)	769186710	5,5 (2)	tolerated (0.21)	benign (0.04)	18.53	No datos
c.3614G>A/ p.(Ser1205Asn)		2.7 (1)	tolerated (0.17)	benign (0.109)	19.12	0	
c.3622G>A/ p.(Glu1208Lys)		2.7 (1)	tolerated (0.05)	benign (0.272)	20.9	No datos	
SMAD2	c.695C>T/ p.(Thr232Ile)		2.7 (1)	tolerated (0.13)	benign (0.005)	21.9	No datos
	c.760C>T/ p.(Leu254Phe)		2.7 (1)	tolerated (0.2)	benign (0.092)	19.86	No datos
SMAD3	c.502G>A/ p.(Ala168Thr)	548756379	2.7 (1)	tolerated (0.47)	benign (0.002)	24.2	0.02
SMAD4	c.10A>T/ p.(Met4Leu)		2.7 (1)	tolerated (0)	benign (0.155)	14.51	No datos
	c.697C>T/ p.(His233Tyr)		2.7 (1)	tolerated (0.84)	benign (0.035)	8.53	No datos
	c.938C>T/ p.(Pro313Leu)		2.7 (1)	tolerated (0.23)	benign (0.003)	21.3	No datos
	c.925G>A/ p.(Ala309Thr)		2.7 (1)	tolerated (0.46)	benign (0)	14.13	No datos
NCLN	c.101C>T/ p.(Ala34Val)		2.7 (1)	tolerated (0.28)	benign (0.002)	21.0	No datos
	c.248G>A/ p.(Arg83His)		2.7 (1)	tolerated (0.18)	benign (0.078)	20.2	No datos
	c.331A>T/ p.(Ile111Phe)		2.7 (1)	tolerated (0.11)	benign (0.013)	17.77	No datos
	c.679G>A/ p.(Ala227Thr)		2.7 (1)	tolerated (0.13)	benign (0.133)	16.37	0
	c.800C>T/ p.(Ala267Val)	764839344	2.7 (1)	tolerated (0.07)	benign (0.027)	22.1	No datos
TGDF1	c.452G>A/ p.(Gly151Asp)		2.7 (1)	tolerated (0.17)	benign (0.086)	9.41	No datos

Como ya describimos anteriormente, con el programa ESE FINDER realizamos un análisis bioinformático de predicción teórica del efecto de los polimorfismos exónicos sobre el procesamiento del RNA mensajero (Tabla 32).

Tabla 32. Predicciones realizadas mediante el programa ESE FINDER en las variantes probablemente neutrales. Marcados con el signo positivo aparecen las zonas en las que la variante analizada puede afectar a la maduración del RNA mensajero debido a que aparece un nuevo motivo ESE y, con signo negativo aparecen las zonas en las que la variante analizada puede afectar a la maduración del RNA mensajero debido a que desaparece el motivo ESE de la secuencia wt. = el programa no predice un cambio en ese sitio.

Gen	Variante	SF2/ASF	SC35	SRp40	SRp55
<i>LEFTY</i>	c.144G>A/ p.(Met48Ile)	=	=	-	=
<i>NOMO</i>	c.31G>A/ p.(Gly11Arg)	+	=	-	+
	c.935C>T/p.(Thr312Ile)	-	=	-	+
	c.967G>A/ p.(Val323Met)	-	=	=	=
	c.2562G>T/ p.(Gln854His)	-	=	-	=
<i>SMAD2</i>	c.760C>T/ p.(Leu254Phe)	=	=	-	=
<i>SMAD3</i>	c.502G>A/ p.(Ala168Thr)	=	+	-	=
<i>SMAD4</i>	c.10A>T/ p.(Met4Leu)	=	-	=	-
	c.697C>T/ p.(His233Tyr)	-	+	=	=
<i>NCLN</i>	c.331A>T/ p.(Ile111Phe)	=	-	=	=
	c.679G>A/ p.(Ala227Thr)	-	+	=	=
<i>TDGF1</i>	c.452G>A/ p.(Gly151Asp)	=	-	=	=

De las 45 variantes de significado clínico desconocido catalogadas como probablemente benignas por los programas de predicción utilizados en este trabajo, el programa ESE FINDER encontró valores significativos para 12 de ellas. En unas variantes se crean nuevos sitios ESE para unas proteínas SR y se eliminan otros sitios existentes en la secuencia germinal para otras proteínas. Estos resultados son orientativos y se requiere una comprobación mediante RT-PCR.

-Variantes de significado incierto

Además, también se identificaron un total de 23 cambios incluidos en el grupo de variantes de significado clínico desconocido donde unos programas de predicción las catalogaba como neutrales mientras que otros lo hacían como mutaciones deletéreas o posiblemente patogénicas (Tabla 33). De todas ellas, solo tres estaban descritas en la base de datos dbSNP y, en ninguna de las tres se alcanzaba una frecuencia alélica superior al 1% en los estudios poblacionales. Por ello, en ninguna de estas variantes se conoce si el cambio nucleotídico produce cambios en la función de la proteína. Para todos estos casos, es necesario realizar estudios funcionales para poderlas clasificar. Mientras tanto se deben de considerar como variantes de significado clínico desconocido.

En este caso, no se detectaron variantes en los genes *TEMEM147*, *SMAD4* y *LEFTY*. Por el contrario, de estos 23 cambios identificados, 9 (40%) de ellos se localizaban en el gen *NOMO*, mientras que los siguientes en frecuencia de aparición de esta subclase de mutaciones fueron los dos genes que codifican para las proteínas que actúan como receptores de la vía Nodal, el receptor tipo I (*ALK4*) y el receptor tipo II (*ActRII*) donde se identificaron un total de 4 (17,4%) variantes de significado clínico desconocido en cada uno de ellos.

Tabla 33. Listado de variantes descritas como neutrales por unos programas de predicción utilizados y como patogénicas por otros programas. En la tabla se describen cada una de las variantes identificadas junto con el valor del rs (las que están descritas en la base de datos dbSNP) y la frecuencia de aparición en los tumores analizados. También se indica la predicción propuesta por cada programa y la frecuencia alélica global de aparición en población ExAC (<http://exac.broadinstitute.org>) en las variantes que aparecen recogidas en esta base de datos.

Gen	Variante	rs	Frec.tu. % (n)	Sift	Polyphen	CADD Score	Frec. pobla(%)
ActRII	c.269T>C/ p.(Met90Thr)		4,3(1)	deleterious (0)	benign (0.004)	22.5	No datos
	c.646G>A/p.(Glu216Lys)		4,3(1)	deleterious (0.02)	benign (0.035)	25.8	No datos
	c.733T>A/ p.(Leu245Ile)		4,3(1)	tolerated (0.07)	possibly damaging (0.88)	23.1	No datos
	c.782G>A/ p.(Gly261Glu)		4,3(1)	tolerated (0.29)	probably damaging (1)	26.0	No datos
ALK4	c.269T>C/p.(Met90Thr)		4,3(1)	deleterious (0)	benign (0.004)	20.35	No datos
	c.386G>A/p.(Gly129Asp)		4,3(1)	tolerated (0.07)	possibly damaging (0.624)	24.1	No datos
	c.1498T>A/ p.(Trp500Arg)		4,3(1)	tolerated (1)	probably damaging (0.999)	22.0	No datos
	c.1629C>A/p.(Asp543Glu)		4,3(1)	tolerated (0.07)	probably damaging (0.999)	23.1	No datos
GPR78	c.952C>T/ p.(Arg318Cys)	61746144	4,3(1)	tolerated (0.1)	possibly damaging (0.513)	10.69	0,44
NOMO	c.714T>G/p.(Phe238Leu)		4,3(1)	deleterious (0.02)	benign (0.437)	15.46	No datos
	c.1807G>A/ p.(Glu603Lys)		4,3(1)	tolerated (0.19)	possibly damaging (0.786)	24.8	No datos
	c.2440G>A/p.(Val814Ile)		4,3(1)	tolerated (0.09)	probably damaging (0.996)	24.4	No datos
	c.2542G>A/p.(Glu848Lys)	200338200	4,3(1)	deleterious (0.01)	benign (0.219)	22.9	0
	c.2716C>T/p.(Leu906Phe)		4,3(1)	tolerated (0.06)	probably damaging (0.931)	24.7	No datos
	c.3025C>A/p.(Leu1009Met)		4,3(1)	tolerated (0.05)	possibly damaging (0.736)	23.8	No datos
	c.3178A>G/p.(Ser1060Gly)	3895684	4,3(1)	tolerated (0.09)	probably damaging (0.997)	15.63	0
	c.3587C>T/p.(Ala1196Val)		30,5(7)	deleterious (0.01)	benign (0.155)	18.37	No datos
c.3665C>T/ p.(Thr1222Met)		4,3(1)	deleterious (0)	probably damaging (0.979)	23.2	No datos	
SMAD2	c.625G>A/ p.(Gly209Arg)		4,3(1)	deleterious (0.01)	benign (0.036)	25.8	No datos
SMAD3	c.967C>A/p.(Arg323Ser)		4,3(1)	tolerated (0.11)	possibly damaging (0.908)	32.0	No datos
NCLN	c.350C>T/ p.(Ala117Val)		4,3(1)	deleterious (0.02)	benign (0.005)	23.2	No datos
	c.1453C>T/p.(Arg485Cys)		4,3(1)	deleterious (0)	benign (0.021)	31.0	No datos
TDGF1	c.425C>T/ p.(Pro142Leu)		4,3(1)	deleterious (0.04)	benign (0.272)	23.1	No datos

En este caso, el programa ESE FINDER identificó un total de 10 variantes que presentaban diferencias estadísticamente significativas en las regiones ESE con respecto a la secuencia germinal (Tabla 34). Como se mencionó anteriormente, estos resultados son orientativos y se requiere una comprobación mediante RT-PCR.

Tabla 34. Predicciones realizadas mediante el programa ESE FINDER en variantes de significado incierto. Marcados con el signo positivo aparecen las zonas en las que la variante analizada puede afectar a la maduración del RNA mensajero debido a que aparece un nuevo motivo ESE y, con signo negativo aparecen las zonas en las que la variante analizada puede afectar a la maduración del RNA mensajero debido a que desaparece el motivo ESE de la secuencia wt. = el programa no predice un cambio en ese sitio.

Gen	Variante	SF2/ASF	SC35	SRp40	SRp55
<i>GPR78</i>	c.952C>T/ p.Arg318Cys	=	+	=	-
<i>NOMO</i>	c.1807G>A/ p.Glu603Lys	-	=	-	=
	c.2440G>A/p.Val814Ile	-	+	=	=
	c.3025C>A/p.Leu1009Met	-	-	=	-
<i>SMAD2</i>	c.625G>A/ p.Gly209Arg	-	=	=	=
<i>SMAD3</i>	c.967C>A/p.Arg323Ser	+	=	+	=
<i>SMAD4</i>	c.10A>T/ p.Met4Leu	=	-	=	-
	c.697C>T/ p.His233Tyr	-	+	=	=
<i>NCLN</i>	c.350C>T/ p.Ala117Val	-	=	=	=
<i>TDGF1</i>	c.425C>T/ p.Pro142Leu	-	=	=	=

-Variantes probablemente patogénicas

El tercer subgrupo de variantes incluidas dentro del grupo de variantes de significado clínico desconocido engloba a todas aquellas catalogadas como deletéreas o probablemente patogénicas por los programas de predicción utilizados en este trabajo. En nuestro estudio, hemos encontrado un total de 44 variantes de significado clínico desconocido incluidas en este grupo (Tabla 35). De todas ellas, solo 4 estaban descritas en la base de datos dbSNP y la frecuencia alélica para los estudios poblacionales no superaba el 1% en ninguna de ellas por lo que, en principio, no se podían catalogar como polimorfismos.

Pero, en este estudio también hemos identificado una serie de variantes en las regiones intrónicas de algunos de estos genes que, por su proximidad a los exones, pueden afectar al procesamiento del RNA mensajero codificado por el gen donde se localiza el cambio y, por tanto, codifican una proteína no funcional, tratándose en este caso de mutaciones patogénicas. Estas mutaciones estaban localizadas en las regiones de *splicing* y en las regiones UTR de los genes analizados y, al no estar descritas en la literatura, las hemos considerado dentro de este subgrupo (Tabla 35).

Tabla 35. Listado de variantes descritas como patogénicas por los dos programas de predicción utilizados. En la tabla se describen cada una de las variantes identificadas junto con el valor del rs (las que están descritas en la base de datos dbSNP) y la frecuencia de aparición en los tumores analizados. También se indica la predicción propuesta por cada programa y la frecuencia alélica global de aparición en población ExAC (<http://exac.broadinstitute.org>) en las variantes que aparecen recogidas en esta base de datos. Frec. tumor.: frecuencia de la variante en los 36 tumores analizados en este estudio. Frec. poblacional: frecuencia de la variante en población control.

Gen	Variante	rs	Frec. tumor n (%)	Sift	Polyphen	CADD Score	Frec. poblacional (%)
ActRII	c.308T>C/ p.(Leu103Pro)		2 (4,5)	deleterious (0)	probably damaging (0.967)	26.7	No datos
	c.353C>T/ p.(Pro118Leu)		2 (4,5)	deleterious (0)	probably damaging (0.961)	32.0	No datos
	c.61G>A/ p.(Gly21Arg)		1 (2,27)	deleterious (0)	probably damaging (1)	22.5	No datos
	c.1139G>T/ p.(Arg380Leu)		1 (2,27)	deleterious (0)	possibly damaging (0.684)	32.0	No datos
	c.779G>A/ p.(Gly260Glu)		1 (2,27)	deleterious (0.02)	probably damaging (0.985)	32.0	No datos
ALK4	c.1162G>A/ p.(Gly388Ser)		1 (2,27)	deleterious (0.05)	probably damaging (0.911)	27.1	No datos
	c.1193C>T/ p.(Ala398Val)		1 (2,27)	deleterious (0)	possibly damaging (0.771)	32.0	No datos
GPR78	c.205C>T/ p.(Arg69Cys)	199555298	1 (2,27)	deleterious (0)	possibly damaging (0.849)	23.3	0,02
	c.-4C>T (5'UTR)	370649634	1 (2,27)				0
NOMO	c.118G>T/ p.(Gly40Cys)		1 (2,27)	deleterious (0)	probably damaging (1)	23.1	No datos
	c.1517C>A/ p.(Ser506Tyr)		1 (2,27)	deleterious (0)	probably damaging (0.999)	28.6	No datos
	c.2261C>A/ p.(Ser754Tyr)		1 (2,27)	deleterious (0)	probably damaging (0.931)	25.6	No datos
	c.2278G>A/ p.(Gly760Arg)		1 (2,27)	deleterious (0.01)	probably damaging (0.992)	29.3	No datos
	c.2433G>T/ p.(Leu811Phe)		1 (2,27)	deleterious (0.01)	probably damaging (0.987)	25.8	No datos
	c.2743A>G/ p.(Thr915Ala)		1 (2,27)	deleterious (0,01)	possibly damaging (0.783)	21.6	No datos
	c.2787G>A/ p.(Met929Ile)		1 (2,27)	deleterious (0.01)	possibly damaging (0.783)	26.4	No datos
	c.2794G>A/ p.(Glu932Lys)		1 (2,27)	deleterious (0)	probably damaging (0.977)	32.0	No datos
	c.3002G>A/ p.Gly1001Asp)		1 (2,27)	deleterious (0.01)	probably damaging (0.998)	25.1	No datos
	c.3019G>A/ p.(Gly1007Arg)		1 (2,27)	deleterious (0)	probably damaging (1)	26.0	No datos
	c.3557A>G/ p.Gln1186Arg)	3895684	1 (2,27)	deleterious (0.05)	possibly damaging (0.78)	22.3	0
	c.3578G>A/ p.(Gly1193Glu)		1 (2,27)	deleterious (0.01)	probably damaging (0.919)	24.9	No datos
	c.3587C>T/ p.(Ala1196Val)		1 (2,27)	deleterious (0.01)	probably damaging (0.931)	24.1	No datos
SMAD2	c.167G>A/ p.(Gly56Glu)		1 (2,27)	deleterious (0.03)	possibly damaging (0.831)	25.9	No datos
	c.329C>T/ p.(Ser110Phe)		1 (2,27)	deleterious (0.02)	probably damaging (0.93)	26.5	No datos
	c.370C>T/ p.(Pro124Ser)		1 (2,27)	deleterious (0.03)	probably damaging (0.973)	26.0	No datos
	c.875C>A/ p.(Ala292Glu)		1 (2,27)	deleterious (0)	probably damaging (0.991)	29.9	No datos
	c.1028G>T/ p.(Gly343Val)		1 (2,27)	deleterious (0)	probably damaging (0.987)	33.0	No datos
	c.1214T>G/ p.(Val405Gly)		1 (2,27)	deleterious (0)	probably damaging (0.991)	27.6	No datos
	c.1391C>T/ p.(Ser464Leu)		1 (2,27)	deleterious (0)	probably damaging (0.963)	32.0	No datos
	SMAD3	c.218G>A/ p.(Gly73Asp)		1 (2,27)	deleterious (0)	possibly damaging (0.776)	31.0
c.658+3G>A			1 (2,27)				No datos
c.1268G>A/ p.(Ser423Asn)			1 (2,27)	deleterious (0.01)	possibly damaging (0.496)	28.3	No datos
SMAD4	c.311T>A/ p.(Leu104His)		1 (2,27)	deleterious (0)	probably damaging (1)	28.5	No datos
	c.965A>G/ p.(Tyr322Cys)		1 (2,27)	deleterious (0.01)	probably damaging (0.979)	27.0	No datos
	c.1081C>T/ p.(Arg361Cys)	80338963	1 (2,27)	deleterious (0)	probably damaging (1)	35.0	0
	c.1204C>T/ p.(Leu402Phe)		2 (4,5)	deleterious (0.01)	probably damaging (0.92)	32.0	No datos
	c.1489C>A/ p.(Arg497Ser)		2 (4,5)	deleterious (0)	probably damaging (1)	31.0	No datos
NCLN	c.-6G>A (5'UTR)		1 (2,27)				No datos
	c.562A>G/ p.(Thr188Ala)		1 (2,27)	deleterious (0,09)	possibly damaging (0.635)	16.37	No datos
	c.974C>T/ p.(Ser325Phe)	771630485	1 (2,27)	deleterious (0)	probably damaging (0.995)	29.0	No datos
	c.1057C>T/ p.(Arg353Trp)		1 (2,27)	deleterious (0)	possibly damaging (0.67)	24.4	0
	c.1562C>T/ p.(Pro521Leu)		1 (2,27)	deleterious (0)	probably damaging (0.998)	33.0	No datos
TDGF1	c.275G>A/ p.(Gly92Glu)		3 (6,77)	deleterious (0)	possibly damaging (0.635)	28.6	No datos
	c.*5C>T (3'UTR)		1 (2,27)				No datos

Al igual que en el caso anterior, el gen en el que encontramos un mayor número de esta clase de variantes fue en *NOMO* donde se identificaron un total de 13 cambios (29,5%). De todas estas variantes de significado incierto/probablemente patogénicas encontradas en *NOMO*, solo una aparecía descrita en la base dbSNP y, en el estudio poblacional ningún probando presentó la mutación. Aun así, ninguna de estas variantes se puede catalogar como mutaciones patogénicas ya que para ello es necesario realizar estudios funcionales en los que se confirme una alteración de la actividad de la proteína que codifica cada gen.

En este caso, el programa ESE FINDER identificó un total de 8 variantes que presentaban diferencias estadísticamente significativas en las regiones ESE con respecto a la secuencia *wild type* (Tabla 36). Como se mencionó anteriormente, estos resultados son orientativos y se requiere una comprobación mediante RT-PCR.

Tabla 36. Predicciones realizadas mediante el programa ESE FINDER en variantes probablemente patogénicas. Marcados con el signo positivo aparecen las zonas en las que la variante analizada puede afectar a la maduración del RNA mensajero debido a que aparece un nuevo motivo ESE y, con signo negativo aparecen las zonas en las que la variante analizada puede afectar a la maduración del RNA mensajero debido a que desaparece el motivo ESE de la secuencia wt. = el programa no predice un cambio en ese sitio.

Gen	Variante	SF2/ASF	SC35	SRp40	SRp55
<i>NOMO</i>	c.118G>T/ p.(Gly40Cys)	-	=	-	=
	c.1517C>A/ p.(Ser506Tyr)	=	-	=	=
	c.3557A>G/ p.(Gln1186Arg)	+	+	=	+
	c.3578G>A/ p.(Gly1193Glu)	-	=	=	+
SMAD2	c.167G>A/ p.(Gly56Glu)	-	=	-	=
SMAD3	c.218G>A/ p.(Gly73Asp)	=	=	+	=
NCLN	c.562A>G/ p.(Thr188Ala)	=	=	-	=
	c.1562C>T/ p.(Pro521Leu)	-	=	=	=

4. 1.3. MUTACIONES PATOGENICAS

En total, hemos identificado 24 mutaciones patogénicas de las cuales 17 son mutaciones *nonsense*, 3 son mutaciones *missense* descritas como patogénicas y 4 son mutaciones intrónicas que se localizan en el primer nucleótido o en el último del intrón y que por tanto afectan al procesamiento del RNA mensajero codificado por el gen donde se localiza el cambio (Tabla 37).

En las mutaciones *nonsense* hay un cambio en una de las bases del DNA de forma que el nuevo triplete que se forma es uno de los tres codones de terminación. Según dónde quede truncada la proteína será capaz de preservar algo de función o no. Pero, en todo caso, como afecta al funcionamiento de la proteína se consideran mutaciones patogénicas. Las mutaciones *missense* consideradas patogénicas en este trabajo, c.1229T>G/ p.Val410Gly (*SMAD3*),

c.1268G>A/p.Ser423Asn (*SMAD3*) y c.1081C>T/p.Arg361Cys (*SMAD4*) son mutaciones descritas por otros grupos como patogénicas (Schepers *et al.* 2018; Houlston *et al.* 1998). En este grupo, también se incluye una variante, c.3651_3652delGA/p.Lys1218AspfsTer31, que se corresponde con una mutación de cambio de fase. Las mutaciones con cambio del marco de lectura implican la inserción o deleción de uno o varios nucleótidos en el que el número de pares de bases eliminado no es divisible por tres. Por tanto, este tipo de mutación interrumpe el marco de lectura y da lugar a una proteína truncada. Y, finalmente, como ya se ha indicado, cuatro variantes intrónicas que afectan al *splicing*.

Tabla 37. Listado de mutaciones patogénicas identificadas en los tumores analizados.

Gen	Variante	Frec. Tumor.n (%)
<i>NOMO</i>	c.1444C>T/p.(Gln482Ter)	1 (4,1)
	c.1692G>A/p.(Trp564Ter)	1 (4,1)
	c.2428G>T/ p.(Glu810Ter)	1 (4,1)
	c.2434G>T / p.(Glu812Ter)	1 (4,1)
	c.2684T>A/p.(Leu895Ter)	1 (4,1)
	c.2833C>T/p.(Gln945Ter)	1 (4,1)
	c.3637C>T/p.(Gln1213Ter)	1 (4,1)
	c.3649C>T/ p.(Gln1217Ter)	1 (4,1)
	c.3651_3652delGA/p.(Lys1218AspfsTer31)	1 (4,1)
<i>SMAD2</i>	c.236+1G>A	1 (4,1)
	c.715C>T/p.(Gln239Ter)	1 (4,1)
	c.923C>A/ p.(Ser308Ter)	1 (4,1)
	c.1389C>A/p.(Cys463Ter)	1 (4,1)
<i>SMAD3</i>	c.76C>T/p.(Gln26Ter)	1 (4,1)
	c.277C>T/ p.(Arg93Ter)	1 (4,1)
	c.1229T>G/p.(Val410Gly)	1 (4,1)
	c.1268G>A/p.(Ser423Asn)	1 (4,1)
	c.*1A>G (3' UTR)	1 (4,1)
<i>SMAD4</i>	c.968G>A/p.(Trp323Ter)	1 (4,1)
	c.1081C>T/p.(Arg361Cys)	1 (4,1)
<i>ActR11</i>	c.601-1G>A	1 (4,1)
<i>NCLN</i>	c.184+1G>A	1 (4,1)
	c.1336C>T/p.(Gln446Ter)	1 (4,1)
<i>ALK4</i>	c.538C>T/ p.(Gln180Ter)	1 (4,1)

Frec. tumor.: frecuencia de la variante en los 36 tumores analizados en este estudio.

El gen en el que se detectó un mayor número de mutaciones patogénicas fue *NOMO* con un total de 9 cambios (37,5%). Todos ellos daban lugar a un codón de parada prematuro generando una proteína truncada. El segundo gen del panel de estudio con mayor prevalencia de mutaciones patogénicas fue *SMAD3* con un total de 5 cambios (20,8%), en este caso se trataba de dos mutaciones *missense*, descritas anteriormente, dos mutaciones *nonsense* dando

lugar al truncamiento de la proteína y una mutación de *splicing*. *SMAD2* presentó un total de 4 mutaciones patogénicas (16,6%), *NCLN* y *SMAD4* presentaron dos en cada uno (8,3) y, finalmente, *ActR11* y *ALK4* con una mutación en cada uno de ellos.

De los 36 tumores analizados, 9 (25%) de ellos tenían mutación patogénica en *NOMO*. Si dividimos estos tumores en función de la edad de diagnóstico y del estado de *NOMO* en la qPCR, se obtuvo (Tabla 38):

-Mayores de 70 años/*NOMO* germinal: se analizó un total de 3 muestras y en uno de ellos se identificó la variante intrónica c.184+1G>A localizada en el gen *NCLN* con una frecuencia alélica de un 15%. En el resto de los tumores no encontramos mutaciones patogénicas en los genes incluidos en el panel de análisis de la vía Nodal.

-Mayores de 70 años/*NOMO* deleciónado heterocigoto: se analizaron un total de 4 muestras pertenecientes a este subgrupo de las cuales, el 50%, presentaban mutación patogénica en el gen *NOMO* (p.Lys1218AspfsTer31 y p.Trp564Ter). Estos dos cambios se presentaban con una frecuencia alélica en el tumor de 46,9% y 12,5% respectivamente. Además, en cada una de estas dos muestras, también se identificó una mutación patogénica en *SMAD3* (p.Ser423Asn y p.Gln26Ter) con frecuencias alélicas de 36,4% y 18,6% respectivamente. Un tercer tumor presentó mutaciones patogénicas en los genes *SMAD4* (p.Trp323Ter, 20%) y *ALK4* (p.Gln180Ter, 21,6%) y, la cuarta muestra perteneciente a este subgrupo presentó mutación patogénica en el gen *NCLN* (p.Gln446Ter, 13,3%).

-Mayores de 70 años/*NOMO* deleciónado homocigoto: en este subgrupo, de las dos muestras iniciales que se incluyeron en el estudio solo se pudo analizar una de ellas ya que la otra no alcanzaba los parámetros óptimos de cobertura para su análisis. En esta muestra se identificó una mutación patogénica en *SMAD3* (p.Val410Gly) con una frecuencia alélica de 40% y la mutación intrónica c.601-1G>A en el gen *ACTR11* con una frecuencia alélica del 10,5%.

-Menores de 45 años/*NOMO* germinal: se han analizado un total de 8 tumores pertenecientes a este subgrupo. De todos ellos, solo 1 (12,5%) presentó mutación patogénica en el gen *NOMO* (p.Gln482Ter, frecuencia alélica en el tumor 12,1%) y otro presentó una variante intrónica, de las que en este trabajo se han considerado patogénicas, en el gen *SMAD2* (c.236+1G>A con una frecuencia alélica del 36,7%). El resto de los tumores no presentó ninguna mutación patogénica en ninguno de los genes analizados en el panel.

-Menores de 45 años/*NOMO* deleciónado heterocigoto: de los 10 tumores pertenecientes a este subgrupo, 5 de ellos (50%) presentaban mutación patogénica en el gen *NOMO* (p.Leu895Ter, frecuencia alélica en el tumor 22,5%; p.Gln1213Ter, frecuencia alélica en

el tumor 27,3%; p.Glu812Ter, frecuencia alélica en el tumor 32,3; p.Gln1217Ter frecuencia alélica en el tumor 23,5 y p.Gln945Ter, frecuencia alélica en el tumor 10,5%), además, uno de estos 5 tumores presentaba una variante intrónica patogénica en el gen *SMAD3* (c.*1A>G (3`UTR), con una frecuencia alélica de 27,3%). En otro tumor se identificó una mutación patogénica en el gen *SMAD3* (p.Ser308Ter, frecuencia alélica en el tumor 17,6%) y, en los 4 restantes, no se identificó ninguna mutación *nonsense* ni variante de cambio de sentido descrita como patogénica. Por otra parte, este fue el subgrupo donde se identificó el mayor número de variantes de significado clínico desconocido catalogadas como patogénicas por los dos programas de predicción utilizados en este trabajo.

-Menores de 45 años/*NOMO* deleciónado homocigoto: dentro de este grupo se han analizado un total de 8 tumores. De todos ellos, tres presentaron mutaciones patogénicas. Uno de ellos en el gen *SMAD2* (p.Cys463Ter, frecuencia alélica en el tumor 31,8%) otro en *SMAD3* (p.Arg93Ter, frecuencia alélica en el tumor 10,5%) y el tercero, mostró dos mutaciones patogénicas en los genes *SMAD2* y *SMAD4* (p.Gln239Ter, frecuencia alélica en el tumor 10,5% y p.Arg361Cys, frecuencia alélica en el tumor 15,6%). Los 5 tumores restantes no presentaron mutaciones patogénicas.

Además, en este estudio también se incluyeron 3 pólipos diagnosticados en pacientes menores de 45 años, todos ellos con *NOMO* normal en el estudio de qPCR y, uno que presentaba pérdida heterocigota en este gen. De estos cuatro pólipos solo se pudieron analizar dos que tenían *NOMO* normal en la qPCR, uno de ellos presentaba una mutación patogénica en el gen *NOMO* (p.Glu810Ter, 37,2%). El otro pólipo no presentó ninguna alteración patogénica ni en *NOMO* ni en ninguno del resto de genes.

Tabla 38. Número de tumores correspondientes a cada uno de los subgrupos establecidos en este trabajo con mutación patogénica en los diferentes genes implicados en la vía de señalización Nodal.

	<i>NOMO</i> n(%)	<i>SMAD2</i> n(%)	<i>SMAD3</i> n(%)	<i>SMAD4</i> n(%)	<i>NCLN</i> n(%)	<i>ActRII</i> n(%)	<i>ALK4</i> n(%)
Mayores de 70 años <i>NOMO</i> normal (n=3)	1 (33,3)						
Mayores de 70 años <i>NOMO</i> delección heterocigota (n=4)	2 (50)		2 (50)	1 (25)	1 (25)		1(25)
Mayores de 70 años <i>NOMO</i> delección homocigota (n=1)			1 (100)			1 (100)	
Menores de 45 años <i>NOMO</i> CNV normal (n=8)	1 (12,5)	1(12,5)					
Menores de 45 años <i>NOMO</i> delección heterocigota (n=10)	5 (50)		2(20)				
Menores de 45 años <i>NOMO</i> delección homocigota (n=8)		1 (12,5)	2 (25)	1(12,5)			
Pólipos menores de 45 años <i>NOMO</i> CNV normal (n=2)	1 (100)						

Es decir, de los 36 tumores analizados, 17 de ellos (47,3%) presentan mutación patogénica en alguno de los genes implicados en la vía Nodal y, de todos estos genes, *NOMO* es el que apareció mutado con mayor frecuencia ya que 9 (53%) de los 17 tumores con mutación, presentaban mutación patogénica en este gen. Además, sorprendió el hecho de que en el 50% de los tumores que presentaban delección heterocigota del gen *NOMO* mediante qPCR, presentaban una mutación patogénica en el otro alelo del gen.

DISCUSIÓN



El CCR en el adulto joven es una enfermedad heterogénea y poco frecuente caracterizada por fenotipos más agresivos del tumor y por un peor pronóstico. En nuestro país representan el 8-10% de los pacientes con CCR y, aunque las tasas de mortalidad han disminuido gracias a los programas de detección precoz y a los avances en terapias dirigidas, la incidencia de esta enfermedad en pacientes jóvenes ha presentado un leve incremento en los últimos años (Ahnen *et al.* 2014).

El CCR surge como resultado de la acumulación de diferentes eventos moleculares que conducen a la pérdida del control sobre la proliferación, diferenciación y/o supervivencia celular. La existencia de diferentes rasgos fenotípicos asociados al CCR en el adulto joven (≤ 45 años) y en el anciano (≥ 70 años), ha llevado a barajar la hipótesis de que tras ellos puedan subyacer diferentes mecanismos moleculares de carcinogénesis (Osorio *et al.* 2014, Kirzin *et al.* 2014).

Existen numerosos estudios genéticos que se basan en la caracterización del perfil molecular del CCR en el adulto mayor, sin embargo, la base molecular de la enfermedad en personas jóvenes continúa siendo poco conocida. Aunque se han sugerido diferencias en los perfiles genómicos de los tumores con diferente edad de diagnóstico, esas diferencias aún no están bien establecidas ni se conocen sus implicaciones (Labots *et al.* 2014).

1. LA REGIÓN 16p13.12-13.11 ESTÁ DELECIÓNADA EN CCR

La variación estructural del genoma humano se debe a cambios genómicos que incluyen deleciones, duplicaciones, inserciones e inversiones. Todos estos cambios genómicos, excepto las inversiones, dan como resultado una variación en el número de copias (CNV) para un segmento genómico dado (Nagamani *et al.* 2011). La disponibilidad de herramientas que permiten estudiar el genoma humano ha revelado hasta qué punto las CNVs juegan un papel crítico en la variación genética humana. Concretamente, la hibridación genómica comparada (aCGH) ha llevado al descubrimiento de muchas alteraciones genómicas (Nagamani *et al.* 2010).

Una de estas CNV con un fenotipo clínico aún no caracterizado es un reordenamiento en la región cromosómica 16p13.11. Si bien las duplicaciones en esta región se asocian con deterioro cognitivo, anomalías de comportamiento y, posiblemente, manifestaciones cardíacas, las deleciones se manifiestan con microcefalia, convulsiones y deterioro cognitivo (Nagamani *et al.* 2011). En cáncer, los cambios en esta región se encontraron en tumores de células epiteloides perivasculares (PEComa) y en cáncer de próstata (Choucair *et al.* 2012).

Nuestro trabajo se basa en un estudio previo de EOCRC empleando la metodología de aCGH, en el que encontramos de manera recurrente alteraciones en la región 16p13.12-p13.11 (Arriba *et al.* 2015).

Aunque existen numerosos estudios en los que se asocian diferentes trastornos con la presencia de alteraciones genómicas en la región 16p13.12-p13.11 (Ramalingam *et al.* 2011), hasta la fecha, estas alteraciones no se han relacionado con la presencia de cáncer. Sengupta *et al.* en 2013 analizaron, mediante aCGH, el perfil cromosómico de tumores EOCRC de pacientes británicos de origen bangladesí, identificando pequeñas deleciones en la región 16p13.2. Esta región del genoma codifica *RBFOX1*, una proteína de unión al RNA altamente conservada implicada en el *splicing* alternativo. Encontraron que, en un elevado número de CCR, las regiones reguladoras del gen y el extremo 5' se perdían, lo que conducía a una expresión de *RBFOX1* aberrante. Además, las anomalías en este gen se han relacionado con otros tipos de tumores como el melanoma. Sin embargo, este gen no se localiza en la región del cromosoma 16 que aparecía delecionada en el 23,3% de los tumores de CCR en nuestro estudio previo.

Los datos obtenidos en nuestro trabajo revelan que las únicas diferencias significativas, al comparar las características clínico-patológicas, moleculares, familiares y de seguimiento, entre el grupo de pacientes que presentaban la deleción en la región cromosómica 16p13.12-p13.11 y el grupo que no presentaba esta anomalía fueron: la localización del tumor, las CNAs y una mayor supervivencia del grupo de pacientes que presentaban la deleción en esta región cromosómica (Tablas 20 y 21). Estos datos sugieren que esta deleción puede ser un factor pronóstico en CCR. Hasta la fecha, existen numerosos factores pronósticos asociados a cáncer colorectal; por ejemplo, hay evidencias de que el diagnóstico a edades tempranas se asocia, por lo general, con un tumor más agresivo y de peor pronóstico (Luy *et al.* 2000). El estadio clínico de la enfermedad es un factor fundamental para el pronóstico, así como el número de ganglios afectados, la localización del tumor, la presencia de perforación intestinal, la invasión a otros órganos etc. (Yang *et al.* 2018).

El EOCRC con deleción de la región 16p13.12-p13.11 presenta anomalías en la longitud telomérica

Los telómeros son regiones de DNA no codificante situados en los extremos de los cromosomas eucariotas formados por repeticiones en tándem de secuencias cortas de nucleótidos. Su principal función es el mantenimiento de la integridad y la estabilidad genómica (Blackburn *et al.* 1991; Rodier *et al.* 2004). La telomerasa es una ribonucleoproteína cuya función enzimática es la de añadir repeticiones teloméricas *de novo* para así poder mantener una

longitud telomérica estable a lo largo de la historia replicativa de determinados tipos celulares (Ulaner *et al.* 1998; Kyo *et al.* 1999). La telomerasa es muy abundante en las células tumorales, donde mantiene el crecimiento y división indefinida de las células cancerosas, convirtiéndolas en inmortales. Las células sanas normalmente apenas expresan la telomerasa y, en consecuencia, sus telómeros se acortan progresivamente en los sucesivos ciclos de división celular, hasta que alcanzan una longitud críticamente corta que desencadena la apoptosis celular o la senescencia (Rodier *et al.* 2004).

En algunas lesiones premalignas e inflamatorias de colon se ha observado una sobreexpresión de la telomerasa. En CCR, en el tejido tumoral la longitud de los telómeros es más corta que en el tejido sano adyacente (Engelhardt *et al.* 1997).

En este trabajo, analizamos la relación entre la media de la longitud telomérica y diferentes subgrupos de pacientes con cáncer colorectal establecidos en función del estado de la región 16p y de la edad (Figura 28). Al comparar la media de la longitud telomérica del grupo de tumores diagnosticados en pacientes ancianos y jóvenes sin la delección 16p13.12-p13.11, la longitud telomérica de los ancianos es significativamente menor que la de los jóvenes. Esto es debido a que la longitud de las repeticiones teloméricas se acorta progresivamente con el aumento de la edad del organismo como consecuencia de la multiplicación celular necesaria para regenerar los tejidos. Esta observación sugiere que las células tumorales mantienen el tamaño del telómero, en el momento en el que se inicia la transformación tumoral, más corto cuanto mayor es la edad del paciente en el que se origina el tumor.

Al comparar la media de la longitud telomérica del grupo de tumores procedentes de ancianos y tumores procedentes de jóvenes con la delección 16p13.12-p13.11, no se obtuvieron diferencias significativas. Es más, cuando se comparó la media de la longitud telomérica de los tumores de jóvenes con la delección y tumores de jóvenes sin la delección, los telómeros del grupo con la delección eran significativamente más cortos que los del grupo sin la delección. Este hecho podría sugerir que la presencia de la delección de la región 16p13.12-p13.11 se asocia a un silenciamiento de la telomerasa en lugar de a una sobreexpresión y, por tanto, a una mayor inestabilidad cromosómica, dato que concuerda con nuestros resultados puesto que el grupo de tumores con la delección mostraba un alto índice de inestabilidad cromosómica (Fang *et al.* 2017). Además, estudios recientes revelan que pacientes con CCR que presentan los telómeros más cortos en las células tumorales tienen un pronóstico más favorable y ninguno de ellos sufrió una recurrencia durante el seguimiento de la enfermedad (Fernández-Marcelo *et al.* 2016). Nuestros resultados apoyan esta observación, puesto que los pacientes que presentan la delección en la región 16p13.12-p13.11, tienen una longitud telomérica significativamente

menor que los que no tenían la deleción y, a su vez, este grupo de pacientes tienen un pronóstico más favorable.

2. EL GEN *NOMO* ESTÁ IMPLICADO EN EOCRC

Mediante hibridación *in situ* por fluorescencia, confirmamos la deleción detectada por aCGH en un total de 5 tumores. Gracias a esta técnica, pudimos observar que los tumores incluidos en el estudio presentaban una mínima región en común, con un tamaño de 614.838 bp, donde había pérdida de material genético. Dentro de esta región, se localizan los genes: *BFAR*, *PLA2G10*, *NPIPA2*, *ABCC6P2*, *NOMO1*, *PDXDC1*, *NTAN1*, *RRN3* y *PKD1P6*. Dado que, de todos estos genes, *NOMO* juega un importante papel como antagonista de una vía implicada en carcinogénesis, la vía de señalización Nodal, se decidió analizar de una manera detallada este gen en tumores de colon.

Aunque en las bases de datos se describen tres genes *NOMO*: *NOMO1*, *NOMO2* y *NOMO3*, nuestro trabajo pone de manifiesto la existencia de un único gen, descartando la existencia de tres proteínas con la misma función como se describe. Al analizar los datos obtenidos en un ensayo de arrays de expresión, observamos que el programa informático tiene registrado los tres genes, sin embargo, no diferencia *NOMO2* de *NOMO3* y los analiza como un único gen porque identifica una única señal (Tabla 27). Además, el hecho de que en diferentes especies como por ejemplo en el ratón, sólo existe un gen *NOMO*, sugiere fuertemente la presencia de un único gen. Por tanto, podría ser que los genes *NOMO2* y *NOMO3* sean el resultado de problemas de anotación en el genoma.

Nosotros hemos observado que el gen *NOMO* no sólo está delecionado en los casos que presentan pérdida de material genético en la región 16p13.12-13.11 por aCGH y FISH, sino también en tumores diagnosticados en pacientes jóvenes que no presentan la alteración citogenética. Así, el 32% de los tumores analizados presentan pérdida del gen, siendo esta deleción más recurrente en la población joven. Más del 70% de los tumores pertenecientes al grupo EOCRC presentan deleción del gen *NOMO*, mientras que en el grupo de los ancianos y en el grupo de edad intermedia, tan sólo la presentan el 5% de los tumores analizados. El bajo número de tumores con pérdida de *NOMO* en el grupo de edades intermedias puede atribuirse a que la mayoría de los pacientes incluidos en nuestro estudio presentaban edades superiores a los 60 años. De hecho, la edad más frecuente del grupo fue 64 años y la media 65,7 por lo que no se puede considerar un grupo representativo de edades intermedias, sino que se asemeja mucho más al grupo de pacientes ancianos.

En este trabajo se describe por primera vez la alta tasa de tumores EO CRC con pérdida del gen *NOMO*, lo que sugiere que este gen podría ser un posible biomarcador en tumores de inicio precoz. Sin embargo, hasta la fecha no hay evidencias de que *NOMO* esté relacionado con el desarrollo de CCR.

Cuando clasificamos los tumores EO CRC en dos grupos: con estabilidad de microsatélite (MSS-EO CRC) y con inestabilidad de microsatélite (MIS-EO CRC), se observó que más del 90% de los tumores con estabilidad de microsatélite tenían delección homocigota de *NOMO* y, por el contrario, en los que había inestabilidad de microsatélite el porcentaje de tumores con *NOMO* homocigoto descendió hasta un 44%. En la interpretación de estos datos, hay que tener en cuenta que de los 75 tumores de los que disponíamos de datos acerca de la inestabilidad de microsatélite, la mayor parte de ellos (59/75; 79%) presentaban estabilidad de microsatélite y, tan sólo un 21% (16/75) eran inestables, por lo que no es un grupo equilibrado. Estos resultados aportarían datos importantes para la caracterización de los tumores MSS-EO CRC ya que la patogénesis de los tumores MSI-EO CRC está bien caracterizada en individuos con Síndrome de Lynch en los que existe una mutación en línea germinal en algún gen del sistema MMR, que se relaciona con un aumento en la susceptibilidad a desarrollar cáncer de colon; sin embargo, el conocimiento acerca de las características moleculares del EO CRC esporádico es bastante más limitado.

Teniendo en cuenta el elevado porcentaje de tumores que mostraban delección homocigota del gen *NOMO* en individuos diagnosticados a edades tempranas, nos planteamos que la pérdida en uno de los alelos podría darse en línea germinal y una segunda pérdida a nivel somático. La ausencia de delección heterocigota del gen *NOMO* en las 13 muestras de sangre periférica procedentes de pacientes con delección homocigota del gen en el tumor, indica que la delección de *NOMO* es de origen somático, por lo que podemos especular con que la pérdida de un alelo del gen *NOMO* puede estar implicado en las primeras etapas de la carcinogénesis colorrectal. El estudio del gen *NOMO* en pólipos adenomatosos mostró delección heterocigota en el 44% de los casos, lo que parece reforzar esta opción. Es necesario aumentar el número de pólipos analizados para poder confirmar estos resultados. De confirmarse, la delección heterocigota en el pólipo adenomatoso podría considerarse un posible marcador predictivo de EO CRC.

El estudio simultáneo de tejido tumoral y tejido sano adyacente mostró pérdida homocigota del gen *NOMO* en DNA extraído de tejido peritumoral. El concepto de cancerización de campo surgió a partir de la observación de que los pacientes de ciertos tipos de cánceres son propensos a desarrollar otras neoplasias malignas del mismo tipo de tejido cerca del cáncer

primario a lo largo del tiempo. (Gallegos-Villanueva et al. 2007; Soo-Kyung et al. 2016). El concepto actual de cancerización de campo comprende el pre-acondicionamiento de un área del epitelio al crecimiento tumoral, ya sea como resultado de una proliferación clonal de células mutantes a través del epitelio o debido a cambios constantes en las células estromales. Se puede observar este fenómeno en casi todas las superficies epiteliales, entre las que se incluyen orofaringe, esófago, estómago, pulmón, colon, ano, cérvix, tejido sanguíneo y piel (Hawtorn et al. 2014). La delección de *NOMO* en tejido adyacente al tumor puede deberse a la cancerización de campo.

Con el fin de determinar si la pérdida de *NOMO* se asociaba al EOCRC exclusivamente, analizamos este gen en otros tipos de tumores: ovario, endometrio, mama, tumores de cabeza y cuello y glioblastomas; todos ellos diagnosticados en pacientes jóvenes. De los 69 tumores analizados, solo uno (1,4%) presentó delección homocigota de *NOMO*.

Diferentes estudios acerca del perfil cromosómico en distintos tipos de tumores revelan que en ninguno existen alteraciones de material genético en la región 16p13.12-13.11. Por ejemplo, en cáncer de mama diagnosticado en pacientes menores de 40 años se han identificado ganancias en las regiones cromosómicas 1q, 6p, 8q, 11q, 16p, 17q, 19 y 20q y pérdidas en 6q, 13q, 16q, 17p y 22q (Ueno et al. 2012; Pascual A. 2017). Un porcentaje importante de tumores de cabeza y cuello incluyen delecciones en las regiones cromosómicas 1p, 4, 5q, 6q, 8p, 9p, 11, 13q, 18q, y 21q y sobreexpresión en las regiones 3q, 8q, 16p, 17q, 19, 20q, y 22q (Bockmühl et al. 1998). En los tumores de ovario las regiones con ganancia de material genético más frecuentes son 1q21.3 y 8q24.3 y, sin embargo, existen muchas regiones con pérdidas de material genético como 3p21.31, 4q12, 5q13.2, 5q23.2, 5q31.1, 7p22.1, 7q11.23, 8p12, 9p22.1, 11p15.1, 12p13.31, 15q11.2, 15q21.2, 18p11.31 y 22q11.21 (Sung et al. 2013). En los tumores de endometrio, se han encontrado pérdidas en las regiones 7p y en el cromosoma 19 y ganancias en 1q 6p y 8q (Flicker et al. 2015).

Estas observaciones sugieren que la delección del gen *NOMO* puede ser patognómica en EOCRC.

En nuestro trabajo hemos inactivado el gen *NOMO* en la línea celular HT29 mediante la tecnología de edición génica CRISPR/Cas9. Para ello, se diseñaron tres guías en el exón 3 del gen ya que, al cortar en una región inicial, queda mucha secuencia para que se genere un codón de parada prematuro y de lugar a una proteína defectuosa. Además, otro motivo por el que se seleccionó este exón es porque el último nucleótido de la región codificante codifica un aminoácido con el primer nucleótido del exón 4 por lo que, al producir un corte en la hebra de

DNA en esta zona, se produce un cambio en el marco de lectura y con ello una proteína truncada. La transfección se hizo con combinación de dos guías por tres motivos principales: el primero es porque al introducir dos guías aumentan las posibilidades de que, al menos una de ellas, produzca el corte; el segundo es porque al estar las dos guías consecutivas y producirse los dos cortes, se pierde el fragmento del gen comprendido entre ellas y el sistema de reparación celular es incapaz de reparar el daño, y el tercer motivo de transfectar las células con combinación de dos guías es porque de este modo se aumenta la especificidad hacia el gen diana evitando las indeseadas alteraciones *of target* (fuera de la secuencia diana).

En su uso como herramienta de edición genómica, el sistema CRISPR/Cas9, puede ser programado para desempeñar distintas funciones con diferentes efectos, principalmente, disrupciones mediante inserciones y deleciones (InDel), modificaciones o ediciones endógenas y activación o represión de la expresión génica (Mojica *et al.* 2013; Ran *et al.* 2013).

En nuestro caso, esta técnica ha sido muy útil, sencilla y económica para inactivar el gen *NOMO* en la línea celular HT29. Sin embargo, la eficiencia final de la técnica fue muy baja, de todas las células separadas con el plásmido transfectado mediante *sorting*, únicamente crecieron 5 clones, es decir, el 0,5% de las células separadas. Hay que tener en cuenta que, aunque la técnica sea sencilla de desarrollar, su eficacia puede alterarse en las diferentes etapas que conlleva su desarrollo: en primer lugar, de todas las células de las que se dispone para transfectar, tan solo se consigue en menos del 10% aun con el método de transfección óptimo; y, en segundo lugar, las posibilidades de que una célula aislada sea capaz de sobrevivir y de dividirse son muy bajas, muriendo la mayoría de ellas.

Es conocido que la edición con el sistema CRISPR-Cas9 no es específico de la zona diana y puede afectar a otras regiones. Por ello, en nuestro trabajo, comprobamos, mediante la técnica aCGH, los posibles efectos *off-target* del sistema CRISPR-Cas9. Comparando el perfil genómico de la línea celular HT29 sin editar genéticamente con la línea HT29 modificada por CRISPR-Cas9, observamos que no sólo hay pérdida de material genético en la región cromosómica donde se localiza *NOMO*, si no que hay deleción heterocigota de todo el cromosoma 16. Además, existen diferencias respecto a ganancias y pérdidas de material genético en diferentes regiones de los cromosomas entre ambas líneas como, por ejemplo, la amplificación de la región Xp22.2-p12.2 y Xp22.11 y pérdidas en 22q11.23 en la línea editada genéticamente. Por tanto, nuestro trabajo confirma que el sistema CRISPR/Cas9 introduce mutaciones no deseadas cuya entidad es considerable, teniendo lugar grandes deleciones y complejos reordenamientos, lo que puede comprometer gravemente la función genética y dificulta el estudio de los efectos biológicos de la edición.

Estos efectos *off-target* se dan porque en el organismo de origen del sistema CRISPR/Cas9, su función de defensa contra elementos nucleotídicos externo, permite el reconocimiento y destrucción de DNA viral o plasmídico hipervariable; sin embargo, esta relativa flexibilidad en el reconocimiento de secuencias variables, cuando el sistema se saca de su contexto original y se dedica a la modificación genómica, genera mutaciones indeseadas más allá de la región diana y resulta en efectos colaterales indeseados (Long *et al.* 2014; Ma *et al.* 2014).

La inactivación de *NOMO* en la línea celular HT29 no modifica la supervivencia celular, sin embargo, induce la formación de esferas y una mayor capacidad migratoria. Este fenotipo de formación de esferas y mayor migración celular se ha asociado en diferentes tipos de tumores a una sobreactivación de la actividad de la vía de señalización Nodal. Por ejemplo, en un estudio realizado por Wenchen *et al.* en 2017, se observó que la línea celular MCF-7 derivada de un cáncer de mama con sobreexpresión de la proteína Nodal formaba mayor número de esferas y, además, estas esferas presentaban mayor tamaño que cuando la línea celular no tenía sobreexpresión de la proteína Nodal. Este fenotipo se revertía en presencia de un inhibidor de la vía. En otros trabajos en cáncer de próstata, páncreas o melanoma, también se ha descrito que la sobreexpresión de la proteína Nodal favorece la formación de un mayor número de esferas y confiere mayor capacidad de migración (Duan *et al.* 2015; Lawrence *et al.* 2012).

Por tanto, el fenotipo de los clones con *NOMO* inactivado se correspondería con una sobreactivación de la vía Nodal lo que concuerda con el papel inhibidor en esta vía del complejo proteico del que forma parte *NOMO*.

3. NCLN Y *NOMO* SON ESENCIALES PARA MANTENER LA INTEGRIDAD DEL COMPLEJO

Las proteínas NCLN, *NOMO* y TEMEM147 forman un complejo transmembrana localizado en el RE. La disminución de los niveles de una de estas proteínas resulta en la desestabilización de las otras dos, siendo considerada hasta el momento NCLN como la reguladora en la formación del complejo ya que, además de controlar los niveles de expresión de *NOMO*, también controla los niveles de expresión de TEMEM147 (Haffne *et al.* 2004). No obstante, un bloqueo en la expresión de cualquiera de las tres proteínas resulta en una fuerte reducción en la expresión de las otras dos (Haffne *et al.* 2007). Sin embargo, una sobreexpresión de *NOMO* y de TEMEM147, solos o en combinación, no produce un aumento en los niveles de expresión de NCLN, lo que sugiere que NCLN es el factor limitante en la formación del complejo proteico (Dettmer *et al.* 2010).

Por todo esto, se ha propuesto un mecanismo de regulación en el que NOMO se une a la NCLN a través de su dominio luminal y, posteriormente, se une TEMEM147 a través del dominio citoplasmático. El exceso de NOMO y de TEMEM147 es sus formas monoméricas es degradado por el proteosoma (Haffne *et al.* 2004; Haffne *et al.* 2007; Dettmer *et al.* 2010).

En este trabajo, hemos analizado los niveles de expresión de las proteínas NCLN y TEMEM147 en una línea celular con *NOMO* inactivado. Nuestros resultados concuerdan con los descritos previamente, observando que el bloqueo en los niveles de expresión de *NOMO* resulta en la desestabilización de las otras dos proteínas, lo que nos permite sugerir que estas proteínas son inestables cuando no están unidas entre si y que la formación del complejo tiene un efecto estabilizador (Dettmer *et al.* 2010). Además, mostramos que no solo la proteína NCLN, sino también NOMO es esencial para la integridad del complejo. Es necesario hacer estudios de localización para confirmar si, al no formar parte del complejo, las proteínas se degradan o quedan secuestradas en algún compartimento celular.

La conservación evolutiva de las proteínas que forman el complejo proteico NCLN/NOMO/TEMEM147 en metazoos y plantas hace pensar que estas desempeñan un papel muy importante. En embriones de pez cebra, la expresión ectópica de NCLN y NOMO, pero no de cada proteína por separado, causan ciclopia (Dettmer *et al.* 2010). La sobreexpresión de ambas proteínas está implicada en un proceso que altera el desarrollo de algunas zonas del cerebro. Además, un bloqueo en la función de NOMO aumenta de manera dependiente de la dosis, la cantidad de derivados mesodérmicos; es decir, el bloqueo de NOMO en embriones de pez cebra, induce un aumento de la actividad de la vía de señalización Nodal (Haffner *et al.* 2004)

Nosotros pensamos que la posible implicación en carcinogénesis del gen y la proteína NOMO pueden deberse a la desestabilización del complejo transmembrana.

4. LA VÍA NODAL NO SE MODIFICA EN AUSENCIA DE NOMO

Los componentes de la vía Nodal rara vez se expresan en tejido adulto (Bodenstine *et al.* 2016). En los últimos años se ha comunicado la desregulación de la vía Nodal en diferentes tipos de tumores como el melanoma, el cáncer de mama, el carcinoma de endometrio, el cáncer de vejiga y el cáncer de próstata, entre otros (Zhong *et al.* 2017). Así, se ha descrito una sobreexpresión de la actividad de esta vía en tumores metastásicos humanos y el bloqueo de esta vía de señalización inhibe la invasión celular, la formación de colonias y la tumorigénesis en algunos tipos de tumores (Topczewska *et al.* 2006). Por ejemplo, se ha observado que un elevado número de tumores de mama tienen sobreexpresión de la proteína Nodal, la cual se correlaciona con una sobreexpresión de diferentes marcadores de células madre tumorales,

sugiriendo la importancia de esta vía en el mantenimiento de las células madre. Además, los pacientes cuyos tumores presentaban sobreactividad de la vía Nodal tienen menor supervivencia (Gong *et al.* 2016; Gong *et al.* 2017). Por el contrario, la reducción de la expresión nodal disminuye la tumorigenicidad y la metástasis de las células de cáncer de mama en múltiples modelos *in vivo* (Gong *et al.* 2017). Es decir, la sobreactividad de esta vía en cáncer de mama se correlaciona con una mayor progresión tumoral, metástasis y un peor pronóstico (Gong *et al.* 2016). En el caso del cáncer de páncreas, un importante número de tumores tienen sobreexpresión de la proteína Nodal y, por tanto, muestran una sobreactividad de esta vía mediada por las proteínas Smads. Esta sobreexpresión es mayor en tumores de páncreas pobremente diferenciados (Jie *et al.* 2014). Es decir, en muchos tumores hay un aumento en la actividad de la vía Nodal que se correlaciona con un fenotipo tumoral más agresivo, pobremente diferenciado e incluso con un fenotipo metastásico (Lawrence *et al.* 2012).

En el caso del CCR, se ha observado una sobreexpresión de varias proteínas que participan en la vía de señalización Nodal como es el caso del receptor ALK4 y, sobre todo, la sobreexpresión de la proteína que actúa como correceptor de la vía, CRIPTO1. La expresión de estas proteínas es mayor en tejido tumoral que en el tejido adyacente no tumoral, lo que refleja que el aumento en la actividad de la vía Nodal está relacionado con la carcinogénesis (Gong *et al.* 2014).

Dado que la vía de señalización Nodal juega un importante papel en la tumorigénesis colorrectal y que el complejo proteico NCLN/NOMO/TEMEM147 funciona como antagonista de esta vía de señalización, nos propusimos analizar la expresión de las proteínas integrantes de la vía Nodal en la línea celular HT29 con el gen NOMO inactivado para ver si el bloqueo de NOMO conlleva a un aumento en la actividad de esta vía. En este trabajo hemos analizado la expresión de las proteínas: ALK4, ACTR2, CRIPTO, SMAD2, P-SMAD2, SMAD3, P-SMAD3 y SMAD4. En ninguno de los casos observamos modificación en la expresión de estas proteínas.

Estos resultados indican que la posible acción carcinogénica de NOMO no está mediada por su efecto sobre la vía Nodal, por lo que es necesario profundizar en los cambios moleculares que acompañan a la pérdida de NOMO para caracterizar la vía por la que este gen está implicado en el desarrollo tumoral. Se ha demostrado la existencia de señalización de la familia de proteínas TGF β (familia a la que pertenecen las proteínas Nodales) a través de proteínas no Smads (señalización no canónica) aunque sigue siendo poco conocida. Esta señalización puede estar mediada por receptores tipo I y tipo II e implica la activación de vías como p38, JNK, ERK, PI3K y Rho. Estas vías de señalización pueden actuar de forma independiente o en cooperación con Smads (Zhang *et al.* 2009). Por ejemplo, se ha demostrado que ERK y TGF β 1 actúan con

otros factores para inducir vías como la transición epitelial a mesenquima (EMT), uno de los procesos más intensamente estudiados en biología del cáncer. Estudios recientes en cáncer de mama han demostrado una relación entre los efectos protumorigénicos de Nodal y la vía ERK. Se demostró en este estudio que la inhibición de ERK reduce la capacidad de Nodal para aumentar la expresión de c-myc y regular negativamente p27, lo que sugiere que Nodal actúa regulando aspectos del ciclo celular a través de ERK, independientemente de la señalización de Smad (Quail *et al.* 2012). Es decir, se ha demostrado que diferentes ligandos de la familia de proteínas TGF β activan otras cascadas de señalización independiente de los Smads por lo que no podemos descartar que las proteínas Nodales estén implicadas en estas vías y el complejo inhibidor NCLN/NOMO/TEMEM147 actúe en alguna de ellas.

Consideramos relevante apuntar que la vía de señalización Nodal solo existe en vertebrados, aunque ortólogos de NCLN, NOMO y TEMEM147 también están presentes en invertebrados, lo que indica que el complejo podría cumplir funciones adicionales como, por ejemplo, participar en el ensamblaje del receptor nicotínico de acetilcolina (nAChR) en invertebrados (Dettmer *et al.* 2010). Sin embargo, debido a las diferencias que existen entre los sistemas de nAChR de vertebrados e invertebrados, la relevancia de estos hallazgos no es concluyente (Haffner *et al.* 2007).

Para seguir profundizando en el papel que juega NOMO en el CCR, se llevó a cabo un análisis del transcriptoma de los dos clones que se obtuvieron con NOMO inactivado. En un análisis preliminar de las rutas desreguladas al bloquear NOMO, llama la atención la vía de NOD. Esta vía está implicada en la enfermedad inflamatoria intestinal de Crohn, aumentando esta enfermedad el riesgo a desarrollar CCR (Huhn *et al.* 2018). Por todo ello, consideramos que es necesario continuar estudiando el papel que juega NOMO en CCR para intentar identificar dónde actúa esta proteína y aportar nuevos avances moleculares en EOCRC.

5. LOS GENES IMPLICADOS EN LA VÍA DE SEÑALIZACIÓN NODAL ESTÁN MUTADOS EN CCR

La incorporación de las nuevas tecnologías de secuenciación masiva ha supuesto una revolución en el abordaje de muchos tipos de enfermedades, entre ellas el cáncer (Calabria *et al.* 2016). El uso de datos genéticos en la práctica clínica ha supuesto una revolución médica y científica en todo el mundo, permitiendo el desarrollo de un modelo de medicina de precisión (Calabria *et al.* 2016). La metodología NGS (Next Generation Sequencing) es idónea para abordar la complejidad mutacional de los tumores; además, permite la detección de alteraciones de tipo somático presentes en porcentajes bajos en las muestras e incluso detectar inversiones, una

clase de variación cuyo estudio resulta muy complicado para la mayoría de las otras técnicas (Rodríguez *et al.* 2012). La NGS es de gran relevancia para fines diagnósticos como alternativa a las técnicas de secuenciación de primera generación, las técnicas citogenéticas moleculares y de CGH arrays.

Respecto a las técnicas convencionales, el avance más importante de la NGS es la enorme cantidad de datos que puede producir, con el consecuente reto de poder almacenarlos y procesarlos. Debido a ello, es imprescindible el análisis de todos estos datos mediante softwares específicos (Wilhelm *et al.* 2009). La correcta interpretación de las variantes encontradas o cambios detectados es clave para el diagnóstico genético; sin embargo, la determinación de variantes por sí misma es insuficiente, siendo necesaria la interpretación de estas para poder determinar sus implicaciones moleculares y clínicas de manera precisa para un óptimo abordaje terapéutico (Matthijs *et al.* 2016; Yang *et al.* 2018).

En nuestro trabajo se diseñó un panel específico para el análisis de variantes en los genes que forman el complejo inhibidor NCLN/NOMO/TEMEM147 y en los genes integrantes de la vía Nodal. Dicho panel no aseguraba una cobertura completa de todos los genes incluidos en el estudio por diferentes motivos: en algunos de los genes (*ALK4*, *ActR1I*, *NOMO*, *TEMEM147*) hay una región de repeticiones que afecta al rendimiento de la secuenciación de algunos amplicones de estos genes y, además, en determinadas regiones de algunos de estos genes, había un alto contenido en CG lo que también afecta al anillamiento de los oligos para la amplificación de los genes. Por todo esto y por los problemas propios asociados a la secuenciación masiva de muestras en parafina, como la concentración y la calidad óptima de las muestras, de los 50 tumores incluidos inicialmente en el estudio, solo 36 de ellos se secuenciaron con la cobertura requerida por la técnica.

Además, en nuestro trabajo solo hemos considerado las variantes con frecuencia igual o superior al 10% en el tumor, localizadas en regiones con cobertura óptima y las que aparecían en las dos cadenas de DNA y, dentro de todas estas, se han seleccionado únicamente aquellas que pasaban todos los filtros de calidad impuestos por los programas de análisis. Es decir, no se puede descartar el hecho de que además de las variantes consideradas en este trabajo existan otras que no se han tenido en cuenta por no cumplir alguno de los requisitos impuestos.

Todas las mutaciones identificadas mediante secuenciación masiva deberían confirmarse a través de otra técnica. Por ejemplo, en los estudios genéticos realizados en sangre periférica, las mutaciones patogénicas detectadas mediante NGS se confirman por Sanger. Sin embargo, en el caso de las mutaciones somáticas, como un elevado porcentaje de ellas tienen

una frecuencia alélica baja, no serían detectables mediante la secuenciación por Sanger. En la actualidad, no disponemos de ninguna técnica capaz de confirmar las mutaciones somáticas que aparecen con baja frecuencia alélica. Por ello, no se puede descartar que, en realidad, alguna de estas mutaciones sea un artefacto (esta opción es poco probable ya que los artefactos se suelen repetir en varias muestras dentro de la misma carrera) o un falso positivo.

De las 286 variantes diferentes identificadas en este estudio, 150 se corresponden con cambios neutrales (Tabla 30). Dentro de este grupo incluimos polimorfismos, en este tipo de variantes hay un cambio de una de las bases de DNA, de tal manera que el triplete de nucleótidos se cambia, pero incluso si se necesita un cambio de aminoácido, el aminoácido que entra en el lugar en cuestión resulta tener poco o ningún impacto en la función de la proteína. Las variantes identificadas como polimorfismos aparecen en más del 1% de la población control (Karki *et al.* 2015). También se incluyeron todas las variantes silenciosas, en este tipo de mutación hay un cambio en una de las bases del DNA de forma que el triplete de nucleótidos se modifica, pero sigue codificando para el mismo aminoácido y, variantes localizadas en las regiones intrónicas alejadas de los genes por lo que no afectan al procesamiento del RNA mensajero (Karki *et al.* 2015).

Cabe señalar que, dado que muchas de las variantes que hemos clasificado como neutrales no superan el 1% de frecuencia poblacional, es necesario realizar estudios funcionales para verificar su neutralidad.

Además, se han identificado 112 variantes de significado clínico desconocido. El desconocimiento del efecto que, sobre la función y la estructura, provocan las variantes de significado desconocido impiden obtener un resultado informativo en el análisis. Puesto que los estudios funcionales resultan demasiado costosos en tiempo y esfuerzo; se suele realizar un estudio aproximativo habitualmente no concluyente. Los estudios de clasificación de las variantes de significado desconocido se centran, principalmente, en estudios de la frecuencia con la que aparece la variante en la población sana, conservación entre especies, pérdida de heterocigosidad, estudios bioinformáticos, etc.

En este trabajo hemos clasificado las variantes de significado clínico desconocido en tres grupos en función de los resultados obtenidos con los programas de predicción utilizados: probablemente neutrales (tipo 2) (Tabla 31), variantes de significado incierto (tipo 3) (Tabla 33) y variantes probablemente patogénicas (tipo 4) (Tabla 35).

Para confirmar todas estas variantes es necesario realizar estudios funcionales para comprobar si dichos cambios afectan o no a la funcionalidad de la proteína y poderlas catalogar definitivamente como neutrales o patogénicas.

Dentro de este grupo de variantes de significado clínico desconocido se han incluido las variantes identificadas en los extremos UTR de alguno de los genes incluidos en el panel. Existen numerosos estudios en los que se demuestra que los extremos UTR pueden jugar un papel importante en el control post-transcripcional de la expresión de un gen, en la localización celular y subcelular del mRNA, y en el control de la eficiencia de traducción del mRNA (Chen et al., 2006). Por todo ello, es necesario realizar estudios sobre el RNA para comprobar que el procesamiento de estos genes es normal y las muestras de parafina incluidas en nuestro trabajo dificultan esta aproximación metodológica.

Algunas de estas variantes de significado clínico desconocido pueden afectar a la función génica alterando la precisión o velocidad en que un mRNA es traducido, plegado o procesado e incluso afecta al plegamiento de la proteína. Hoy día sabemos que algunas de estas secuencias pueden modular el procesamiento del RNA al estar localizadas en regiones reguladoras del procesamiento como los sitios ESE (*exonic splicing enhancer*) o ISE (*intronic splicing enhancer*). La regulación que ejercen las secuencias ESE se produce, entre otras, a través de las proteínas SR, que tienen la capacidad de reconocer y unirse a los sitios ESE y reclutar a los componentes de la maquinaria de maduración del RNA (Supek *et al.* 2014). Las mutaciones que alteren alguna de estas secuencias pueden afectar a la unión de las proteínas SR y, consecuentemente, disminuir la eficacia en la maduración del RNA mensajero. La mayor densidad de los sitios ESE importantes se localizan tanto en los primeros 20-25 nucleótidos del extremo 5' como en los del extremo 3'; en consecuencia, mutaciones localizadas en estas zonas tienen una mayor probabilidad de resultar dañinas para el procesamiento del RNA. Sin embargo, los SNPs aparecen distribuidos por los exones con menor frecuencia en las regiones próximas a las zonas de consenso de procesamiento del RNA mensajero y con una alta frecuencia en el interior en un intento de mantener la estabilidad del procesamiento del RNA (Fairbrother *et al.* 2004). En nuestro trabajo, el programa ESE FINDER encontró que, de las 122 variantes de significado clínico desconocido, 30 generan algún nuevo sitio ESE de unión de alguna de las proteínas SR o eliminan un sitio ESE de la secuencia germinal al introducir el cambio (Tabla 32, Tabla 34 y Tabla 36), por lo que estas mutaciones podrían modificar el nivel del procesamiento del RNA. Sin embargo, conviene resaltar que estos resultados son orientativos y se requiere una comprobación mediante RT-PCR.

Es necesario tener en cuenta que la mayoría de los genes incluidos en el panel cuentan con muy pocos estudios publicados hasta el momento y que la mayoría de las variantes encontradas en este trabajo no están caracterizadas. En la actualidad, debido al impulso de los avances tecnológicos, la capacidad de secuenciar es mayor que la capacidad para interpretar las variantes que se detectan. De hecho, los efectos de las variantes de significado incierto y de las variantes en regiones intrónicas plantean un desafío para su interpretación. Además, los métodos utilizados en la alineación o el filtrado de variantes influyen considerablemente en su detección. Por tanto, se recomienda utilizar más de un protocolo de análisis independiente y reevaluar los casos no resueltos para minimizar los falsos negativos (Le Gallo *et al.* 2017). Por otro lado, confiar exclusivamente en las herramientas de interpretación *in silico* es peligroso, no son el reemplazo de los estudios funcionales y, por tanto, por si solas no pueden clasificar una variante.

Por último, en este estudio hemos encontrado 24 mutaciones patogénicas, de las cuales 3 son mutaciones de cambio de sentido (*missenses*) descritas por otros grupos como patogénicas. Las variantes **c.1229T>G/ p.Val410Gly** y **c.1268G>A/p.Ser423Asn** se localizan en el dominio MH2 del gen *SMAD3*, una región muy conservada responsable de la oligomerización de *SMAD2* y *SMAD3* con *SMAD4* activando así la transcripción (Schepers *et al.* 2018). El otro cambio es **c.1081C>T/p.Arg361Cys** que resulta en una sustitución de aminoácidos en la posición 361 del gen *SMAD4* de una arginina a una cisteína. Al igual que en el caso anterior, este cambio se localiza dentro del dominio MH2 por lo que impide la formación del complejo transcripcional (Houlston *et al.* 1998).

Además, hemos identificado 17 mutaciones sin sentido (*nonsense*), que generan un codón de parada prematuro y dan lugar a una proteína truncada.

También se han incluido en este grupo 4 mutaciones localizadas en las regiones de procesamiento del RNA (*splicing*). Estas mutaciones son: **c.236+1G>A** (*SMAD2*), **c.*1A>G** (*SMAD3*), **c.601-1G>A** (*ActR1I*), **c.184+1G>A** (*NCLN*). Al estar localizadas en el primer nucleótido del intrón, afectan al procesamiento del RNA mensajero y, por tanto, dan lugar a una proteína defectuosa.

De los 36 tumores analizados, 17 (47,2%) presentaban mutación patogénica en alguno de los genes implicados en la vía Nodal. Esta cifra puede ser superior si alguna de las variantes de significado incierto se confirma que es patogénica. Es decir, tanto los genes del complejo inhibidor como los genes integrantes de la vía Nodal presentan mutaciones en casi la mitad de

los CCR analizados, lo que confirma la implicación tanto de la vía Nodal como del complejo inhibidor en el desarrollo de cáncer de colon.

El gen en el que identificamos un mayor número de mutaciones patogénicas fue *NOMO* con 9 mutaciones (45%), seguido por *SMAD3*, *SMAD2* y *SMAD4* con 4,3 y 2 mutaciones patogénicas respectivamente. Nos llamó la atención que, de los subgrupos establecidos en este trabajo en función de la edad y del estado de *NOMO* en la qPCR, aquellos en los que se identificó un mayor número de mutaciones en este gen fueron en los tumores que presentaron *NOMO* con pérdida heterocigota independientemente de la edad. Tanto en el subgrupo de pacientes mayores de 70 años como menores de 45 con *NOMO* deleciónado en un alelo, el 50% de los tumores presentaba mutación patogénica en el otro alelo de este gen, lo que sugiere que la pérdida de función del gen es muy frecuente en homocigosis y puede remedar el comportamiento de los genes supresores de tumores.

Hasta la fecha, dentro de la vía de señalización Nodal, solo los genes *SMAD2*, *SMAD3* y *SMAD4* se han asociado al CCR. Los genes *SMAD2* y *SMAD4* están localizados en el cromosoma 18q, una región frecuentemente deleciónada en adenocarcinomas colorrectales (Grady *et al.* 2002). Las mutaciones germinales en *SMAD4* se asocian con el síndrome de poliposis juvenil, y aparece mutado en el 10-15% de los CCR esporádicos. *SMAD2* y *SMAD3* aparecen mutados en menos del 10% de CCR esporádicos (Bellam *et al.* 2010). En ambos genes, las mutaciones más frecuentes se localizan en el dominio MH2 impidiendo la unión con *SMAD4*, por lo que no se forma el complejo transcriptor (Nicholas *et al.* 2012). La secuenciación del exoma ha permitido la identificación de mutaciones en genes menos estudiados en CCR como es el caso del gen *ACTVR1B* (*ALK4*) que aparece mutado en aproximadamente el 5% de adenomas de colon de alto grado (Lee *et al.* 2017). Diferentes estudios también establecen una relación entre mutaciones en el receptor tipo II (*ActR2*) con tumores de colon de mal pronóstico (Zhuo *et al.* 2018). Los resultados obtenidos en nuestro trabajo indican que los genes Smads son, después de *NOMO*, los que presentan mayor frecuencia de mutaciones patogénicas: *SMAD2* aparece mutado en el 5,5% de los tumores, *SMAD3* en el 20% y *SMAD4* en el 5,5%. También se identificó una mutación patogénica en *ALK4* y otra en *ActR2*.

En este trabajo describimos por primera vez mutaciones en los genes *NOMO* y *NCLN*, complejo regulador de la vía nodal y confirmamos la presencia de mutaciones en genes de la vía nodal en CCR, principalmente en pacientes menores de 45 años

Nuestro trabajo muestra por primera vez la implicación del complejo *NOMO/NCNL/TEMEM147* en cáncer colorrectal, aunque no podemos confirmar que sea un

efecto driver. Para confirmarlo, hemos generado un ratón condicionado y tras anular la expresión de NOMO de manera específica en epitelio intestinal de ratón, estamos esperando la evolución de los mismo para confirmar su posible efecto patogénico. Por otra parte, nuestro trabajo indica que, de estar implicado en carcinogénesis colorectal, el complejo NCLN/NOMO/TEMEM147, no lo hace a través de la vía Nodal, por lo que son necesarios estudios buscando el mecanismo patogénico que explique la implicación de NOMO. Finalmente, nuestro trabajo confirma la implicación de genes de la vía Nodal en la carcinogénesis tumoral. Dado que la mayor parte de las alteraciones encontradas en nuestro trabajo aparecen en sujetos con cáncer colorectal de aparición precoz (EOCRC), nuestro trabajo confirma la existencia de alteraciones moleculares específicas de este grupo de pacientes que pueden utilizarse como biomarcadores para el diagnóstico y seguimiento, y como posibles dianas terapéuticas.

CONCLUSIONES



CONCLUSIONES

1.-Nuestro trabajo pone de manifiesto por primera vez que solo existe un gen *NOMO* en el genoma humano.

2.-El gen *NOMO* presenta delección homocigota en más del 30% de los tumores de CCR, asociada principalmente al grupo de EO CRC. La presencia de mutaciones somáticas en los casos con delección heterocigota sugiere que este gen puede estar implicado en la génesis del cáncer colorrectal en pacientes jóvenes, siendo necesarios más estudios para confirmarlo.

3.-La presencia de la delección del gen *NOMO* en la mitad de los pólipos adenomatosos analizados en este trabajo, permite considerar la detección de esta delección como un posible marcador predictivo de EO CRC.

4.-La proteína codificada por el gen *NOMO* interviene en la estabilización del complejo NCLN/*NOMO*/TEMEM147 y la pérdida de este complejo podría ser el mecanismo por el que *NOMO* está implicado en carcinogénesis.

5.-La falta del complejo NCLN/*NOMO*/TEMEM147 no modifica los niveles de las proteínas de la vía Nodal, lo que sugiere que la actividad carcinogénica de *NOMO* debe de ser mediada por una vía accesoria no descrita hasta el momento.

6.-Nuestro trabajo confirma la existencia de alteraciones moleculares específicas no solo en *NOMO* sino en otros genes de la vía Nodal en el grupo de pacientes jóvenes, sugiriendo que estas vías pueden estar implicadas en la carcinogénesis colorrectal en jóvenes.

BIBLIOGRAFÍA



BIBLIOGRAFÍA

1. Aaltonen L, Johns L, Järvinen H, Mecklin JP, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res*. 2007. 13: 356-361.
2. Ahnen DJ, Wade SW, Jones WF, Sifri R, Mendoza Silveiras J, Greenamyer J, Guiffre S, Axilbund J, Spiegel A, You YN. The increasing incidence of young-onset colorectal cancer: a call to action. *Mayo Clin Proc*. 2014. 89: 216-224.
3. Al-Sohaily S, Biankin A, Leong R, Kohonen-Corish M, Warusavitarne J. Molecular pathways in colorectal cancer. *J Gastroenterol Hepatol*. 2012. 27: 1423-31.
4. Amersi F, Agustin M, Clifford Y. Colorectal Cancer: Epidemiology, Risk Factors, and Health Services. *Clin Colon Rectal Surg*. 2005. 18(3): 133–140.
5. Anderson JC. Pathogenesis and management of serrated polyps: current status and future directions. *Gut Liver*. 2014. 8(6):582-9.
6. Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, Cuatrecasas M, Bujanda L, Giraldez MD, Takahashi M, Cabanne A, Barugel ME, Arnold M, Roca EL, Andreu M, Castellvi-Bel S, Llor X, Jover R, Castells A, Boland CR, Goel A. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PLoS One* 2012. 7: e45357.
7. Arriba M, García JL, Rueda D, Pérez J, Brandariz L, Nutu O, Alonso L, Rodríguez Y, Urioste M, González-Sarmiento R, Perea J. Unsupervised Analysis of Array Comparative Genomic Hybridization Data from Early-Onset Colorectal Cancer Reveals Equivalence with Molecular Classification and Phenotypes. *Neoplasia*. 2016. Volume 19 Number xx. pp. 28–34 28.
8. Arriba M, García JL, Inglada-Pérez L, Rueda D, Osorio I, Rodríguez Y, Álvaro E, Sánchez R, Fernández T, and Pérez J, et al. DNA copy number profiling reveals different patterns of chromosomal instability within colorectal cancer according to the age of onset. *Mol Carcinog*. 2015. 55(5):705-16. doi: 10.1002/mc.22315.
9. Arvelo F, Sojo F, Cotte C. Biology of colorectal cancer. *Ecancermedicalscience*. 2015. 9;9:520.
10. Arnold M, Pandeya N, Byrnes G, Renehan AG, Stevens GA, Ezzati M. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol*. 2015. 16(1):36–46.
11. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg* 1954. 139:846-52.
12. Bae SI, Kim YS. Colon cancer screening and surveillance in inflammatory bowel disease. *Clin Endosc*. 2014. 47(6):509-15.

13. Bakircioglu M, Carvalho OP, Khurshid M et al: The essential role of centrosomal NDE1 in human cerebral cortex neurogenesis. *Am J Hum Genet.* 2011. 88: 523–535.
14. Balemans W, Van Hul W. Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. 2002.*Dev Biol* 250: 231–250.
15. Ballester V, Rashtak S, Boardman L. Clinical and molecular features of young-onset colorectal cancer. *World J Gastroenterol.* 2016. 7; 22(5): 1736-1744.
16. Balmaña J, Balaguer F, Cervantes A, Arnold D; ESMO Guidelines Working Group. Familial risk-colorectal cancer: ESMO Clinical Practice Guidelines. *Ann Oncol.* 2013 Oct;24 Suppl 6:vi73-80.
17. Banerjee A, Hands RE, Powar MP, Bustin SA, Dorudi S. Microsatellite and chromosomal stable colorectal cancers demonstrate poor immunogenicity and early disease recurrence. *Colorectal Dis* 2009. 11: 601-608.
18. Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer.2013. *Gut.* 62(6):933–47.
19. Bas B, Dinc B, Oymaci E, Mayir B, Gunduz UR. What are the Endoscopic and Pathological Characteristics of Colorectal Polyps? *Asian Pac J Cancer Prev.* 2015.16(13):5163-7.
20. Baltimore David, Berg Paul, Botchan Michael, Carroll Dana, R. Alta Charo, George Church, Jacob E. Corn, George Q. Daley, Jennifer A. Doudna, Marsha Fenner, Henry T. Greely, Martin Jinek, G. Steven Martin, Edward Penhoet, Jennifer Puck, Samuel H. Sternberg, Jonathan S. Weissman, Keith R. Yamamoto., “A prudent path forward for genomic engineering and germline gene modification”, *Science.* 2015. vol. 348, n. 6230, p. 38.
21. Bellam & Pasche. TGF- β Signaling Alterations and Colon Cancer. *Cancer Genetics.* 2010. 85–103.
22. Berg M, Agesen TH, Thiis-Evensen E; INFAC-study group, Merok MA, Teixeira MR, Vatn MH. Distinct high resolution genome profiles of early onset and late onset colorectal cancer integrated with gene expression data identify candidate susceptibility loci. *Mol Cancer.* 2010. 6;9:100.
23. Berger NA, Savvides P, Koroukian SM, Kahana EF, Deimling GT, Rose JH, Bowman KF, Miller RH. Cancer in the elderly. *Trans Am Clin Climatol Assoc.* 2006. 117:147–55.
24. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. *Obes Rev.*2002. 3(1):5–8.
25. Bianco C, Strizzi , Rehman A. A Nodal- and ALK4-independent signaling pathway activated by Cripto-1 through Glypican-1 and c-Src. *Cancer Res.*2003. 63, 1192-1197.
26. Biesecker LG, Sapp JC. Proteus Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. *GeneReviews*® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2019.2012 [updated 2019 Jan 10].
27. Binefa G, Rodríguez-Moranta F, Teule A, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. *World J Gastroenterol.* 2014. 14;20(22):6786-808.

28. Blackburn, E. H. Structure and function of telomeres. *Nature*. 1991. 350, 569-573.
29. Blanco I, Teulé A, González S: Aspectos clínicos de los síndromes polipósicos. En: Alonso MA, Balmaña J, Benavides M, et al, Eds. *Cáncer Hereditario II Ed.* SEOM 2010. pp. 559-90.
30. Bockmühl U, Wolf G, Schmidt S, Schwendel A, Jahnke V, Dietel M, Petersen I. Genomic alterations associated with malignancy in head and neck cancer. *Head Neck*. 1998. 20(2):145-51.
31. Bodenstine TM, Chandler GS, Seftor RE, Seftor EA, Hendrix MJ. Plasticity underlies tumor progression: role of Nodal signaling. *Cancer Metastasis Rev*. 2016. 35(1):21–39.
32. Botteri E, Iodice S, Bagnardi V. Smoking and colorectal cancer: a metaanalysis. *JAMA* 2008. 300:2765-78.
33. Boyle P y Langman LS. ABC of colorectal cancer. *Epidemiology. Br Med J* 2000. 321:805-808.
34. Calabria I, Pedrola L, Berlanga P, Aparisi MJ, Sánchez D, Cañete A, Cervera J, Millán JM, Castel V. El nuevo reto en oncología: la secuenciación NGS y su aplicación a la medicina de precisión. *AnPediatr(Barc)*.2016. 85(5):273.e1-273.e7.
35. Cao L, Li L, Li Y, Zhuang J, Chen Y, Luo S, Wang Y, Shi Y, Chen J, Gao J, Zhu X, Wan Y, Li F, Yuan W, Mo X, Ye X, Wang X, Zhou Z, Dai G, Jiang Z, Zhu P, Wu X, Fan X. Loss of the Nodal modulator Nomo results in chondrodysplasia in zebrafish. *Curr Mol Med*. 2018.
36. Carro A, Rico D, Rueda OM, Diaz-Uriarte R, Pisano DG. WaviCGH: a web application for the analysis and visualization of genomic copy number alterations. *Nucleic Acids Res*. 2010. 38.
37. Cartegni L., Wang J., Zhu Z., Zhang M. Q., Krainer A.R. ESE finder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003. 31:3568-3571.
38. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012. 18;487(7407):330-7.
39. Capps B, Chadwick R, Joly Y, Mulvihill J.J, Lysaght T, Zwart H. Falling giants and the rise of gene editing: ethics, private interests and the public good. *Human Genomics*. 2017. 11 (1): 20.
40. Chen, J. M., Férec, C., & Cooper. D. N. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: The importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. *Human Genetics*. 2006. 120(3), 301–333.
41. Cho, S.W. Analysis of off-target effects of CRISPR Cas-derived RNA-guided endonucleases and nickases. *Genome Research*.2014. 24, pp.132–141.

42. Choucair KA, Guérard KP, Ejdelman J, Chevalier S, Yoshimoto M, Scarlata E, Fazli L, Sircar K, Squire JA, Brimo F, Cunha IW, Aprikian A, Gleave M. The 16p13.3 (PDPK1) Genomic Gain in Prostate Cancer: A Potential Role in Disease Progression. *Transl Oncol.* 2012. 5:453–60.
43. Colorectal Cancer: Overview [Internet]. *Cancer.net*; 2015. Available from: <https://www.cancer.net/cancer-types/colorectal-cancer/overview>.
44. Cong, L. & Zhang, F. Genome engineering using CRISPR-Cas9 system. *Methods Mol. Biol.* 2015. 1239, 197–217.
45. Corral J, Borràs JM, Chiarello P, García-Alzoriz E, Macià F, Reig A, et al. Estimation of hospital costs of colorectal cancer in Catalonia (Spain). *Gac Sanit.* 2015 Aug 26.
46. Cunningham D, Atkin W, Lenz HJ, Lynch Ht, Minsky B, Nordlinger B, et al. Colorectal cancer. *Lancet.* 2010. 20;375.
47. Database of Single Nucleotide Polymorphisms (dbSNP). National Center for Biotechnology Information, U.S. National Library of Medicine (Bethesda, MD, US). <http://www.ncbi.nlm.nih.gov/SNP/>.
48. D'Ottavio G, Wulfson A, Rojman J. CARCINOGENÉISIS COLORRECTAL Nuevas perspectivas e implicancias clínicas para su detección. *Rev. Méd. Rosario.* 2014. 80: 63-74.
49. Duan W, Li R, Ma J, Lei J, Xu Q, Jiang Z, Nan L, Li X, Wang Z, Huo X, Han L, Wu Z, Wu E and Ma Q. Overexpression of Nodal induces a metastatic phenotype in pancreatic cancer cells via the Smad2/3 pathway. *Oncotarget* 2015; 6: 1490-1506.
50. Dettmer U, Kuhn P, Haffner C. Transmembrane Protein 147 (TMEM147) is a novel component of the Nicalin-NOMO Protein Complex. *J Biol Chem.* 2010. 285(34).
51. Driss Ait Ouakrim, Cécile Pizot, Magali Boniol, Matteo Malvezzi, Mathieu Boniol, Eva Negri, Maria Bota, Mark A Jenkins, Harry Bleiberg, Philippe Autier. Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. *BMJ.* 2015. 351:h4970.
52. Edge SB, Byrd DR, Compton CC, et al. editors. *AJCC Cancer Staging Manual.* 7th ed. New York, NY: SpringerVerlag, 2009.
53. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, editores. *AJCC Cancer Staging Manual.* Octava edición. EE.UU.: Springer. 2010.
54. Eissing M, Ripken L, Schreibelt G, Westdorp H, Ligtenberg M, Netea-Maier R, Netea MG, de Vries IJM, Hoogerbrugge N. PTEN Hamartoma Tumor Syndrome and Immune Dysregulation. *Transl Oncol.* 2019. 12(2):361-367.
55. Engelhardt M, Drullinsky P, Guillem J, Moore MA. Telomerase and telomere length in the development and progression of premalignant lesions to colorectal cancer. *Clin Cancer Res.* 1997.1931-41.

56. European Colorectal Cancer Screening Guidelines Working Group, von Karsa L, Patnick J, Segnan N, Atkin W, Halloran S, Lansdorp-Vogelaar I, Malila N, Minozzi S, Moss S, Quirke P, Steele RJ, Vieth M, Aabakken L, Altenhofen L, Ancelle-Park R, Antoljak N, Anttila A, Armaroli P, Arrossi S, Austoker J, Banzi R, Bellisario C, Blom J, Brenner H, Bretthauer M, Camargo Cancela M, Costamagna G, Cuzick J, Dai M, Daniel J, Dekker E, Delicata N, Ducarroz S, Erfkamp H, Espinàs JA, Faivre J, Faulds Wood L, Flugelman A, Frkovic-Grazio S, Geller B, Giordano L, Grazzini G, Green J, Hamashima C, Herrmann C, Hewitson P, Hoff G, Holten I, Jover R, Kaminski MF, Kuipers EJ, Kurtinaitis J, Lambert R, Launoy G, Lee W, Leicester R, Leja M, Lieberman D, Lignini T, Lucas E, Lynge E, Mádai S, Marinho J, Maučec Zakotnik J, Minoli G, Monk C, Morais A, Muwonge R, Nadel M, Neamtiu L, Peris Tuser M, Pignone M, Pox C, Primic-Zakelj M, Psaila J, Rabeneck L, Ransohoff D, Rasmussen M, Regula J, Ren J, Rennert G, Rey J, Riddell RH, Risio M, Rodrigues V, Saito H, Sauvaget C, Scharpantgen A, Schmiegel W, Senore C, Siddiqi M, Sighoko D, Smith R, Smith S, Suchanek S, Suonio E, Tong W, Törnberg S, Van Cutsem E, Vignatelli L, Villain P, Voti L, Watanabe H, Watson J, Winawer S, Young G, Zaksas V, Zappa M, Valori R. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. *Endoscopy* 2013; 45: 51-59.
57. Fang X, Hu T, Yin H, Yang J, Tang W, Hu S, Xu X. Differences in telomerase activity and the effects of AZT in aneuploid and euploid cells in colon cancer. *Int J Oncol.* 2017.525-532.
58. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell.* 1990 1;61(5):759-67.
59. Feldcamp LA, Boutros PC, Raymond R, Fletcher PJ, Nobrega JN, Wong AHC. Pdx1c1 modulates prepulse inhibition of acoustic startle in the mouse. *Translational Psychiatry.* 2017. 7.
60. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. Lyon, France: International Agency for Research on Cancer, 2013.
61. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer.* 2013. 49:1374-403.
62. Fernández-Marcelo T, Sánchez-Pernaute A, Pascua1 I, De Juan C, Head J, Torres-García A, Iñiesta P. Clinical Relevance of Telomere Status and Telomerase Activity in Colorectal Cancer. *Plos One.* 2016.
63. Flicker K., Smolle E., Haybaeck J. Moinfar F. Genomic characterization of endometrial stromal sarcomas with array comparative genomic hybridization. *Experimental and Molecular Pathology.* 2015. 98(3), 367–374.
64. Francescangeli F, Contavalli P, De Angelis ML, Baiocchi M, Gambarà G, Pagliuca A, Fiorenzano A, Prezioso C, Boe A, Todaro M, Stassi G, Castro NP, Watanabe K, Salomon DS, De Maria R, Minchiotti G, Zeuner A. Dynamic regulation of the cancer stem cell compartment by Cripto-1 in colorectal cancer. *Cell Death Differ.* 2015. 1700-13. doi: 10.1038/cdd.2015.19.
65. Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*, 32(3), pp.279–84.
66. Galceran J, Ameijide A, Carulla M, Mateos A, Quirós J. A, Rojas D, Alemán A, Torrella A, Chico M, Vicente M, Díaz J.M, Larrañaga N, Marcos-Gragera R, Sánchez M.J, Perucha J, Franch P, Navarro C,

- Ardanaz E, Bigorra J, Rodrigo P, Peris Bonet R. REDECAN Working Group. Cancer incidence in Spain. *Clin Transl Oncol*. 2015.
67. Gallegos-Villanueva MJ, Chimenos-Küstner E, López-López J, Roselló-Llabrés X. Cancerización de campo: revisión del concepto. *Av Odontostomatol*. 2007. Vol.23 no.1.
68. Giang H, Nguyen VT, Nguyen SD, Nguyen HP, Vo BT, Nguyen TM, Nguyen NH, Truong KD, Do TT, Phan MD, Nguyen HN. Detection of a heterozygous germline APC mutation in a three-generation family with familial adenomatous polyposis using targeted massive parallel sequencing in Vietnam. *BMC Med Genet*. 2018. 19(1):188.
69. Giorgio E, Liguoro A, D'Orsi L et al. Cripto haploinsufficiency affects in vivo colon tumor development. *Int J Oncol*. 2014; 45: 31-40.
70. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. ARCI: OMS; [último acceso el 13/01/18]. Accesible en: <http://globocan.iarc.fr/Default.aspx>
71. Gong, Y., Guo, Y., Hai, Y., Yang, H., Liu, Y., Yang, S., ... He, Z. (2014). Nodal Promotes the Self-Renewal of Human Colon Cancer Stem Cells via an Autocrine Manner through Smad2/3 Signaling Pathway. *BioMed Research International*. 2014. 1–11.
72. Gong W, Sun B, Zhao X, et al. Nodal signaling promotes vasculogenic mimicry formation in breast cancer via the Smad2/3 pathway. *Oncotarget*. 2016. 7(43):70152–70167.
73. Gonzalez MB, Hernandez JM, Garcia JL, Lumbreras E, Castellanos M, Hernández JM, Fernández-Calvo J, Gutiérrez NC, San Miguel JF. The value of fluorescence in situ hybridization for the detection of 11q in multiple myeloma. *Haematologica*. 2004. 89:1213–1218.
74. Grady WM, Markowitz SD. Genetic and epigenetic alterations in colon cancer. *Annu.Rev Genom Hum Genet*. 2002. 3:101–128
75. Greaney ML, Puleo E, Sprunck-Harrild K, Syngal S, Gonzalez EG, Emmons KM. Changes in colorectal cancer screening intention among people aged 18–49 in the United States. *BMC Public Health*. 2014. 14:901.
76. Gryfe R. Inherited Colorectal Cancer Syndromes. *Clinics in colon and rectal cancer surgery*. 2009. 22:4.
77. Guilinger, J.P, Thompson, D.B. & Liu, D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology*. 2014. 32(6), pp.577–82.
78. Guillermo E. D'Ottavio, Adolfo M. Wulfson, José A. Rojman. CARCINOGENESIS COLORRECTAL Nuevas perspectivas e implicancias clínicas para su detección. *Rev. Méd. Rosario*. 2014. 80: 63-74.
79. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015. 21:1350-6.

80. Haffner C, Dettmer U, Weiler T, Haass C. The Nicastrin-like protein Nicalin regulates assembly and stability of the Nicalin-nodal modulator (NOMO) membrane protein complex. *The Journal of Biological Chemistry*. 2007. Vol. 282, NO. 14, pp. 10632–10638.
81. Haffner C, Frauli M, Topp S, Irmeler M, Hofmann K, Regula JT, Bally-Cuif L, Haass C. Nicalin and its binding partner Nomo are novel Nodal signaling antagonists. *EMBO J*. 2004. 4;23(15):3041-50.
82. Haq AI, Schneeweiss J, Kalsi V, Arya M. The Dukes staging system: a corner stone in the clinical management of colorectal cancer. *Lancet Oncol*. 2009;10(11):1128.
83. Harms PW, Chang C. Tomoregulin-1 (TMEFF1) inhibits nodal signaling through direct binding to the nodal coreceptor Cripto. *Genes*. 2003. Dev 17: 2624–2629
84. Harriss DJ, Atkinson G, Batterham A, George K, Cable NT, Reilly T. Colorectal Cancer, Lifestyle, Exercise and Research Group. Lifestyle factors and colorectal cancer risk (2): a systematic review and meta-analysis of associations with leisure-time physical activity. *Colorectal Dis*. 2009. 11(7):689–701.
85. Hartz A, He T, Ross JJ. Risk factors for colon cancer in 150.912 postmenopausal women. *Cancer Causes Control*. 2012. 23:1599-1605.
86. Hashimoto H, Rebagliati M, Ahmad N, Muraoka O, Kurokawa T, Hibi M, Suzuki T The Cerberus/Dan-family protein Charon is a negative regulator of Nodal signaling during left–right patterning in zebrafish. *Development*. 2004. 131: 1741–1753.
87. Hause R, Pritchard C, Shendure J, Salipante S. Classification and characterization of microsatellite instability across 18 cancer types. *Nature Medicine*. 2016.
88. Hav M, Eav S, Ky V, Cuvelier C, In S, Kong R, Kheang Y, et al. Colorectal cancer in young Cambodians. *Asian Pac J Cancer Prev*. 2011.12(4):1001-5.
89. Hawtorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. *Genomics*.2014. 103, 211–221.
90. Heald B, Rigelsky C, Moran R, LaGuardia L, O'Malley M, Burke CA, Zahka K. Prevalence of thoracic aortopathy in patients with juvenile Polyposis Syndrome-Hereditary Hemorrhagic Telangiectasia due to SMAD4. *Am J Med Genet A*. 2015. 167A(8):1758-62.
91. Helbig I, Mefford, H.C, Sharp, A.J, Guipponi, M, Fichera, M, Franke, A, Muhle, H, de Kovel, C, Baker, C, von Spiczak, S. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet*. 2009. 41, 160–162.
92. Hennenon P, Frank M, Litvinova E, Miranda S, Messas E, Darnige L. [Gardner-Diamond syndrome in a young man: A case report and literature review]. *Rev Med Interne*. 2017. 38(9):623-627.
93. Henson DE, Hueman MT, Chen D, Patel JA, Wang H, Schwartz AM. The anatomy of the TNM for colon cancer. *J Gastrointest Oncol* 2017. 8(1):12-19.

94. Houlston, R., Bevan, S., Williams, A., Young, J., Dunlop, M., Rozen, P., Tomlinson, I. Mutations in DPC4 (SMAD4) cause juvenile polyposis syndrome, but only account for a minority of cases. *Human Molecular Genetics*. 1998. 7(12), 1907–1912.
95. Huhn, S, da Silva Filho, M. I., Sanmuganatham, T., Pichulik, T., Catalano, C., Pardini, B. Försti, A. Coding variants in NOD-like receptors: An association study on risk and survival of colorectal cancer. *PLOS ONE*. 2018. 13/6.
96. Hsu, P.D. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*. 2013. 31(9), pp.827–32.
97. Hsiao YH, Wei CH, Chang SW, Chang L, Fu YW, Lee HC, Liu HL, Yeung CY. Juvenile polyposis syndrome: An unusual case report of anemia and gastrointestinal bleeding in Young infant. *Medicine (Baltimore)*. 2016. 95(37).
98. IARC Handbooks of Cancer Prevention, Volume 6. Weight control and physical activity. 2002. Lyon, France: IARC Press.
99. Ingason A, Rujescu D, Cichon S et al: Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry*. 2011. 16: 17–25.
100. Iratni R, Yan YT, Chen C, Ding J, Zhang Y, Price SM, Reinberg D, Shen MM. Inhibition of excess nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. *Science*. 2002. 298: 1996–1999
101. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Hisopathology*. 2007. 50:113–30.
102. Jass JR, Whitehall VLJ, Young J, Leggett BA. Emerging concepts in colorectal neoplasia. *Gastroenterology*. 2003.123: 862-76.
103. Johnson DS, Li YM, Pettersson M, St George-Hyslop PH. Structural and Chemical Biology of Presenilin Complexes. *Cold Spring Harb Perspect Med*. 2017. 1;7(12).
104. Ju, J., Kim D.H., Bi, L., Meng, Q., Bai, X., Li, Z., Li, X., Marma, M.S., Shi, S., Wu, J., Edwards, J.R., Romu, A., Turro, N.J. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. U.S.A.* 2006. 103: 19635-19640.
105. Kamat S, Yeola S, Zhang W, Bianchi L, Driscoll M. NRA-2, a nicalin homolog, regulates neuronal death by controlling surface localization of toxic *Caenorhabditis elegans* DEG/ENaC channels. *J Biol Chem*. 2014. 25;289(17).
106. Kanwal Tariq, Kulsoom Ghias. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol Med* 2016.

107. Karki, R., Pandya, D., Elston, R. C., & Ferlini, C. Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC Medical Genomics*. 2015. 8(1), 1–7.
108. Kentaro Inamura. Colorectal Cancers: An Update on Their Molecular Pathology. *Cancers*. 2018. 10, 2.
109. Kettunen J¹, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikäinen LP, Kangas AJ, Soininen P, Würtz P, Silander K, Dick DM, Rose RJ, Savolainen MJ, Viikari J, Kähönen M, Lehtimäki T, Pietiläinen KH, Inouye M, McCarthy MI, Jula A, Eriksson J, Raitakari OT, Salomaa V, Kaprio J, Järvelin MR, Peltonen L, Perola M, Freimer NB, Ala-Korpela M, Palotie A, Ripatti S. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet*. 2012. 29;44(3):269-76.
110. Kim SH, Park KH, Shin SJ, Lee KY, Kim TI, Kim NK, Rha SY, Ahn JB. CpG Island Methylator Phenotype and Methylation of Wnt Pathway Genes Together Predict Survival in Patients with Colorectal Cancer. *Yonsei Med J*. 2018. 59(5):588-594.
111. Kirzin S, Marisa L, Guimbaud R, De Reynies A, Legrain M, Laurent-Puig P, et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. *PLoS One*. 2014. 9(8):e103159.
112. Koh KJ, Park HN, Kim KA. Gardner syndrome associated with multiple osteomas, intestinal polyposis, and epidermoid cysts. *Imaging Sci Dent*. 2016. 46(4):267-272.
113. Koh KS, Telisinghe PU, Bickle I, Abdullah MS, Chong CF, Chong VH. Characteristics of young colorectal cancer in brunei darussalam: an epidemiologic study of 29 years (1986-2014). *Asian Pac J Cancer Prev*. 2015;16(8):3279-83.
114. Kyo, S., Kanaya, T., Takakura, M., Tanaka, M. & Inoue, M. Human telomerase reverse transcriptase as a critical determinant of telomerase activity in normal and malignant endometrial tissues. *International journal of cancer. Journal international du cancer*. 1999. 80, 6063.
115. Labianca R, Beretta GD, Kildani B, Milesi L, Merlin F, Mosconi S. Colon Cancer. *Crit Rev Oncol Hematol*. 2010. 74 (2): 106-133.
116. Labots M, Buffart TE, Haan JC, van Grieken NC, Tijssen M, van de Velde CJ, et al. High-level copy number gains of established and potential drug target genes in gastric cancer as a lead for treatment development and selection. *Cell Oncol (Dordr)*. 2014. 37(1):41-52.
117. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970. 227, 680–685.
118. Laish I, Mizrahi J, Naftali T, Konikoff FM. Diabetes Mellitus and Age are Risk Factors of Interval Colon Cancer: A Case-Control Study. *Dig Dis*. 2019. 1-6.
119. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr*. 2007. 86(3):556–65.

120. Laurent S, Franchimont D, Coppens JP, Leunen K, Macken L, Peeters M, et al. Familial adenomatous polyposis: clinical presentation, detection and surveillance. *Acta Gastroenterol Belg.* 2011 Sep;74(3):415-20.
121. Lawrence MG, Margaryan NV, Loessner D, Collins A, Kerr KM, Turner M, Seftor EA, Stephens CR, Lai J and Postovit LM. Reactivation of embryonic nodal signaling is associated with tumor progression and promotes the growth of prostate cancer cells. *Prostate* 2011; 71: 1198-209.
122. Lee KL, Lim SK, Orlov YL, Yitl Y, Yang H, Ang LT, Poellinger L, Lim B. Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLOS GENET.* 2011. 7:e1002130.
123. Lee SH, Jung SH, Kim TM, Rhee JK, Park HC, Kim MS, Kim SS, An CH, Lee SH, Chung YJ. Whole-exome sequencing identified mutational profiles of high-grade colon adenomas. *Oncotarget.* 2017. 8(4):6579-6588.
124. Le Gallo M, Lozy F, Bell DW. Next-Generation Sequencing. *Adv Exp Med Biol.* 2017;943:119-148.
125. Leonardo Saúl Lino-Silva, Alberto León-Takahashi, Horacio López-Basave, Alejandro Padilla-Rosciano, Gerardo Miranda-Dévora, Martín Granados-García, Rosa Salcedo-Hernández. Colorectal carcinoma molecular classification. A short review. *Gac Med Mex.* 2018; 154:507-512.
126. Li H, Shi H, Huo K. p75NTR signal transduction suppressed by BFAR and p75NTR interactions. *Sci China Life Sci.* 2012. 55(4):367-74.
127. Li J, Guo B-C, Sun L-R, Wang J-W, Fu X-H, Zhang S-Z, et al. TNM staging of colorectal cancer should be reconsidered by T stage weighting. *World J Gastroenterol.* 2014. 7;20(17):5104–12.
128. Li Y., Zhong, W., Zhu, M., Hu, S., & Su, X. Nodal regulates bladder cancer cell migration and invasion via the ALK/Smad signaling pathway. *OncoTargets and Therapy.* 2018. Volume 11, 6589–6597.
129. Liang JT, Huang KC, Cheng AL, Jeng YM, Wu MS, Wang SM. Clinicopathological and molecular biological features of colorectal cancer in patients less than 40 years of age. *Br J Surg* 2003; 90: 205-214.
130. Libutti SK, Saltz LB and Willet CG and Levine RA: Cancer of the Colon. En: De Vita VT, Hellman S and Rosenberg SA: *Cancer: Principles & Practice of Oncology.* 10th Ed. Wolters Kluwer Health. Philadelphia. 2015; pp 768-812.
131. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer- -analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000. 343: 78-85.
132. Lin, Y. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research.* 2014. 42(11), pp.7473–7485.

133. Lonardo E, Hermann P, Mueller MT, Huber S, Balic A, Miranda-Lorenzo I, Zagorac S, Alcalá S, Rodríguez-Arabaolaza I and Ramirez JC. Nodal/activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell Stem Cell* 2011; 9: 433-446.
134. Long, C. Prevention of muscular dystrophy in mice by CRISPR/Cas9– mediated editing of germline DNA. *Science*. 2014. 345(6201), pp.1184–1188.
135. Losi L, Di Gregorio C, Pedroni M, Ponti G, Roncucci L, Scarselli A, Genuardi M, Baglioni S, Marino M, Rossi G, Benatti P, Maffei S, Menigatti M, Roncari B, Ponz de Leon M: Molecular genetic alterations and clinical features in early-onset colorectal carcinomas and their role for the recognition of hereditary cancer syndromes. *Am J Gastroenterol* 2005. 100:2280e2287.
136. Luy G, Maldonado G, Chinga E, Luy S, Peinado J. Características clínicas del cáncer colorectal en el Hospital E. Rebagliati Martins 1995-1999; *Rev Gastroenterol Perú*. 2000. 20: 406 – 13.
137. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, and Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.* 2009. 76(1): 1–18.
138. Ma, S. CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of *BmKu70* in *Bombyx mori*. *Scientific reports*. 2014. 4, p.4489.
139. Mali, P., Yang, L. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*. 2013. 339(6121), pp.823–6.
140. Mancino M, Strizz L, Wechselberge C, et al. Regulation of human cripto-1 gene expression by TGF- β 1 and BMP-4 in embryonal and colon cancer cells. *J Cell Physiol*. 2008. 215: 192–203
141. Marley AR, Nan H. Epidemiology of colorectal cancer. *Int J Mol Epidemiol Genet*. 2016. 30;7(3):105–114.
142. Martin J, Han C, Gordon LA, Terry A, Prabhakar S, She X, Xie G, Hellsten U, Chan YM, Altherr M, Couronne O, Aerts A. The sequence and analysis of duplication-rich human chromosome 16. *Nature*. 2004. Dec 23;432(7020):988-94.
143. Martin SA, Hewish M, Lord CJ, Ashworth A. Genomic instability and the selection of treatments for cancer. *J Pathol*. 2010. 220(2):281-9.
144. Martínez De LaPiscina Martín, Pérez de Nanclares Leal. Nuevas tecnologías aplicadas en la detección de alteraciones genéticas. *Rev Esp Endocrinol Pediatr*. 2018. 9 (Suppl 1). 10.3266.
145. Martinez-Delgado, B. et al. Shorter telomere length is associated with increased ovarian cancer risk in both familial and sporadic cases. *Journal of medical genetic*. 2012. s 49, 341-344.
146. Martínez-Urueña N, Macias L, Perez-Carbonero L, et al. Incidence of -93 MLH1 promoter polymorphism in familial and sporadic colorectal cancer. *Colorectal Dis*. 2013. 15:e118-23.

147. Massague, J., Seoane, J. and Wotton, D. Smad transcription factors. *Genes*. 2005. *Dev*. 19, 2783-2810.
148. Matsubara N. Epigenetic regulation and colorectal cancer. *Dis Colon Rectum*. 2012. 55: 96-104.
149. Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet*. 2016. 24:2-5.
150. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011. 12, R41.
151. Moghaddam AA, Woodward M, Huxley R. Obesity and risk of colorectal cancer: a metaanalysis of 31 studies with 70,000 events. *Cancer Epidemiol Biomarkers*. 2007 *Prev*. 16(12):2533-47.
152. Mojica, F.J.M. y Garret, R.A. Discovery and Seminal Developments in the CRISPR Field. En *CRISPR-Cas Systems RNA-Mediated Adaptive Immunity in Bacteria and Archaea*. Barrangou, R. y van der Oost, J., 2013. eds. Springer, p. 299.
153. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* 65, 55-63 (1983).
154. Mulvihill J.J., Capps B., Joly Y., Lysaght T., Zwart H.A., Chadwick R., International Human Genome Organization (HUGO), Committee of Ethics, Law, and Society (CELS). Ethical issues of CRISPR technology and gene editing through the lens of solidarity. *British Medical Bulletin*. 2017. 122 (1): 17-29.
155. Nagamani SC, Erez A, Bader P, Lalani SR, Scott DA, Scaglia F, Plon SE, Tsai CH, ReMISchisel T, Roeder E, Malphrus AD, Eng PA, Hixson PM, et al. Phenotypic manifestations of copy number variation in chromosome 16p13.11. *Eur J Hum Genet*. 2011. 19:280-6.
156. Nardon E, Glavac D, Benhattar J, y col. A multicenter study to validate the reproducibility of MIS testing with a panel of 5 quasimonomorphic mononucleotide repeats. *Diagn Mol Pathol*. 2010. 19: 236-42.
157. Nayak SS, Roy P, Arora N, Arun I, Roy MK, Banerjee S, Mallick I, Mallath MK. Prevalence estimation of microsatellite instability in colorectal cancers using tissue microarray based methods - A tertiary care center experience. *Indian J Pathol Microbiol*. 2018.
158. Nishimasu, Hl. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. 2014. 156, 935-49.
159. Norat T, Scoccianti C, Boutron-Ruault MC, Anderson A, Berrino F, Cecchini M, et al. European Code against Cancer 4th edition: Diet and cancer. *Cancer Epidemiol*. 2015. pii: S1877-7821(15)00070-3.
160. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular Binary Segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004. 5(4):557-72.

161. Ogino S, Goel A. Molecular classification and correlatos in colorectal cancer. *J Mol Diagn* 2008. 10:13-27.
162. Orive M, Aguirre U, Gonzalez N, Lázaro S, Redondo M, Bare M, Anula R, Briones E, Escobar A, Sarasqueta C, García-Gutierrez S, Quintana JM. Risk factors affecting hospital stay among patients undergoing colon cancer surgery: a prospective cohort study. *REDISSEC-CARESS/CCR group. Support Care Cancer*. 2019 . doi: 10.1007/s00520-019-04683-7.
163. Osorio I, Rueda D, Rodríguez Y, García JL, de la Cruz Vigo F, et al. Early-onset colorectal cancer: A separate subset of colorectal cancer. *World J Gastroenterol*. 2014. 14; 20(46): 17288–17296.
164. Palacio Rúa, Muñetón Peña. Bases moleculares del cáncer colorectal. *Iatreia*. 2012. Vol. 25 (2): 137 – 148.
165. Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu Rev Genet*. 2008.42, 301-334.
166. Park Y, Hunter DJ, Spiegelman D, Bergkvist L, Berrino F, Van den Brandt PA, et al. Dietary fiber intake and risk of colorectal cancer: A pooled analysis of prospective cohort studies. *JAMA*. 2005. 294: 2849-57.
167. Parajuli R, Bjerkaas E, Tverdal A, Le Marchand L, Weiderpass E, Gram IT. Smoking increases rectal cancer risk to the same extent in women as in men: results from a Norwegian cohort study. *BMC Cancer*. 2014. 6;14:321.
168. Pascual A. (2017). *Nuevas aportaciones a la caracterización del cancer de mama en mujeres jóvenes*. (Tesis doctoral). Universidad de Salamanca, Salamanca, España.
169. Perea J, Cano JM, Rueda D, et al. Classifying early-onset colorectal cancer according to tumor location: new potencial subcategories to explore. *Am J Cancer Res* 2015; 5:2308-13.
170. Perea J, García JL, Pérez J, Rueda D, Arriba M, Rodríguez Y, Urioste M, González-Sarmiento R. *NOMO-1* gene is deleted in early-onset colorectal cancer. *Oncotarget*. 2017. 8(15):24429-24436.
171. Perea J, Lomas M, Hidalgo M. Molecular basis of colorectal cancer: Towards an individualized management? *Rev Esp Enferm Dig*. 2011; 103:29-35.
172. Perea J, Rueda D, Canal A, Rodríguez Y, Alvaro E, Osorio I, et al. Age at onset should be a major criterion for subclassification of colorectal cancer. *J Mol Diagn*. 2014. 16(1):116-26.
173. Pérez Segura P: Alteraciones moleculares en CCHNP. En: Alonso MA, Balmaña J, Benavides M, et al, Eds. *Cáncer Hereditario II Ed. SEOM*. 2010; pp. 479-497.

174. Pillozzi E, Maresca C, Duranti E, Giustiniani MC, Catalanotto C, Lucarelli M, Cogoni C, Ferri M, Ruco L, Zardo G. Left-sided early-onset vs late-onset colorectal carcinoma: histologic, clinical, and molecular differences. *Am J Clin Pathol*. 2015; 143:374–84.
175. Pinheiro M, Pinto C, Peixoto A, Veiga I, Lopes P, Henrique R, et al. Target gene mutational pattern in Lynch syndrome colorectal carcinomas according to tumour location and germline mutation. *Br J Cancer*. 2015. 11;113(4):686-92.
176. Poulogiannis G, Frayling IM, Arends MJ. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology*. 2010. 56(2):167-79.
177. Qaiyoume A, Khursheed A. Colorectal cancer in younger population: Our experience. *J Pak Med Assoc*. 2013. 63: 1275-7.
178. Quail D, Siegers GM, Jewer M, Postovit L_M, Nodal signalling in embryogenesis and tumourigenesis. *Int J Biochem Cell Biol*. 2013. 45: 885-898.
179. Quail DF, Zhang G, Walsh LA, et al. Embryonic morphogen nodal promotes breast cancer growth and progression. *PLoS One*. 2012; 7:e48237.
180. Quirke P, Risio M, Lambert R, von Karsa L, Vieth M. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Quality assurance in pathology in colorectal cancer screening and diagnosis. *Endoscopy*. 2012. 44
181. Ramalingam A, Zhou X, Fiedler S, Brawner S, Joyce J, Liu H, Yu S, 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders, *Journal of Human Genetics*. 2011. 56, 541–544
182. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc*. 8. 2013. 2281–2308.
183. Ran, F.A. Double nicking by RNA- guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013.154(6).
184. REDECAN. Red Española de Registros de Cáncer. [último acceso el 13/01/18]. Accesible en: <http://redecn.org/es/index.cfm>
185. Reissmann E., Jornvall H. et al. The orphan receptor ALK7 and the Activin receptor ALK4 mediate signaling by Nodal proteins during vertebrate development. *Genes Dev*. 2001. 15: 2010-2022.
186. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*.2008. 371(9612):569–78.
187. Rex DK, Ahnen DJ, Baron JA, Batts KP, Burke CA, Burt RW, et al. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol*. 2012. 107(9):1315-29.

188. Rodier, F., Kim, S. H., Nijjar, T., Yaswen, P. & Campisi, J. Cancer and aging: the importance of telomeres in genome maintenance. *The international journal of biochemistry & cell biology*. 2004. 37, 977-990.
189. Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J Natl Cancer Inst*. 1997; 89:175862.
190. Rodriguez-Santiago B, Armengol L. Next generation sequencing technology in pre- and postnatal genetic diagnosis. *Diagn Prenat* 2012. 23:56-66.
191. Roepman, Schlicker, Tabernero, Majewski, Tian, Moreno, Snel, Chresta, Rosenberg, Nitsche, Macarulla, Capella, Salazar, Orphanides, Wessels, Bernards, and Simon. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int J Cancer*. 2014. 134(3): 552–562.
192. Rubin H. Fields and field cancerization: The preneoplastic origins of cancer. *BioEssays*. 2011. 33(3):224–31.
193. Saeki T, Salomon D, Gullick W, et al. Expression of cripto-1 in human colorectal adenomas and carcinomas is related to degree of dysplasia. *Int J Oncol*. 1994; 5: 445-51
194. Sáez-López P, Filipovich Vegas E, Martínez Peromingo J, Jiménez Mola S. Cáncer colorrectal en el anciano. Tratamiento quirúrgico, quimioterápico y aportación desde la geriatría. Colorectal cancer in the elderly. Surgical treatment, chemotherapy, and contribution from geriatrics. *Revista Española de Geriatría y Gerontología*. September–October. 2017. Volume 52, Issue 5, Pages 261-270.
195. Sahnane N, Bernasconi B, Carnevali I, Furlan D, Viel A, Sessa F, Tibiletti MG. Disruption of the APC gene by t(5;7) translocation in a Turcot family. *Cancer Genet*. 2016. 209(3):107-11.
196. Salnikow E, Drung B, Fabre G, Itkin A, Otyepka M, Dencher NA, Schmidt B, Hauß T, Trouillas P, Bechinger B. Lipid bilayer position and orientation of novel carprofens, modulators of γ -secretase in Alzheimer's disease. *Biochim Biophys Acta Biomembr*. 2018.1860(11):2224-2233.
197. Sameer AS, Nissar S, Fatima K. Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *Eur J Cancer Prev*. 2014. 23(4):246-57.
198. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol*. 2014. 32, 347–55.
199. Sandomenico A, Ruvo M. Targeting Nodal and Cripto-1: perspectives inside dual potential theranostic cancer biomarkers. *Curr Med Chem*. 2018. doi: 10.2174/0929867325666180912104707.
200. Sarvepalli S, Burke CA, Monachese M, Lopez R, Leach BH, Laguardia L, O'Malley M, Kalady MF, Church JM. Web-Based Model for Predicting Time to Surgery in Young Patients with Familial Adenomatous Polyposis: An Internally Validated Study. *Am J Gastroenterol*. 2018.

201. Schepers, D., Tortora, G., Morisaki, H., MacCarrick, G., Lindsay, M., Liang, D. Loeyes, B. A mutation update on the LDS-associated genes TGF β 2/3 and SMAD2/3 . *Human Mutation*. 2018. 39(5), 621–634. doi:10.1002/humu.23407
202. Schier AF. Nodal signaling in vertebrate development. *Annu Rev Cell Dev Biol*. 2003. 19:589–621.
203. Schneider NI, Langner C. Prognostic stratification of colorectal cancer patients: current perspectives. *Cancer Manag Res*. 2014. 6:291-300.
204. Schneider R, Schneider C, Jakobeit C, Fürst A, Möslin G. Gender-Specific Aspects of Lynch Syndrome and Familial Adenomatous Polyposis. *Viszeralmedizin*. 2014. 30(2):82-8.
205. Sengupta N, Yau C, Sakthianandeswaren A, Mouradov D, Gibbs P, Suraweera N, Cazier JB, Polanco-Echeverry G, Ghosh A, Thaha M, Ahmed S, Feakins R, Propper D, et al. Analysis of colorectal cancers in British Bangladeshi identifies early onset, frequent mucinous histotype and a high prevalence of RFX1 deletion. *Mol Cancer*. 2013. 12:1.
206. Shani, G., Fischer, W. H., Justice, N. J., Kelber, J. A., Vale, W., & Gray, P. C. GRP78 and Cripto Form a Complex at the Cell Surface and Collaborate To Inhibit Transforming Growth Factor Signaling and Enhance Cell Growth. *Molecular and Cellular Biology*. 2007. 28(2), 666–677.
207. Sharma AK, Nymark M, Sparstad T, Bones AM, Winge P. Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation. *Sci Rep*. 2018, 8(1):14401.
208. Shen MM. Nodal signaling: developmental roles and regulation. *Development*. 2007. 134, 1023-1034.
209. Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*. 2003. 113: 685–700
210. Siegel RL, Jemal A, Ward EM. Increase in incidence of colorectal cancer among young men and women in the United States. *Cancer Epidemiol Biomarkers Prev*. 2009. 18(6):1695-8.
211. Silla IO, Rueda D, Rodríguez Y, García JL, de la Cruz Vigo F, Perea J. Early-onset colorectal cancer: a separate subset of colorectal cancer. *World J Gastroenterol*. 2014. 20: 17288-17296.
212. Silva SM, Rosa VF, Santos AC, Almeida RM, Oliveira PG, Sousa JB. Influence of patient age and colorectal polyp size on histopathology findings. *Arq Bras Cir Dig*. 2014 Apr-Jun;27(2):109-13.
213. Smith TR, Levine EA, Freimanis RI, Akman SA, Allen GO, Hoang KN, Liu-Mares W, Hu JJ. Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. *Carcinogenesis*. 2008. 29(11):2132-8.
214. Sociedad Española de Oncología Médica. “Las cifras del cáncer en España, 2018”. https://seom.org/seomcms/images/stories/recursos/Las_Cifras_del_cancer_en_Espana2018.pdf.

215. Solera Albero J, Tárraga López PJ, López Cara MA, Celada Rodríguez A, Cerdán Oliver M, Ocaña López JM. Influencia de la dieta y los estilos de vida en el cáncer colorrectal. *Rev Esp Enferm Dig.* 2007; 99: 190- 200.
216. Song YK, Park YS, Seon CS, Lim HJ, Son BK, Ahn SB, et al. Alcohol drinking increased the risk of advanced colorectal adenomas. *Intest Res.* 2015. 13(1):74-9.
217. Soo-Kyung Park, Chang Seok Song, Hyo-Joon Yang, Yoon Suk Jung, Kyu Yong Choi, Dong Hoe Koo, Kyung Eun Kim, Kyung Uk Jeong, Hyung Ook Kim, Hungdai Kim, Ho-Kyung Chun, Dong Il Park. Field Cancerization in Sporadic Colon Cancer. *Gut and Liver.* 2016. Vol. 10, No. 5,
218. Stepanchick A, Zhi H, Cavanaugh A, Rothblum K, Schneider D, Rothblum L. DNA Binding by the Ribosomal DNA Transcription Factor Rrn3 Is Essential for Ribosomal DNA Transcription. *J Biol Chem.* 2013. 29; 288(13): 9135–9144.
219. Sternberg S. H., Redding S., Jinek M., Greene E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature.* 2014. 507, 62–67.
220. Stigliano V, Sanchez-Mete L, Martayan A, Anti M. Early-onset colorectal cancer: a sporadic or inherited disease? *World J Gastroenterol.* 2014. 21;20(35):12420-30.
221. Stigliano V, Sanchez-Mete L, Martayan A, Diodoro M, Casini B, Sperduti I, et al. Early-onset colorectal cancer patients without family history are “at very low risk” for lynch syndrome. *J Exp Clin Cancer Res.* 2014. 2;33:1.
222. Strizzi L, Bianco C, Normanno L, et al. Epithelial mesenchymal transition is a characteristic of hyperplasias and tumors in mammary gland from MMTV-Cripto-1 transgenic mice. *J. Cell. Physiol.* 2004. 201: 266-276.
223. Sung, C. O., Choi, C. H., Ko, Y.-H., Ju, H., Choi, Y.-L., Kim, N., Kim, B.-G. Integrative analysis of copy number alteration and gene expression profiling in ovarian clear cell adenocarcinoma. *Cancer Genetics.* 2013. 206(5), 145–153.
224. Supek, F., Minana, B., Valcarcel, J., Gabaldon, T., and Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell.* 2014. 156, 1324-1335.
225. Sweetser S, Smyrk TC, Sinicrope FA. Serrated colon polyps as precursors to colorectal cancer. *Clin Gastroenterol Hepatol.* 2013. 11(7):760-7.
226. Tamás Arányi, Marcin Ratajewski, Viola Bardóczy, Lukasz Pulaski, András Bors, Attila Tordai and András Váradi. Identification of a DNA methylation-dependent activator sequence in the pseudoxanthoma elasticum gene, ABCC6. *J. Biol. Chem.* 2005.
227. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, Tusquets I, Albanell J, Rodon J, Taberero J, Torres C de, Dienstmann R, et al. 2018. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* 10.

228. Tamborero Noguera D, Rubio Pérez C, Déu Pons J, Schroeder MP, Vivancos Prellezo A, Rovira Guerín A, Tusquets I, Albanell Mestres J, Rodon J, Tabernero Cartula J, Dienstman R, González-Pérez A, et al. 2016. OncodriveMUT.
229. Topczewska JM, Postovit LM, Margaryan NV, et al. Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat Med*. 2006. 12(8):925–932.
230. Tsai, S.Q. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*. 2014. 32(6), pp.569–576.
231. Thinakaran G, Harris C, Tamara Ratovitski, Frances Davenport, Hilda H. Slunt, Donald L. Price, David R. Borchelt. 1997. Evidence That Levels of Presenilins (PS1 and PS2) Are Coordinately Regulated by Competition for Limiting Cellular Factors*. *The Journal of Biological Chemistry*. 1997. Vol. 272, No. 45, pp. 28415–28422.
232. Tieu AH, Edelstein D, Axilbund J, Romans KE, Brosens LA, Wiley E, Hylind L, Giardiello FM. Clinical Characteristics of Multiple Colorectal Adenoma Patients Without Germline APC or MYH Mutations. *J Clin Gastroenterol*. 2016. 50(7):5848.
233. Topczewska, J. M., Postovit, L. M., Margaryan, N. V., Sam, A., Hess, A. R., Wheaton, W. W., Nickoloff, B. J., Topczewski, J. and Hendrix, M. J. Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat. Med*. 2006.12, 925-932.
234. Tropeano M, Ahn JW, Dobson RJB et al: Male-biased autosomal effect of 16p13.11 copy number variation in neurodevelopmental disorders. *PLoS One*. 2013. 8: e61365.
235. Tuefferd, M. et al. Genome-wide copy number alterations detection in fresh frozen and matched FFPE samples using SNP 6.0 arrays. *Genes. Chromosomes Cancer*. 2008.47, 957–964.
236. Ueno, T., Emi, M., Sato, H., Ito, N., Muta, M., Kuroi, K., Toi, M. Genome-wide copy number analysis in primary breast cancer. *Expert Opinion on Therapeutic Targets*. 2012. 16(sup1), S31–S35.
237. Ulaner, G. A., Hu, J. F., Vu, T. H., Giudice, L. C. & Hoffman, A. R. Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts. *Cancer research*. 1998. 58, 4168-4172.
238. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004. 96:261-8.
239. Urnov, F.D. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*. 2010. 11, pp.636–646.
240. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. *Cell*. 2011. 147: 275-92.

241. Valle L. Genetic predisposition to colorectal cancer: Where we stand and future perspectives. *World Journal of Gastroenterology*. 2014. 20(29): 9828-9849.
242. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum*. 1991. 34:424-5.
243. Vasen HF, Watson P, Mecklin JP and Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology*. 1999. 116:1453-6.
244. Velter C, Caussade P, Fricker JP, Cribier B. [Muir-Torre syndrome and Turcot syndrome]. *Ann Dermatol Venerol*. 2017. 144(8-9):525-529.
245. Venugopal A, Stoffel EM. Colorectal Cancer in Young Adults. *Curr Treat Options Gastroenterol*. 2019.
246. Wang H, Chen Z, Chang H, et al. Expression of glia maturation factor γ is associated with colorectal cancer metastasis and its downregulation suppresses colorectal cancer cell migration and invasion in vitro. *Oncol Rep*. 2017.37(2):929–936.
247. Wang, H. One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*. 2013.153, pp.910–918.
248. Wang, J., Duncan, D., Shi,Z. & Zhang, B. WEB-based Gene SeT AnaLysis Toolkit (WebGestalt): update 2013". 2013. *Nucleic Acids Research*. 41:W77-83.
249. Wang J-Y, Wang Y-H, Jao S-W, y col. Molecular mechanisms underlying the tumorigenesis of colorectal adenomas: correlation to activated K-ras oncogene. *Oncol Rep*. 2006. 16: 1245-52.
250. Wang, S. J. et al. The relationship between telomere length and telomerase activity in gynecologic cancers. *Gynecologic oncology*. 2002. 84, 81-84, doi:10.1006/gyno.6483.
251. Wenchen Gong, Baocun Sun, Xiulan Zhao, Danfang Zhang, Junying Sun, Tiejun Liu, Qiang Gu, Xueyi Dong, Fang Liu, Yong Wang, Xian Lin, Yanlei Li. Nodal signaling promotes vasculogenic mimicry formation in breast cancer via the Smad2/3 pathway. *Oncotarget*. 2016; 7:70152-70167.
252. Wilhelm J. Next-generation DNA sequencing techniques. *N Biotechnol* 2009;25.
253. Wong HL, Peters U, Hayes RB, Huang WY, Schatzkin A, Bresalier RS, et al. Polymorphisms in the adenomatous polyposis coli (APC) gene and advanced colorectal adenoma risk. *Eur J Cancer*. 2010.
254. Wood JP, Howells LM, Brown K, Thomas AL. Assessing barriers to a rational chemoprevention trial design in young patients with familial adenomatous polyposis. *Eur J Cancer Prev*. 2017. 26(4):277-284.
255. Worthley DL, Whitehall VL, Spring KJ, Leggett BA. Colorectal carcinogenesis: road maps to cancer. *World J Gastroenterol*. 2007. 13(28):3784-91.

256. Wu QJ, Vogtman E, Zhang W, Xie L, Yang WS, Tan YT, et al. cancer incidence among adolescents and young adults in urban Shanghai, 1973–2005. *PLoS One*. 2012.7(8):e42607.
257. Xu P, Huang S, Li J, Zou Y, Gao M, Kang R, Yan J, Gao X, Gao Y. A novel splicing mutation in the PKD1 gene causes autosomal dominant polycystic kidney disease in a Chinese family: a case report. *BMC Med Genet*. 2018. 13;19(1):198. doi: 10.1186/s.
258. Yamamoto, T. Targeted Genome Editing Using Site-Specific Nucleases ZFNs, TALENs, and the CRISPR/Cas9 System. Springer. 2015. pp. 205.
259. Yamane L, Scapulatempo-Neto C, Reis RM, Guimarães DP. Serrated pathway in colorectal carcinogenesis. *World J Gastroenterol*. 2014. 20(10):2634-40.
260. Yang Y, Chu FH, Xu WR, Sun JQ, Sun X, Ma XM, Yu MW, Yang GW, Wang XM. Identification of regulatory role of DNA methylation in colon cancer gene expression via systematic bioinformatics analysis. *Medicine (Baltimore)*. 2017. 96(47):e8487.
261. Yang W, Ma J, Zhou W, Li Z, Zhou X, Cao B, Zhang Y, Liu J, Yang Z, Zhang H, Zhao Q, Hong L, Fan D. Identification of hub genes and outcome in colon cancer based on bioinformatics analysis. *Cancer Manag Res*. 2018. 11:323-338.
262. Yeo SA, Chew MH, Koh PK, Tang CL. Young colorectal carcinoma patients do not have a poorer prognosis: a comparative review of 2,426 cases. *Tech Coloproctol*. 2013. 17: 653-661.
263. Yiu R, Qiu H, Lee SH, García-Aguilar J. Mechanisms of microsatellite instability in colorectal cancer patients in different age groups. *Dis Colon Rectum*. 2005. 48: 2061-2069.
264. Yuehua Gong, Ying Guo, Zuping He. Nodal Promotes the Self-Renewal of Human Colon Cancer Stem Cells via an Autocrine Manner through Smad2/3 Signaling Pathway. *BioMed Res Int*. 2014. 364134.
265. Zhang H, Xu C, Sun W. Silencing of nodal modulator 1 inhibits the differentiation of P19 cells into cardiomyocytes. *Experimental Cell Research*. 2015. Volume 331, Issue 2, Pages 369–376.
266. Zhang YE. Non-Smad pathways in TGF-beta signaling. *Cell Research*. 2009; 19:128–39. [PubMed: 19114990].
267. Zhong W, Lai Y, Yu T, et al. Wnt and Nodal signaling simultaneously induces definitive endoderm differentiation of mouse embryonic stem cells. *Rom J Morphol Embryol*. 2017. 58(2):527–535.
268. Zhuo, C., Hu, D., Li, J., Yu, H., Lin, X., Chen, Y. Yang, C. Downregulation of Activin A Receptor Type 2A Is Associated with Metastatic Potential and Poor Prognosis of Colon Cancer. *Journal of Cancer*. 2018. 9(19), 3626–3633.
269. Zhu B, Jiang L, Huang T, Zhao Y, Liu T, Zhong Y, Li X, Campos A, Pomeroy K, Masliah E, Zhang D, Xu H. ER-associated degradation regulates Alzheimer's amyloid pathology and memory function by modulating γ -secretase activity. *Nat Commun*. 2017. 13;8(1):1472.

ANEXOS



Anexo 1. Datos de la media de los valores de C_T obtenidos para las curvas de amplificación tanto de *TEL1* (telómero) como de *36B4* (control endógeno) para cada tumor analizado.

MUESTRA	C_T TEL1	C_T 36B4	EDAD	DELECIÓN 16p13.12-13.11
CASO 1	32,7228546	23,9174213	>70	NO
CASO 2	34,9638367	28,0150013	>70	NO
CASO 3	31,8219452	24,809782	>70	NO
CASO 4	32,8172531	26,0898361	>70	NO
CASO 5	30,5389061	24,0592957	>70	NO
CASO 6	30,9576244	24,2174435	>70	NO
CASO 7	35,4708786	28,4204292	>70	NO
CASO 8	30,35392	22,9889393	>70	NO
CASO 9	33,1305656	24,3427429	>70	NO
CASO 10	32,0359993	24,3416348	>70	NO
CASO 11	29,283741	25,6784401	>70	NO
CASO 12	30,73773	28,7873802	>70	NO
CASO 13	27,7372475	25,982851	>70	NO
CASO 14	27,3012981	26,2872219	>70	NO
CASO 15	28,5079994	26,9000111	>70	NO
CASO 16	29,9923267	26,8315983	>70	NO
CASO 17	29,7106514	28,1005974	>70	NO
CASO 18	31,9935665	26,6266994	>70	NO
CASO 19	28,8125424	25,5506229	>70	NO
CASO 20	29,6389046	26,5137272	>70	NO
CASO 21	32,6069565	30,1904526	>70	NO
CASO 22	30,9827995	27,1442738	>70	NO
CASO 23	30,9844265	27,4294949	>70	NO
CASO 24	32,8518448	29,6471786	>70	NO
CASO 25	30,9374046	28,6842537	>70	NO
CASO 26	31,9787178	27,0097179	>70	NO
CASO 27	29,969223	26,1863461	>70	NO
CASO 28	29,1663761	26,2418461	>70	NO
CASO 29	29,8781834	26,255476	>70	NO
CASO 30	29,3479347	27,2089748	>70	NO
CASO 31	29,9779816	27,3255463	>70	NO
CASO 32	33,3193169	31,5799141	>70	NO
CASO 33	29,8130856	26,3496742	>70	NO
CASO 34	31,3334827	26,5676422	>70	NO
CASO 35	32,428833	26,7369938	>70	NO
CASO 36	28,3155746	25,3262558	>70	SI
CASO 37	26,2654705	22,7158413	>70	SI
CASO 38	23,1368599	22,2935371	>70	SI
CASO 39	33,4560661	26,1830807	>70	SI
CASO 40	29,9244556	27,2004604	>70	SI
CASO 41	30,3374977	26,9279404	>70	SI
CASO 42	30,8683033	27,1572037	>70	SI
CASO 43	30,3022156	27,7963486	>70	SI
CASO 44	28,3775425	26,3787384	>70	SI
CASO 45	30,0705948	25,9521828	>70	SI
CASO 46	31,3713131	25,3140717	>70	SI

Anexo 2. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes jóvenes (qPCR).

MUESTRA	EDAD	ALTERACION 16p	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	<45	SI	0,013	0,017	0,011	Homocigoto
CASO 2	<45	SI	0,074	0,083	0,063	Homocigoto
CASO 3	<45	SI	0,078	0,098	0,062	Homocigoto
CASO 4	<45	SI	0,057	0,067	0,048	Homocigoto
CASO 5	<45	SI	0,127	0,155	0,103	Homocigoto
CASO 6	<45	SI	0,272	0,308	0,24	Homocigoto
CASO 7	<45	SI	0,047	0,052	0,043	Homocigoto
CASO 8	<45	SI	0,105	0,157	0,071	Homocigoto
CASO 9	<45	SI	0,121	0,136	0,108	Homocigoto
CASO 10	<45	SI	0,045	0,053	0,038	Homocigoto
CASO 11	<45	SI	0,069	0,082	0,059	Homocigoto
CASO 12	<45	SI	0,086	0,101	0,073	Homocigoto
CASO 13	<45	SI	0,027	0,033	0,023	Homocigoto
CASO 14	<45	SI	0,038	0,046	0,032	Homocigoto
CASO 15	<45	SI	0,02	0,023	0,018	Homocigoto
CASO 16	<45	SI	0,056	0,071	0,044	Homocigoto
CASO 17	<45	SI	0,071	0,077	0,066	Homocigoto
CASO 18	<45	SI	0,094	0,108	0,082	Homocigoto
CASO 19	<45	SI	0,037	0,106	0,013	Homocigoto
CASO 20	<45	SI	0,179	0,224	0,143	Homocigoto
CASO 21	<45	NO	0,044	0,056	0,034	Homocigoto
CASO 22	<45	NO	0,006	0,008	0,005	Homocigoto
CASO 23	<45	NO	0,022	0,032	0,015	Homocigoto
CASO 24	<45	NO	0,014	0,017	0,011	Homocigoto
CASO 25	<45	NO	0,059	0,073	0,047	Homocigoto
CASO 26	<45	NO	0,92	1,113	0,76	Homocigoto
CASO 27	<45	NO	0,084	0,096	0,073	Homocigoto
CASO 28	<45	NO	0,187	0,201	0,174	Homocigoto
CASO 29	<45	NO	0,054	0,08	0,036	Homocigoto
CASO 30	<45	NO	0,098	0,114	0,083	Homocigoto
CASO 31	<45	NO	0,063	0,079	0,051	Homocigoto
CASO 32	<45	NO	0,031	0,054	0,022	Homocigoto
CASO 33	<45	NO	0,074	0,089	0,061	Homocigoto
CASO 34	<45	NO	0,114	0,12	0,108	Homocigoto

Anexo 3. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes jóvenes (qPCR).

MUESTRA	EDAD	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	<45	0,013	0,017	0,011	Homocigoto
CASO 2	<45	0,074	0,083	0,063	Homocigoto
CASO 3	<45	0,078	0,098	0,062	Homocigoto
CASO 4	<45	0,057	0,067	0,048	Homocigoto
CASO 5	<45	0,127	0,155	0,103	Homocigoto
CASO 6	<45	0,272	0,308	0,24	Homocigoto
CASO 7	<45	0,047	0,052	0,043	Homocigoto
CASO 8	<45	0,105	0,157	0,071	Homocigoto
CASO 9	<45	0,121	0,136	0,108	Homocigoto
CASO 10	<45	0,045	0,053	0,038	Homocigoto
CASO 11	<45	0,069	0,082	0,059	Homocigoto
CASO 12	<45	0,086	0,101	0,073	Homocigoto
CASO 13	<45	0,027	0,033	0,023	Homocigoto
CASO 14	<45	0,038	0,046	0,032	Homocigoto
CASO 15	<45	0,02	0,023	0,018	Homocigoto
CASO 16	<45	0,056	0,071	0,044	Homocigoto
CASO 17	<45	0,071	0,077	0,066	Homocigoto
CASO 18	<45	0,094	0,108	0,082	Homocigoto
CASO 19	<45	0,037	0,106	0,013	Homocigoto
CASO 20	<45	0,179	0,224	0,143	Homocigoto
CASO 21	<45	0,044	0,056	0,034	Homocigoto
CASO 22	<45	0,006	0,008	0,005	Homocigoto
CASO 23	<45	0,022	0,032	0,015	Homocigoto
CASO 24	<45	0,014	0,017	0,011	Homocigoto
CASO 25	<45	0,059	0,073	0,047	Homocigoto
CASO 26	<45	0,92	1,113	0,76	Homocigoto
CASO 27	<45	0,084	0,096	0,073	Homocigoto
CASO 28	<45	0,187	0,201	0,174	Homocigoto
CASO 29	<45	0,054	0,08	0,036	Homocigoto
CASO 30	<45	0,098	0,114	0,083	Homocigoto
CASO 31	<45	0,063	0,079	0,051	Homocigoto
CASO 32	<45	0,031	0,054	0,022	Homocigoto
CASO 33	<45	0,074	0,089	0,061	Homocigoto
CASO 34	<45	0,114	0,12	0,108	Homocigoto
CASO 35	<45	0,06	0,071	0,05	Homocigoto
CASO 36	<45	0,594	0,728	0,37	Heterocigoto
CASO 37	<45	0,507	0,547	0,469	Heterocigoto
CASO 38	<45	0,745	0,857	0,647	Normal
CASO 39	<45	0,274	0,685	0,252	Homocigoto
CASO 40	<45	0,299	0,574	0,255	Homocigoto
CASO 41	<45	1,074	1,841	0,627	Normal
CASO 42	<45	0,409	0,437	0,383	Heterocigoto
CASO 43	<45	0,645	1,015	0,412	Heterocigoto

CASO 44	<45	0,421	0,655	0,335	Heterocigoto
CASO 45	<45	1,284	1,581	1,017	Normal
CASO 46	<45	1,254	1,429	1,167	Normal
CASO 47	<45	1,197	1,556	0,889	Normal
CASO 48	<45	1,1	1,271	0,952	Normal
CASO 49	<45	1,385	1,442	1,274	Normal
CASO 50	<45	0,813	0,991	0,675	Normal
CASO 51	<45	0,888	1,016	0,776	Normal
CASO 52	<45	0,749	0,84	0,667	Normal
CASO 53	<45	0,05	0,06	0,042	Homocigoto
CASO 54	<45	0,139	0,148	0,131	Homocigoto
CASO 55	<45	0,034	0,037	0,031	Homocigoto
CASO 56	<45	1,19	1,374	0,975	Normal
CASO 57	<45	0,96	1,112	0,828	Normal
CASO 58	<45	1,054	1,16	0,862	Normal
CASO 59	<45	0,213	0,318	0,154	Homocigoto
CASO 60	<45	0,407	0,459	0,363	Heterocigoto
CASO 61	<45	0,413	0,438	0,39	Heterocigoto
CASO 62	<45	0,023	0,032	0,016	Homocigoto
CASO 63	<45	0,044	0,083	0,024	Homocigoto
CASO 64	<45	0,533	0,857	0,425	Heterocigoto
CASO 65	<45	0,021	0,049	0,009	Homocigoto
CASO 66	<45	0,41	0,644	0,349	Heterocigoto
CASO 67	<45	0,511	0,647	0,497	Heterocigoto
CASO 68	<45	0,222	0,47	0,105	Homocigoto
CASO 69	<45	0,022	0,111	0,004	Homocigoto
CASO 70	<45	0,091	0,256	0,015	Homocigoto
CASO 71	<45	0,947	1,516	0,795	Normal
CASO 72	<45	0,006	0,008	0,005	Homocigoto
CASO 73	<45	0,429	0,593	0,31	Heterocigoto
CASO 74	<45	0,097	0,247	0,009	Homocigoto
CASO 75	<45	0,05	0,185	0,037	Homocigoto
CASO 76	<45	0,072	0,103	0,05	Homocigoto
CASO 77	<45	0,03	0,036	0,025	Homocigoto
CASO 78	<45	0,126	0,345	0,098	Homocigoto
CASO 79	<45	0,004	0,007	0,003	Homocigoto
CASO 80	<45	0,065	0,243	0,012	Homocigoto
CASO 81	<45	0,587	0,653	0,529	Heterocigoto
CASO 82	<45	0,42	0,594	0,283	Heterocigoto
CASO 83	<45	0,016	0,019	0,013	Homocigoto
CASO 84	<45	0,045	0,051	0,04	Homocigoto
CASO 85	<45	0,098	0,168	0,057	Homocigoto
CASO 86	<45	0,052	0,254	0,006	Homocigoto
CASO 87	<45	0,097	0,114	0,082	Homocigoto
CASO 88	<45	0,018	0,021	0,015	Homocigoto

CASO 89	<45	0,158	0,247	0,137	Homocigoto
CASO 90	<45	0,077	0,152	0,039	Homocigoto
CASO 91	<45	0,093	0,162	0,054	Homocigoto
CASO 92	<45	0,001	0,022	0,003	Homocigoto
CASO 93	<45	0,147	0,228	0,074	Homocigoto
CASO 94	<45	0,079	0,124	0,057	Homocigoto
CASO 95	<45	0,156	0,261	0,094	Homocigoto

Anexo 4. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes ancianos (qPCR).

MUESTRA	EDAD	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	>70	0,594	0,728	0,37	Heterocigoto
CASO 2	>70	0,507	0,547	0,469	Heterocigoto
CASO 3	>70	0,745	0,857	0,647	Normal
CASO 4	>70	0,274	0,685	0,252	Homocigoto
CASO 5	>70	0,299	0,574	0,255	Homocigoto
CASO 6	>70	1,074	1,841	0,627	Normal
CASO 7	>70	0,409	0,437	0,383	Heterocigoto
CASO 8	>70	0,645	1,015	0,412	Heterocigoto
CASO 9	>70	0,421	0,655	0,335	Heterocigoto
CASO 10	>70	1,284	1,581	1,017	Normal
CASO 11	>70	1,254	1,429	1,167	Normal
CASO 12	>70	1,197	1,556	0,889	Normal
CASO 13	>70	1,1	1,271	0,952	Normal
CASO 14	>70	1,385	1,442	1,274	Normal
CASO 15	>70	0,813	0,991	0,675	Normal
CASO 16	>70	0,888	1,016	0,776	Normal
CASO 17	>70	0,749	0,84	0,667	Normal
CASO 18	>70	1,15	1,388	0,953	Normal
CASO 19	>70	0,96	1,112	0,828	Normal
CASO 20	>70	1,125	1,376	0,92	Normal
CASO 21	>70	1,048	1,105	0,994	Normal
CASO 22	>70	1,382	1,582	1,288	Normal
CASO 23	>70	0,696	0,771	0,529	Normal
CASO 24	>70	0,938	0,981	0,896	Normal
CASO 25	>70	1,21	1,454	1,006	Normal
CASO 26	>70	0,712	0,824	0,617	Normal
CASO 27	>70	1,326	1,553	1,131	Normal
CASO 28	>70	0,889	1,066	0,742	Normal
CASO 29	>70	0,656	0,81	0,532	Normal
CASO 30	>70	1,477	1,56	1,399	Normal
CASO 31	>70	0,822	0,955	0,707	Normal
CASO 32	>70	0,416	0,488	0,354	Heterocigoto
CASO 33	>70	0,891	0,987	0,804	Normal
CASO 34	>70	0,893	1,012	0,789	Normal
CASO 35	>70	1,17	1,561	0,876	Normal
CASO 36	>70	0,873	1,043	0,73	Normal
CASO 37	>70	0,781	0,921	0,663	Normal
CASO 38	>70	0,946	1,022	0,855	Normal
CASO 39	>70	1,067	1,221	0,721	Normal

CASO 40	>70	0,214	0,487	0,175	Homocigoto
CASO 41	>70	0,914	1,362	0,714	Normal
CASO 42	>70	1,201	1,478	1,07	Normal
CASO 43	>70	0,892	1,141	0,746	Normal
CASO 44	>70	0,815	0,953	0,732	Normal
CASO 45	>70	0,441	0,63	0,398	Heterocigoto
CASO 46	>70	0,521	0,754	0,439	Heterocigoto
CASO 47	>70	1,228	1,325	1,098	Normal
CASO 48	>70	0,795	1,276	0,596	Normal
CASO 49	>70	1,205	1,285	1,124	Normal
CASO 50	>70	0,808	1,022	0,715	Normal
CASO 51	>70	0,939	1,245	0,873	Normal
CASO 52	>70	0,828	1,251	0,715	Normal
CASO 53	>70	1,015	1,143	0,806	Normal
CASO 54	>70	1,281	1,303	1,025	Normal
CASO 55	>70	0,756	0,521	0,932	Normal
CASO 56	>70	0,817	1,043	0,722	Normal
CASO 57	>70	0,954	1,018	0,814	Normal
CASO 58	>70	1,324	1,407	1,162	Normal
CASO 59	>70	0,514	0,658	0,331	Heterocigoto
CASO 60	>70	0,825	0,914	0,801	Normal
CASO 61	>70	0,954	1,32	0,755	Normal
CASO 62	>70	0,893	1,022	0,732	Normal
CASO 63	>70	1,287	1,405	1,124	Normal
CASO 64	>70	1,141	1,353	0,962	Normal
CASO 65	>70	1,23	1,452	0,980	Normal
CASO 66	>70	0,783	0,905	0,614	Normal
CASO 67	>70	0,981	1,24	0,732	Normal

Anexo 5. Resultados normalizados de la cantidad de DNA (RQ) amplificada para tumores diagnosticados en pacientes de edades comprendidas entre los 46 y los 69 años (grupo intermedio) (qPCR).

MUESTRA	EDAD	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	46-69	1,375	1,485	1,123	NORMAL
CASO 2	46-69	0,851	1,19	0,633	NORMAL
CASO 3	46-69	1,31	1,456	1,115	NORMAL
CASO 4	46-69	1,117	1,32	0,867	NORMAL
CASO 5	46-69	0,842	1,247	0,615	NORMAL
CASO 6	46-69	0,732	0,958	0,617	NORMAL
CASO 7	46-69	0,488	0,64	0,396	Heterocigoto
CASO 8	46-69	0,621	0,804	0,514	Heterocigoto
CASO 9	46-69	0.603	0,732	0,487	Heterocigoto
CASO 10	46-69	1,356	1,522	1,119	NORMAL
CASO 11	46-69	0,803	0,966	0,682	NORMAL
CASO 12	46-69	0,237	0,462	0,171	Homocigoto
CASO 13	46-69	0,745	0,857	0,647	NORMAL
CASO 14	46-69	1,288	1,527	1,087	NORMAL
CASO 15	46-69	1,1	1,271	0,952	NORMAL
CASO 16	46-69	1,016	1,139	0,906	NORMAL
CASO 17	46-69	1,227	1,544	1,065	NORMAL
CASO 18	46-69	0,782	0,951	0,632	NORMAL
CASO 19	46-69	1,143	1,323	0,814	NORMAL
CASO 20	46-69	0,774	1,116	0,521	NORMAL
CASO 21	46-69	0,895	1,224	0,603	NORMAL
CASO 22	46-69	0,852	1,117	0,756	NORMAL
CASO 23	46-69	0,195	0,248	0,107	Homocigoto
CASO 24	46-69	0,894	1,024	0,755	Normal
CASO 25	46-69	0,998	0,814	1,223	Normal
CASO 26	46-69	1,255	1,474	1,099	Normal
CASO 27	46-69	1,109	1,246	0,932	Normal
CASO 28	46-69	1,272	1,369	1,115	Normal
CASO 29	46-69	0,834	0,954	0,706	Normal
CASO 30	46-69	0,921	1,314	0,762	Normal
CASO 31	46-69	0,716	0,924	0,532	Normal
CASO 32	46-69	0,215	0,372	0,181	Homocigoto
CASO 33	46-69	0,447	0,382	0,602	Heterocigoto
CASO 34	46-69	0,552	0,739	0,342	Heterocigoto
CASO 35	46-69	1,244	1,481	1,023	Normal
CASO 36	46-69	1,154	1,339	1,017	Normal
CASO 37	46-69	1,297	1,528	0,659	Normal
CASO 38	46-69	1,13	1,341	0,852	Normal
CASO 39	46-69	0,766	0,944	0,614	Normal

CASO 40	46-69	0,204	0,142	0,367	Homocigoto
CASO 41	46-69	0,744	1,094	0,615	Normal
CASO 42	46-69	0,923	1,225	0,741	Normal
CASO 43	46-69	0,884	1,116	0,791	Normal
CASO 44	46-69	1,104	1,362	0,928	Normal
CASO 45	46-69	0,487	0,603	0,391	Heterocigoto
CASO 46	46-69	0,629	0,734	0,506	Heterocigoto
CASO 47	46-69	0,824	1,117	0,698	Normal
CASO 48	46-69	0,736	0,877	0,604	Normal
CASO 49	46-69	1,271	1,496	0,972	Normal
CASO 50	46-69	0,991	1,277	0,855	Normal
CASO 51	46-69	1,331	1,457	1,11	Normal
CASO 52	46-69	0,973	1,162	0,796	Normal
CASO 53	46-69	0,882	0,975	0,688	Normal
CASO 54	46-69	1,227	1,355	1,1	Normal
CASO 55	46-69	1,333	1,559	1,025	Normal
CASO 56	46-69	0,739	0,991	0,549	Normal
CASO 57	46-69	0,836	0,911	0,796	Normal
CASO 58	46-69	0,822	1,138	0,751	Normal
CASO 59	46-69	0,588	0,674	0,362	Heterocigoto
CASO 60	46-69	0,774	1,232	0,588	Normal
CASO 61	46-69	0,699	0,993	0,517	Normal
CASO 62	46-69	1,044	1,274	0,832	Normal
CASO 63	46-69	1,321	1,554	1,114	Normal
CASO 64	46-69	1,298	1,344	0,875	Normal
CASO 65	46-69	0,914	1,122	0,868	Normal
CASO 66	46-69	1,154	1,26	0,962	Normal
CASO 67	46-69	0,847	1,114	0,663	Normal

Anexo 6. Resultados normalizados de la cantidad de DNA (RQ) amplificada para sangre periférica en pacientes jóvenes con delección homocigota en el tumor (qPCR).

MUESTRA	EDAD	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	<45	0,773	1,17	0,542	NORMAL
CASO 2	<45	0,977	0,985	0,853	NORMAL
CASO 3	<45	1,139	1,316	0,986	NORMAL
CASO 4	<45	1,185	1,234	0,978	NORMAL
CASO 5	<45	1,245	1,389	1,098	NORMAL
CASO 6	<45	1.14	1.38	0,845	NORMAL
CASO 7	<45	0,832	0,967	0,715	NORMAL
CASO 8	<45	0,82	0,971	0,692	NORMAL
CASO 9	<45	1,29	0,814	1,525	NORMAL
CASO 10	<45	0,932	1,233	0,705	NORMAL
CASO 11	<45	1,278	1,356	1,117	NORMAL
CASO 12	<45	1,18	1,277	0,984	NORMAL
CASO 13	<45	0,856	0,997	0,735	NORMAL

Anexo 7. Resultados normalizados de la cantidad de DNA (RQ) amplificada en pólipos adenomatosos en pacientes jóvenes con delección homocigota en el tumor (qPCR).

MUESTRA	EDAD	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	<45	0,95	1.189	0,897	NORMAL
CASO 2	<45	1.103	1,354	0,986	NORMAL
CASO 3	<45	0,875	1,104	0,745	NORMAL
CASO 4	<45	1,255	1,354	0,958	NORMAL
CASO 5	<45	0,587	0,653	0,529	Heterocigoto
CASO 6	<45	0,773	1,17	0,542	NORMAL
CASO 7	<45	0,529	0,844	0,374	Heterocigoto
CASO 8	<45	0,524	0,657	0,389	Heterocigoto
CASO 9	<45	0,441	1,089	0,178	Heterocigoto
CASO 10	<45	0,457	0,378	0,615	Heterocigoto
CASO 11	<45	1,118	1,246	1,007	NORMAL
CASO 12	<45	1,074	1,227	0,934	NORMAL
CASO 13	<45	0,556	0,627	0,415	Heterocigoto
CASO 14	46-69	0,486	0,365	0,678	Heterocigoto
CASO 15	46-69	1,124	1,341	0,862	NORMAL
CASO 16	46-69	1,326	1,539	0,975	NORMAL

Anexo 8. Resultados normalizados de la cantidad de DNA (RQ) amplificada en muestra peritumoral en pacientes jóvenes con delección homocigota en el tumor (qPCR).

MUESTRA	EDAD	RQ	RQ MAX.	RQ MIN.	NOMO
CASO 1	<45	0,184	0,265	0,114	Homocigoto
CASO 2	<45	0,143	0,327	0,126	Homocigoto
CASO 3	<45	0,275	0,389	0,187	Homocigoto

Anexo 9. Resultados normalizados de la cantidad de DNA (RQ) amplificada en muestras de otros tumores procedentes de pacientes jóvenes (qPCR).

N LAB.	RQ	RQ MAX.	RQ MIN.	NOMO	TUMOR
CASO 1	0,897	0,105	0,795	NORMAL	Endometrio
CASO 2	0,955	1.201	0,844	NORMAL	Endometrio
CASO 3	1,256	1,07	1,311	NORMAL	Endometrio
CASO 4	1,17	1,451	1,055	NORMAL	Endometrio
CASO 5	0,915	1,023	0,819	NORMAL	Endometrio
CASO 6	1,402	1,679	1,165	NORMAL	Endometrio
CASO 7	1,279	1,601	1,022	NORMAL	Endometrio
CASO 8	1,005	1,257	0,926	NORMAL	Endometrio
CASO 9	1,14	1,365	0,988	NORMAL	Endometrio
CASO 10	0,755	1,188	0,682	NORMAL	Endometrio
CASO 1	1,327	1,517	1,157	NORMAL	Glioblastoma
CASO 2	0,578	0,639	0,526	HETEROCIGOTO	Glioblastoma
CASO 3	0,757	0,905	0,636	NORMAL	Glioblastoma
CASO 4	0,878	0,936	0,824	NORMAL	Glioblastoma
CASO 5	1,066	1,305	0,855	NORMAL	Glioblastoma
CASO 6	0,825	1,003	0,678	NORMAL	Glioblastoma
CASO 7	1,255	1,304	1,188	NORMAL	Glioblastoma
CASO 8	0,774	1,232	0,698	NORMAL	Glioblastoma
CASO 9	0,854	1,189	0,752	NORMAL	Glioblastoma
CASO 10	0,766	1,05	0,659	NORMAL	Glioblastoma
CASO 11	1,023	1,096	0,912	NORMAL	Glioblastoma
CASO 12	0,748	0,837	0,668	NORMAL	Glioblastoma
CASO 13	1,033	1,163	0,918	NORMAL	Glioblastoma
CASO 14	0,949	1,196	0,752	NORMAL	Glioblastoma
CASO 15	1,227	1,352	1,077	NORMAL	Glioblastoma
CASO 1	0,713	0,891	0,575	NORMAL	Mama
CASO 2	0,888	1,016	0,776	NORMAL	Mama
CASO 3	0,749	0,84	0,667	NORMAL	Mama
CASO 4	0,767	0,86	0,685	NORMAL	Mama
CASO 5	0,896	1,032	0,779	NORMAL	Mama
CASO 6	0,556	0,709	0,508	HETEROCIGOTO	Mama
CASO 7	0,783	0,876	0,703	NORMAL	Mama
CASO 8	1,115	1,247	0,997	NORMAL	Mama
CASO 9	0,711	1,029	0,515	NORMAL	Mama
CASO 10	0,109	0,218	0,054	Homocigoto	Mama
CASO 11	0,717	0,884	0,582	NORMAL	Mama
CASO 12	0,886	1,05	0,748	NORMAL	Mama
CASO 13	0,824	1,11	0,725	NORMAL	Mama
CASO 14	1,05	1,146	0,873	NORMAL	Mama
CASO 15	0,954	1,144	0,822	NORMAL	Mama
CASO 16	0,516	0,709	0,375	HETEROCIGOTO	Mama
CASO 17	0,806	1,238	0,627	NORMAL	Mama

CASO 18	0,854	1,024	0,773	NORMAL	Mama
CASO 19	1,122	1,238	0,808	NORMAL	Mama
CASO 20	0,666	0,829	0,535	NORMAL	Mama
CASO 21	0,516	1,021	0,33	HETEROCIGOTO	Mama
CASO 22	1,033	1,429	0,812	NORMAL	Mama
CASO 1	0,774	1,646	0,364	NORMAL	Cabeza y cuello
CASO 2	1,244	1,471	1,102	NORMAL	Cabeza y cuello
CASO 3	0,958	1,276	0,719	NORMAL	Cabeza y cuello
CASO 4	1,415	1,847	1,22	NORMAL	Cabeza y cuello
CASO 5	0,508	0,631	0,41	HETEROCIGOTO	Cabeza y cuello
CASO 6	0,718	0,82	0,632	NORMAL	Cabeza y cuello
CASO 7	0,664	0,766	0,434	NORMAL	Cabeza y cuello
CASO 8	0,481	0,649	0,331	HETEROCIGOTO	Cabeza y cuello
CASO 9	0,675	0,893	0,495	NORMAL	Cabeza y cuello
CASO 10	1,061	1,651	0,603	NORMAL	Cabeza y cuello
CASO 11	1,161	1,535	0,878	NORMAL	Cabeza y cuello
CASO 12	1,347	1,879	0,965	NORMAL	Cabeza y cuello
CASO 13	0,968	1,12	0,837	NORMAL	Cabeza y cuello
CASO 1	0,632	1,082	0,369	NORMAL	Ovario
CASO 2	1,255	1,455	1,033	NORMAL	Ovario
CASO 3	0,725	1,348	0,499	NORMAL	Ovario
CASO 4	0,907	1,209	0,68	NORMAL	Ovario
CASO 5	1,465	1,981	1,038	NORMAL	Ovario
CASO 6	0,484	0,635	0,369	HETEROCIGOTO	Ovario
CASO 7	0,917	1,084	0,776	NORMAL	Ovario
CASO 8	1,312	1,867	0,985	NORMAL	Ovario
CASO 9	1,088	1,384	0,855	NORMAL	Ovario

Anexo 10. Valores de expresión de los genes integrantes de la vía NOD.

Gene Symbol	Description	1.C1	2.C2	3.NORMAL	4.PX458	Fold Change
CASP8	caspase 8, apoptosis-related cysteine peptidase	9,452231	9,838135	10,86774	10,82261	0,4352777
CCL5	chemokine (C-C motif) ligand 5	4,011779	5,07951	6,292878	7,334744	0,20759355
CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma gro	5,563004	4,063625	3,50814	3,599765	2,39389853
CXCL8	chemokine (C-X-C motif) ligand 8	8,961668	5,583646	4,848224	5,11036	4,90198137
CYBA	cytochrome b-245, alpha polypeptide	11,24666	11,57961	12,60719	12,80084	0,40870166
GBP1	guanylate binding protein 1, interferon-inducible	4,237097	4,707614	4,96305	6,389652	0,43407147
GBP3	guanylate binding protein 3	8,568121	9,728691	10,28279	10,60053	0,40802968
IRF7	interferon regulatory factor 7	4,080855	5,18391	6,919415	7,084414	0,19350839
IRF9	interferon regulatory factor 9	5,860827	11,0807	11,30939	11,46014	0,13267776
JUN	jun proto-oncogene	11,19918	10,31521	10,13747	8,957494	2,31291621
MAPK3	mitogen-activated protein kinase 3	9,675529	11,88581	12,61086	12,65814	0,27665684
OAS1	2-5-oligoadenylate synthetase 1	8,896036	13,67212	15,49894	16,21406	0,04203043
OAS2	2-5-oligoadenylate synthetase 2	3,637315	9,049929	14,215	14,93835	0,00332356
OAS3	2-5-oligoadenylate synthetase 3	7,225666	9,755989	11,6384	12,18522	0,09336447
PLCB4	phospholipase C, beta 4	3,3725	3,168012	5,623867	4,751193	0,2647543
STAT1	signal transducer and activator of transcription 1	11,56574	13,6932	15,21744	16,2788	0,11513114
STAT2	signal transducer and activator of transcription 2	5,739167	7,404936	7,563964	8,769703	0,33107225
TAB3	TGF-beta activated kinase 1/MAP3K7 binding pro	9,573515	10,18897	8,445028	8,374645	2,77292001
TMEM173	transmembrane protein 173	8,764249	8,330131	9,424758	9,974742	0,44982633
TRAF5	TNF receptor-associated factor 5	5,307869	5,721229	6,414405	6,876963	0,4565564
TXNIP	thioredoxin interacting protein	5,407413	13,71055	12,41692	10,34055	0,28326937
VDAC3	voltage-dependent anion channel 3	10,7671	10,31176	9,773973	9,174581	2,09239177

Anexo 11. Listado de artículos publicados.

1. Pérez J, Perea J, Arriba M, García JL, García P, González Sarmiento R. Quantitative PCR shows a single NOMO gene. *Genes*. (Proceso de aceptación)
2. Álvaro E, Cano JM, García JL, Brandáriz L, Olmedillas-López S, Arriba M, Rueda D, Rodríguez Y, Cañete Á, Arribas J, Inglada-Pérez L, **Pérez J**, Gómez C, García-Arranz M, García-Olmo D, Goel A, Urioste M, González-Sarmiento R, Perea J. Clinical and Molecular Comparative Study of Colorectal Cancer Based on Age-of-Onset and Tumor Location: Two Main Criteria for Subclassifying Colorectal Cancer. *Int J Mol Sci*. 2019. 20(4). pii: E968.
3. Perea J, García JL, Corchete L, Lumbreras E, Arriba M, Rueda D, Tapial S, **Pérez J**, Vieiro V, Rodríguez Y, Brandáriz L, García-Arranz M, García-Olmo D, Goel A, Urioste M, Sarmiento RG. Redefining synchronous colorectal cancers based on tumor clonality. *Int J Cancer*. 2019. 144(7):1596-1608.
4. Brandariz L, Arriba M, García JL, Cano JM, Rueda D, Rubio E, Rodríguez Y, **Pérez J**, Vivas A, Sánchez C, Tapial S, Pena L, García-Arranz M, García-Olmo D, Urioste M, González-Sarmiento R, Perea J. Differential clinicopathological and molecular features within late-onset colorectal cancer according to tumor location. *Oncotarget*. 2018. 9(20):15302-15311.
5. Tapial S, Rueda D, Arriba M, García J. L, Brandáriz L, **Pérez J**, Rodríguez Y, García-Olmo D, González-Sarmiento R, Urioste M, Perea J. Comment on 'Distinct clinical outcomes of two CIMP-positive colorectal cancer subtypes based on a revised CIMP classification system'. *Br J Cancer*. 2018. 118(3):e3.
6. Perea J, García JL, **Pérez J**, Rueda D, Arriba M, Rodríguez Y, Urioste M, González-Sarmiento R. NOMO-1 gene is deleted in early-onset colorectal cancer. *Oncotarget*. 2017. 24429-24436.
7. Arriba M, García JL, Rueda D, **Pérez J**, Brandariz L, Nutu OA, Alonso L, Rodríguez Y, Urioste M, González-Sarmiento R, Perea J. Unsupervised Analysis of Array Comparative Genomic Hybridization Data from Early-Onset Colorectal Cancer Reveals Equivalence with Molecular Classification and Phenotypes. *Neoplasia*. 2017. 19(1):28-34.
8. Perea J, Arriba M, Rodríguez Y, Rueda D, García JL, **Pérez J**, González-Sarmiento R, Urioste M. Frequency and impact of KRAS mutation in early onset colorectal cancer. *Hum Pathol*. 2017. 61:221-222.
9. Arriba M, Sánchez R, Rueda D, Gómez L, García JL, Rodríguez Y, Pajares JA, **Pérez J**, Urioste M, Sarmiento RG, Perea J. Toward a Molecular Classification of Synchronous Colorectal Cancer: Clinical and Molecular Characterization. *Clin Colorectal Cancer*. 2017. 16(1):31-37.
10. Perea J, Arriba M, Rueda D, Sánchez R, García JL, **Pérez J**, Rodríguez Y, González-Sarmiento R, Urioste M. Comment on 'Wild-type APC prediction of poor prognosis in microsatellite-stable proximal colorectal cancer differs according to the age of onset'. *Br J Cancer*. 2016. 114(10):e7.

11. Perea J, Cano JM, Rueda D, García JL, Inglada L, Osorio I, Arriba M, **Pérez J**, Gaspar M, Fernández-Miguel T, Rodríguez Y, Benítez J, González-Sarmiento R, Urioste M. Classifying early-onset colorectal cancer according to tumor location: new potential subcategories to explore. Am J Cancer Res. 2015. 5(7):2308-13.

12. Arriba M, García JL, Inglada-Pérez L, Rueda D, Osorio I, Rodríguez Y, Álvaro E, Sánchez R, Fernández T, **Pérez J**, Hernández JM, Benítez J, González-Sarmiento R, Urioste M, Perea J. DNA copy number profiling reveals different patterns of chromosomal instability within colorectal cancer according to the age of onset. Mol Carcinog. 2016. 55(5):705-16.

DNA Copy Number Profiling Reveals Different Patterns of Chromosomal Instability Within Colorectal Cancer According to the Age of Onset

María Arriba,¹ Juan L. García,^{2,3} Lucía Inglada-Pérez,^{4,5} Daniel Rueda,⁶ Irene Osorio,⁷ Yolanda Rodríguez,⁸ Edurne Álvaro,⁹ Ricard Sánchez,¹ Tamara Fernández,⁷ Jessica Pérez,^{2,3} Jesús M. Hernández,² Javier Benítez,^{5,10} Rogelio González-Sarmiento,^{2,3} Miguel Urioste,^{5,11} and José Perea^{1,7*}

¹Centre for Biomedical Research of 12 de Octubre University Hospital, Madrid, Spain

²Biomedical Research Institute of Salamanca (IBSAL), University Hospital of Salamanca-USAL-CSIC, Salamanca, Spain

³Institute of Molecular and Cellular Biology of Cancer (IBMCC), University of Salamanca-CSIC, Salamanca, Spain

⁴Hereditary Endocrine Cancer Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

⁵Center for Biomedical Network Research on Rare Diseases (CIBERER), Institute of Health Carlos III, Madrid, Spain

⁶Molecular Biology Laboratory, 12 de Octubre University Hospital, Madrid, Spain

⁷Department of Surgery, 12 de Octubre University Hospital, Madrid, Spain

⁸Department of Pathology, 12 de Octubre University Hospital, Madrid, Spain

⁹Department of Surgery, Infanta Leonor University Hospital, Madrid, Spain

¹⁰Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

¹¹Familial Cancer Clinical Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

Chromosomal instability resulting in copy number alterations is a hallmark of colorectal cancer (CRC). However, few studies have attempted to characterize the chromosomal changes occurring in early-onset CRC in order to compare them with those taking place within the more extensively studied late-onset CRC subset. Our aim was to characterize the genomic profiles of these two groups of colorectal tumors and to compare them to each other. Array comparative genomic hybridization profiling of 146 colorectal tumors (60 early-onset and 86 late-onset) in combination with an unsupervised analysis was used to define common and specific copy number alterations. We found a number of important differences between the chromosomal instability profiles of each age subset. Thus, losses at 1p36, 1p12, 1q21, 9p13, 14q11, 16p13, and 16p12 were significantly more frequent in younger patients, whereas gains at 7q11 and 7q22 were more frequent in older patients. Moreover, the unsupervised analysis stratified the tumors into two clusters, each one of which was enriched in patients from one of the age subsets. Our findings confirm the existence of substantial differences between the chromosomal instability profiles of the two groups which are more important from a qualitative point of view. Further studies are needed to understand the clinicopathological implications of these dissimilarities.

© 2015 Wiley Periodicals, Inc.

Key words: early-onset colorectal cancer; late-onset colorectal cancer; array-based comparative genomic hybridization

INTRODUCTION

In Spain, colorectal cancer (CRC) is the third most frequent neoplasm in men (after prostate cancer and lung cancer) and the second most frequent neoplasm in women (after breast cancer) [1]. It is caused by a progressive accumulation of genetic and epigenetic alterations that lead to the transformation of normal colonic epithelial cells into colon adenocarcinoma cells, in a process called the adenoma-carcinoma sequence [2].

Two distinct mechanisms of genomic instability may be operative in CRC: chromosomal instability (CIN) and microsatellite instability (MSI) [3]. Up to several years ago, these two genetic mechanisms were correlatively associated with the two main carcinogenesis pathways as the “suppressor” and the

Abbreviations: aCGH, array comparative genomic hybridization; CBS, circular binary segmentation; CIN, chromosomal instability; CNA, copy number alteration; CRC, colorectal cancer; FDR, false discovery rate; GII, genomic instability index; GNL, gain normal, loss; MACS, microsatellite and chromosome stable tumors; Mb, megabase; MMR, mismatch repair; MSI, microsatellite instability; MSS, microsatellite stability; SOTA, self organizing tree algorithm.

María Arriba and Juan L. García contributed equally to this work.

Grant sponsor: Spanish Ministry of Health and Consumer Affairs; Grant number: PI10/0683

*Correspondence to: Department of Surgery, Hospital Universitario 12 de Octubre, C/Rosas de Aravaca, 82A, 1^oDcha, E-28023, Madrid, Spain.

Received 12 December 2014; Revised 19 February 2015; Accepted 21 February 2015

DOI 10.1002/mc.22315

Published online 25 March 2015 in Wiley Online Library (wileyonlinelibrary.com).

“mutator” pathways, respectively [4,5]. The CIN pathway accounts for the majority of spontaneous CRCs (85%) [6] and it is considered the result of chromosomal missegregation due to a number of events at the DNA level, including mutations in mitotic checkpoint genes, microtubule spindle defects, and telomere dysfunction. On the other hand, the MSI pathway accounts for approximately 15% [6] of sporadic CRCs and tumors develop by inactivation of the mismatch repair (MMR) system. Initially, MSI and CIN pathways were considered to be mutually exclusive, so it was thought that MSI+ tumors would be CIN-, whereas CIN+ tumors would show microsatellite stability (MSS). However, more recent studies have demonstrated that up to 50% of MSS tumors are CIN- (also called microsatellite and chromosome stable tumors [MACS]), and that about 15% of MSI colorectal tumors are CIN+ [7-9].

Although young age at diagnosis is a characteristic of hereditary forms of cancer, early-onset CRC can also be sporadic. Early-onset MSI tumors are generally related with Lynch syndrome, but early-onset MSS tumors remain still poorly characterized. In comparison with late-onset CRC (patients >70 years old at the time of diagnosis), early-onset MSS tumors are more frequently associated with poor prognosis, invasive phenotype, early metastasis, and familial clustering, and patients with such tumors are in general considered as a high-risk group within CRC [10-12].

It has been proposed that early-onset and late-onset CRCs might evolve in distinct ways, and that there may be a different molecular basis for each age at onset [12]. This would imply different molecular mechanisms of tumorigenesis, which would mean the existence of specific altered loci that might include important genes for tumor progression in the different age groups [12,13]. For this reason, copy number alterations (CNAs) detected by array comparative genomic hybridization (aCGH) may contribute to the identification of new genomic regions involved in tumorigenesis and ultimately lead to the discovery of new cancer-causing genes in patients without known hereditary CRC syndromes.

Some studies have described differences between the CNAs arising in sporadic early-onset and late-onset CRCs, reporting a small number of them specifically associated with each group [14,15]. However, the published results are not unanimous in their conclusions, possibly because of the low frequency of CRC in young people (2-8%) [16,17]. We used aCGH to compare a large group of early-onset CRC with a group of late-onset CRC. Our purpose was to define the differential features of early-onset CRC according to CIN status, as well as to identify potential susceptibility loci mainly for young people.

MATERIALS AND METHODS

Patients, Samples, and Data Collection

We compared two groups of patients differing in the age at onset. A total of 88 individuals with CRC diagnosed at an age of 45 years or younger (range: 16-45 years) were collected at the 12 de Octubre University Hospital in Madrid. We also collected 96 individuals diagnosed during the same period, but at an age of 70 years or older (range: 70-92). All patients, or a first-degree relative in case of death of the index case, provided written consent and the study was approved by the Ethics Committee of our Institution.

Six patients from the early-onset group were excluded because familial adenomatous polyposis was diagnosed. Material for aCGH analysis could be obtained from 60 of the remaining 82 young patients (73.3% with the disease in the colon and 26.7% in the rectum). Of the late-onset group, a total of 86 samples could be adequately processed (61.6% of tumors located in the colon and 38.4% of tumors located in the rectum). Additional clinical, pathological and familial features of the studied patients are shown in Supplementary Table S1.

DNA Isolation and Labeling

A pathologist performed the microscopic inspection of the tumor tissue of paraffin-embedded samples from the index cases. The proportion of tumor cells in the neoplastic material used for DNA extraction exceeded 70% in all cases. DNA was extracted using proteinase K digestion, phenol/chloroform extraction, Phase Lock Gel Light (Eppendorf AG, Hamburg, Germany), and ethanol precipitation. One microgram of genomic DNA (gDNA) from each test and reference sample (Human Genomic DNA [Male G1471/Female G1521] from Promega Corporation, Madison, WI) was labeled using the NimbleGen Dual-Color DNA Labeling Kit and hybridized to a NimbleGen Human CGH 12 × 135K array (Roche NimbleGen, Inc., Reykjavik, Iceland). Extracted gDNA from samples and Promega control DNAs were denatured in the presence of Cy3- (test) or Cy5- (control) labeled random primers and incubated with the Klenow fragment of DNA polymerase, together with 5mM of each dNTP at 37°C for 2h. The reaction was stopped by the addition of 0.5M EDTA, prior to isopropanol precipitation and ethanol washing. After quantification, test and sex matched control samples were combined in equimolar amounts and applied to one of the 12 arrays on the microarray slide.

Hybridizations and Image Acquisition

Comparative genomic hybridization was performed using NimbleGen oligonucleotide microarrays (Roche NimbleGen, Inc.). The whole-genome oligonucleotide microarray contained 135,000

isothermal probes varying in length from 45 to 75 bases and which spanned the entire human genome with a median probe spacing of 12,524 bp. Hybridization was carried out in a Roche NimbleGen Hybridisation Chamber (Roche NimbleGen, Inc.) for 48 h. Slides were washed and scanned using a NimbleGen MS 200 Microarray Scanner (Roche NimbleGen, Inc.).

Preprocessing and Processing

Array image files produced by the MS 200 Data Collection Software were imported into DEVA v1.2.1 (Roche NimbleGen Inc.) for analysis. In addition, segmentation analysis was performed in all cases by using circular binary segmentation (CBS) available in waviCGH software (<http://wavi.bioinfo.cnio.es/>), a web-server application for the analysis and comparison of array-CGH data [18]. Each genomic region exhibiting a copy number change was examined using the UCSC genome browser (<http://genome.ucsc.edu/>) to determine the location and significance of the change.

Degree of Genomic Instability: Number and Length of Alterations

The degree of genomic instability for each tumor was estimated by determining the fraction of altered genome as Bonnet et al. described in 2012 [19]. This index was computed in two different ways. On the one hand, genomic instability was computed as the fraction of altered oligonucleotides where the corresponding segment value was used to determine the altered status of each oligonucleotide, that is, as the fraction of altered genome based on the number of normalized segment values in each case [19]. On the other hand, genomic instability was computed by obtaining the copy number states for each of the windowed probes and determining the number of those assigned as altered in copy number against the total number of windowed probes. This measure, referred to as the genomic instability index (GII), represents the fractions of aberrant genome, and it is calculated by summing the number of probes that are not in the “normal” state as follows:

$$\text{GII}^{\text{total}}: \frac{1}{n} \sum_{i=1}^n (x_i \neq 0) = 1 - \frac{1}{n \sum_{i=1}^n (X_i = 0)},$$

where X_1, \dots, X_n are the probes of the aCGH data for a given patient [19]. The ordinal scale of the aCGH data allows this index to be decomposed into several related scores. For instance, it could be of interest to know the fraction of the genome that is gained and overlook the other types of aberrations. This information can be obtained by considering only probes carrying “+1” labels as follows:

$$\text{GII}^{\text{gain}}: \frac{1}{n} \sum_{i=1}^n (X_i = +1).$$

Similarly, the fraction of deleted genome can be obtained by considering probes carrying “-1” labels as follows:

$$\text{GII}^{\text{loss}}: \frac{1}{n} \sum_{i=1}^n (X_i = -1).$$

Common and Potentially Specific Regions of Copy Number Changes

Frequency plots and a list of the recurrent minimal common regions altered were generated for each tumor subtype. Minimum regions were codified as -1 (losses), zero, or +1 (gains). In order to identify significant minimum regions, univariate analyses were carried out by performing an unconditional logistic regression for each candidate region. A false discovery rate (FDR) was calculated for each P -value as Benjamini et al. described [20], and regions with an $\text{FDR} < 0.09$ were considered significant. All analyses were performed using R Statistical Software [21] (<http://www.r-project.org/>).

Unsupervised Cluster Analysis

For a first unsupervised analysis of the CRCs, all samples were clustered based on “gain = +1, normal = 0, loss = -1.” The cluster was performed using self organizing tree algorithm (SOTA) algorithms (Euclidean squared distances) as implemented in Multi Experiment Viewer 4.8.1 (www.tm4.org/mev.html).

Statistical Analysis

Statistical analysis was performed using SPSS 17.0 (SPSS Inc., Chicago, IL). Comparison of continuous variables (number and size of alterations between different tumor groups and clusters) was done using Student’s two-tailed t -test (for variables of normal distribution) or the Mann-Whitney U test (nonparametric distributions). Differences were considered statistically significant when bilateral P -values were < 0.05 .

RESULTS

Patients and Tumors

The average age at diagnosis for the patients included in the early-onset group was 39.17 ± 5.04 years old, whereas it was 78.07 ± 5.69 for the late-onset group. Clinicopathological and familial characteristics of studied cases are summarized in Supplementary Table S1.

Number and Length of Copy Number Alterations Across Tumor Subtypes

Of the 146 tumors analyzed, 142 (97.26%) showed genomic changes by aCGH, and only in 4 cases (1 early-onset and 3 late-onset) no genomic changes were detected.

The two groups were characterized by a high level of genomic instability (Figure 1 and Supplementary

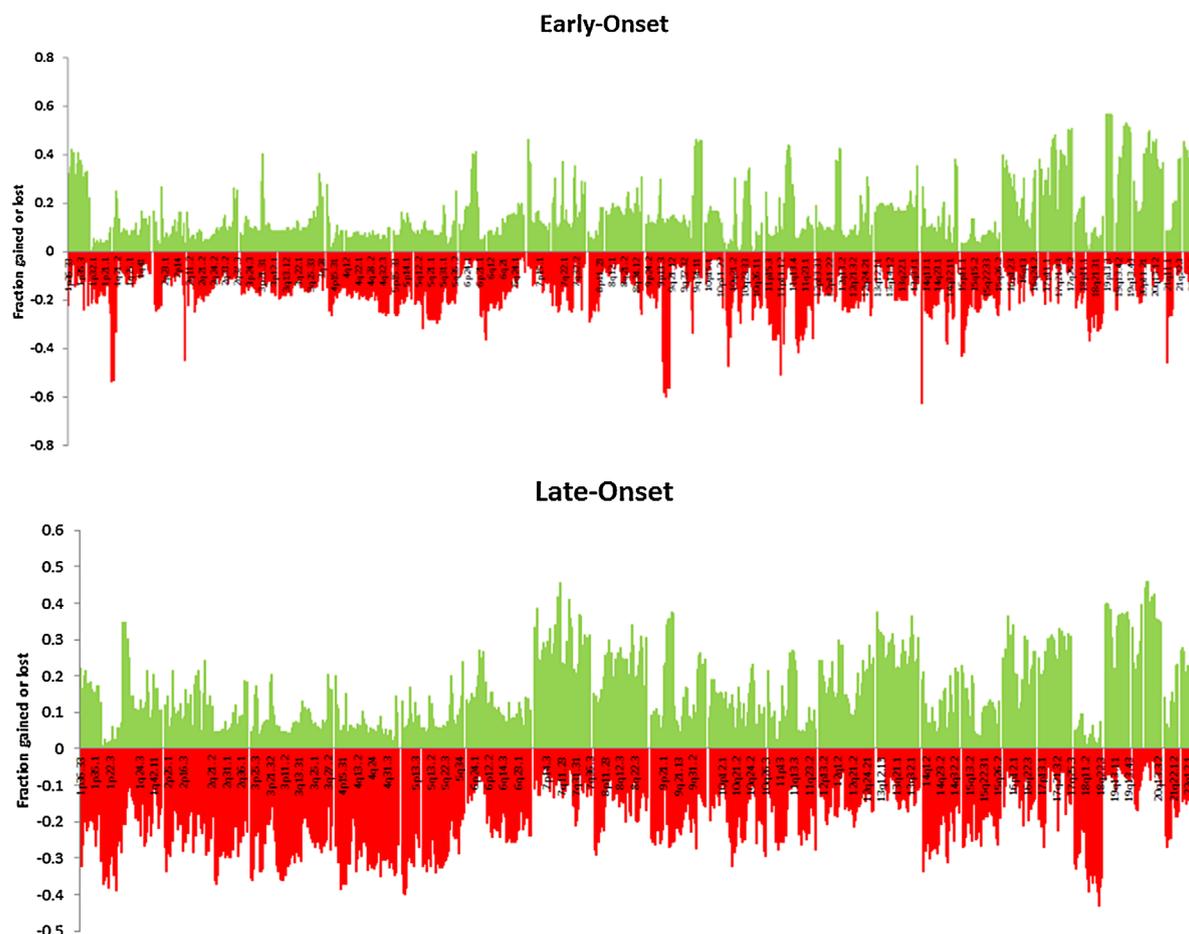


Figure 1. Frequency plots of copy number gains (above zero) and losses (below zero) defined for each subgroup. The fraction gained or lost is plotted on the y-axis versus genomic location on the x-axis.

Figure S1); this was exemplified in the early-onset group by an average of 95.80 altered segments per tumor (45.60 gained vs. 50.20 lost), and in the late-onset group by an average of 150.62 altered segments per tumor (75.87 gained vs. 74.41 lost) (Table 1 and Figure 2A). When these alterations were divided into fragments gained or lost smaller than 1Mb and larger than 1Mb (Figure 2B,C), we found that gains <1 Mb were significantly more common within the late-onset group ($P=0.007$; Figure 2B).

Several of the altered segments showed frequencies greater than or equal to 30%, with losses being more common than gains in both groups (Figure 2D). In the early-onset subset there were 147 recurrent alterations (77 losses vs. 70 gains) that involved 636.92 Mb of the genome, whereas in the late-onset subset there were 135 recurrent alterations (78 losses vs. 57 gains) that involved 590.87 Mb of the genome (Table 1, Figure 2D, and Supplementary Tables S2 and S3). Besides, the GII showed association between the degree of genomic instability and age. Although both groups had a

higher GII for losses than for gains, the difference between both values was more marked within the late-onset group (Table 1 and Figure 3).

Aneuploidy

Alterations in the number of intact chromosomes, known as whole chromosome missegregation, displayed different patterns that could be associated with age at onset. For the entire set of CRCs (146 tumors), 98 cases (67.12%) showed gains or losses of whole chromosomes. The proportion of aneuploidies was similar in both groups (68% in early-onset and 66% in late-onset) with a higher number of gains and losses in late-onset CRC (157 chromosomes lost and 132 gained) than in early-onset CRC (76 chromosomes lost and 102 gained). Considering all samples together, the most frequent losses of whole chromosomes were observed for chromosomes 4, 5, 14, 15, and 18, whereas the most frequent gains corresponded to chromosomes 7, 13, 17, 19, 20, and 22 (Figure 4A).

In the early-onset group the most commonly gained chromosomes were 19 (35%), 22 (33%), 17

Table 1. Global Description of the Genomic Instability Observed in the Early-Onset and Late-Onset CRC Cases

	Early-onset (<45 years old)	Late onset (>70 years old)	P-value ^a
Number of patients	60	86	—
Average number of altered segments per tumor:	95.80 (72.50)	150.62 (52.50)	NS
Total Gains	45.60 (26.00)	75.87 (16.00)	NS
Gains >1 Mb	35.00 (20.00)	43.29 (11.00)	NS
Gains <1 Mb	10.80 (7.50)	32.99 (9.00)	0.007
Total Losses	50.20 (44.00)	74.41 (23.00)	NS
Losses >1 Mb	31.58 (18.00)	46.95 (15.00)	NS
Losses <1 Mb	18.62 (13.00)	27.45 (6.00)	NS
Number of recurrent aberrations ^b :	147	135	NS
Gains	70	57	NS
Losses	77	78	NS
Total length of recurrent aberrations, Mb [SD]:	636.92 [6.64]	590.87 [6.42]	NS
Gains	379.26 [6.70]	293.09 [5.99]	NS
Losses	257.66 [6.48]	297.78 [6.71]	NS
Average GII:			
Gains	0.151062 (0.116119)	0.147055 (0.687489)	NS
Losses	0.165460 (0.091511)	0.200192 (0.103392)	0.004
Normal	0.683470 (0.744741)	0.652745 (0.670491)	NS
Average whole chromosome alterations:			
Gains of chromosomes	1.70 (1.00)	1.53 (1.00)	NS
Losses of chromosomes	1.27 (0.00)	1.83 (0.00)	NS

Mb: megabasepairs; NS: not significant; SD: standard deviation; GII: genomic instability index. Data shown in parenthesis represent median values.

^aStatistical comparison was performed using Student's *t*-test.

^bAberrations displaying a frequency of 30% or higher.

(20%), and 13 (18%), while chromosomes 18 (17%), 4 (15%), and 15 (15%) were the most commonly lost (Figure 4B). On the other hand, the most frequent aneuploidy events in the late-onset group were gains of chromosomes 19 (31%), 20 (20%), 13 (19%), 7 (18%), and 22 (18%), and losses of chromosomes 18 (24%) and 4 (16%) (Figure 4C).

Common and Group-Specific CNAs

Global genomic changes are represented in Figure 1, and Supplementary Tables S2 and S3 summarize the specific chromosomal changes observed with a recurrence greater than or equal to 30% in both tumor subgroups. In the early-onset group, the most recurrent alterations were losses at chromosomes 1p36 (61%), 14q11 (61%), 5q13 (57%), 9p13–q21 (57%), 1p12–q21 (53%), and 11p11–q12 (50%), and gains at chromosomes 19p13–p12 (57%), 19q13 (52%), and 17q24–q25 (50%). However, there were no alterations with a frequency greater than or equal to 50% in the late-onset group. The most recurrent alterations in this last subset were losses at 14q11 (45%) and 5q13 (44%), and gains at 19p13–p12 (47%), 20p11–q11 (47%), 19q13 (42%), 7q21–q22 (42%), and 7q11 (40%).

Table 2 and Supplementary Table S4 show the 24 chromosomal segments occurring with statistically significantly different frequencies in both groups.

There were only three regions more frequently observed in the late-onset group and they corresponded to gains at chromosome 7 (7q11.23, 7q22.1, and 7q11.22). The remaining differential regions were more common within the early-onset tumors; gains at 1p35.2–p35.1 and 3p21.31–p21.1 and losses at 1p35.2–p35.1, 7q11.21, 9p13.1–p11.2, 16p13.12–p13.11, 16p13.11, and 16p12.3 were predominant in the younger patients and very uncommon in the older ones. Moreover, losses at chromosome 1 (p36.33, p12–p11.2, q21.1, and q21.1–q21.2), 3p21.31–p21.1, 7q22.1, 7q35, and 14q11.1–q11.2 were observed in both groups but with a higher frequency in the early-onset group.

Unsupervised Analysis

In addition to a systematic comparison of CNAs in the different age groups, we carried out an unsupervised analysis of the aCGH data to unveil possible associations between particular patterns of genomic changes and the early-onset or late-onset status of the tumors. The unsupervised clustering was performed based on the “gain, normal, loss” (GNL) profile of each tumor (Figure 5), and SOTA algorithms were used.

Unsupervised hierarchical clustering stratified tumors based on their copy number changes into two main clusters. The first cluster (cluster A), was

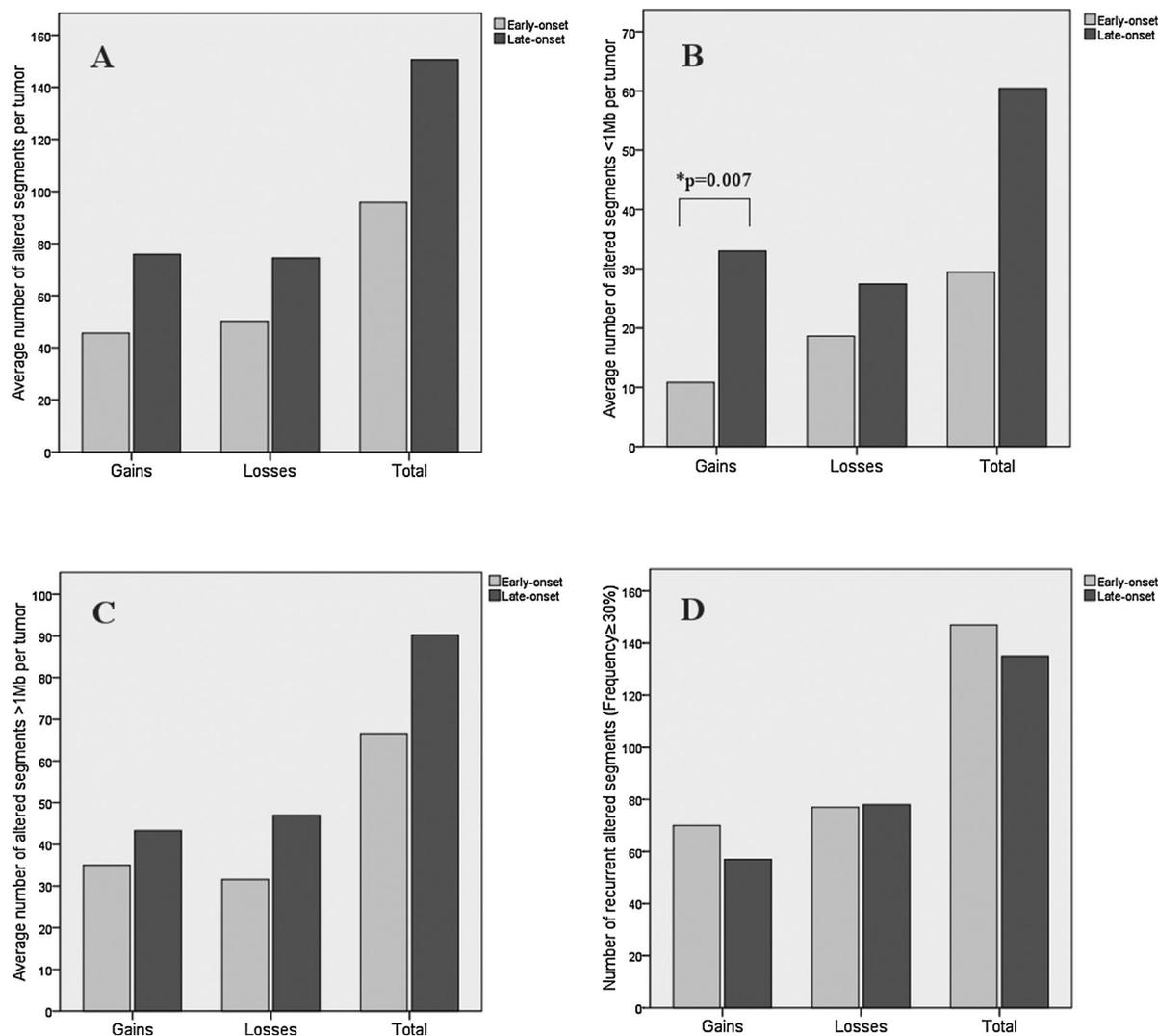


Figure 2. Average number of altered segments per tumor: (A) general; (B) smaller than 1 Mb; (C) larger than 1 Mb; and (D) copy number alterations with a frequency greater than or equal to 30% in both age groups.

composed of 55 cases, 41 of which corresponded to the early-onset subset. That is to say, 68% of early-onset tumors were included in cluster A. The second group (cluster B), contained 91 tumors and was enriched in late-onset cases, including 84% of the carcinomas from the elderly subset (Figure 5). This unsupervised clustering confirms that genomic profiles of CRC may be associated with age.

DISCUSSION

The clinicopathological features of sporadic early-onset CRC differ remarkably from those observed in late-onset CRC, and therefore it has been suggested that both types of CRC may have a different molecular basis [12].

We aimed to determine the genomic instability in early-onset and late-onset tumors, in order to define common and/or distinct events occurring in the initiation and evolution of these neoplasms. For this purpose, we used a high-resolution aCGH platform that allowed us to analyze the patterns of global genomic instability present in tumors from late-onset patients (>70 years old), and in tumors from the particularly poorly characterized early-onset subset (<45 years old). We also performed unsupervised hierarchical clustering to investigate a possible relationship between the profile of genomic changes and the age at onset of the disease.

We observed that the number of aberrations does not differ significantly between both groups of samples. In our series, the late-onset subset showed

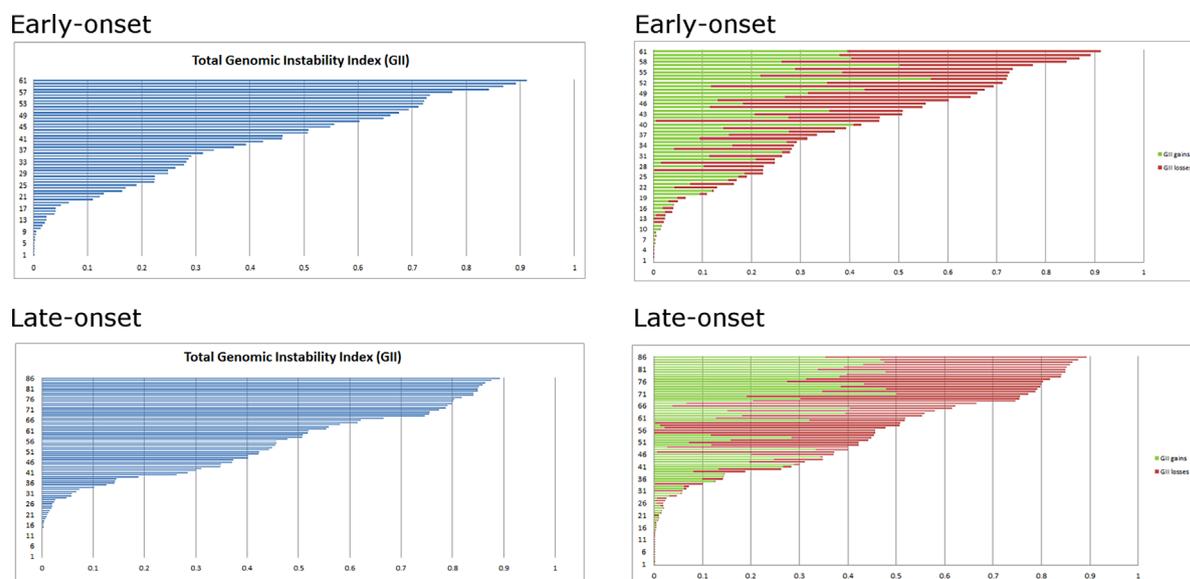


Figure 3. Genomic Instability index (GII) plots for each group of onset. On the left, the overall GII is plotted. On the right, gains (light color) and losses (dark color) are differentiated.

a higher average number of altered segments per tumor (150.62 vs. 95.80), although statistical significance was not reached. However, when we divided the altered segments into those bigger and those smaller than 1 Mb, we found that late-onset tumors had a significantly larger number of the smaller segments than early-onset tumors ($P = 0.007$).

We found differences regarding the CNA frequencies with respect to other published series [22–26], which may be explained by the types of specimen, the method used to assess CNAs, or the different inclusion criteria. On the one hand, we compared patients without known hereditary CRC syndromes whereas other studies were mainly performed in individuals from high-risk families. On the other hand, the present study included a larger number of patients with advanced stages of the disease as well as a larger rectal cancer cohort than others (34% of all tumors were located in the rectum). In our series, a number of altered segments displayed a frequency greater than or equal to 30% (147 in early-onset CRC and 135 in late-onset CRC). Curiously, the early-onset group showed more extreme values in the frequencies of these recurrent alterations, reaching percentages of 61%. By contrast, no alteration showed a frequency greater than 47% within the late-onset group. A possible explanation for this is that, although early-onset CRCs have a lower average number of altered segments per tumor, changes in these segments could be more liable to adjust to a particular pattern.

Another aspect that should be noted is the profile of gains or losses of whole chromosomes, because this could be at least in part nonrandom. Our results show that gains of chromosomes 7 and 20 and losses of

chromosome 18 were more frequent within the late-onset group, supporting the hypothesis that age may eventually determine the signature of genomic imbalances [27,28].

The recurrently gained or lost regions for each group of tumors suggest the involvement of genes whose gain or loss of function might be crucial for CRC development and/or progression. In line with this, the late-onset group showed a higher proportion of losses in 18q21.2, where *DCC* and *SMAD4* are located (36%), and in 5q22.2, where *APC* is located (30%); this could be pointing towards the possibility that older patients may be more predisposed to showing alterations in those genes described in the traditional genetic model for colorectal tumorigenesis [2].

In our series, a total of twenty-four chromosomal segments from six different chromosomes (Table 2) showed significant differences between early-onset and late-onset CRC. Among these altered regions, two were predominantly altered within the early-onset group and very uncommon within the late-onset group, which showed both gains and losses of genomic material. Interestingly, the chromosomal segments located at 1p35.2–p35.1 and 3p21.31–p21.1 harbor genes previously related with the development of malignancy and growth progression of certain tumors such as *PTP4A2*, *KHDRBS1*, *SMARCC1*, *CDC25A*, *IMPDH2*, *HYAL3*, *HYAL2*, *BAP1*, and *RPL29*. Moreover, we observed two other loci present in both groups, but with a higher frequency in the early-onset cases, that only showed loss of genetic material. They were located at 1p36.33 and 1q21.2, where *OR4F29* and

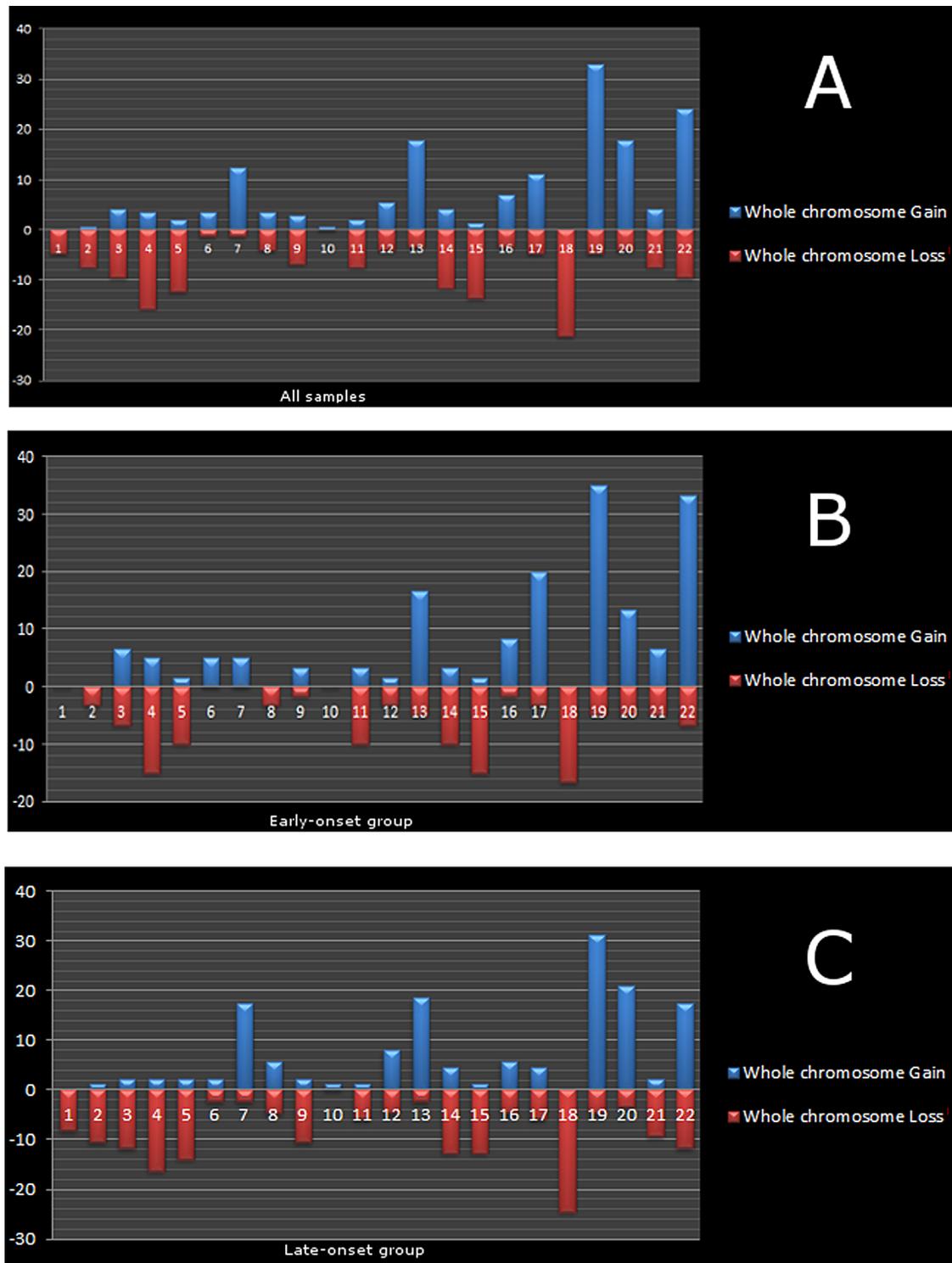


Figure 4. Gains (above zero) and losses (below zero) of whole chromosomes for the entire set of CRC and for early-onset and late-onset groups separately.

the well-known oncogene *BCL9* are included, respectively. Losses at 1q21.2 have been previously described as more frequent in young patients [15]. In consonance with that, 48% of early-onset and

32% of late-onset patients from our series displayed loss at this locus ($P = 0.0005$).

Most of our data were not in consonance with the fifteen regions previously suggested as being useful

Table 2. Summary of the Alterations Occurring With Statistically Significantly Different Frequencies in Both Groups

Cytoband	Region	Frequency early-onset ^a	Frequency late-onset ^a	P-value	FDR	No. of genes	Genes of interest ^b	Other genes in the region ^c
Gains								
1p35.2-p35.1	31075659-3366490	0.333	0.058	0.0001	0.052	47	PTP4A2, KHDRBS1	LAPTM5, PUM1, FABP3, SERINC2, PEF1, HCRTR1, BAI2, TXLNA, EIF31, LCK, BSDC1, HDAC1, RBBP4, TRIM62, ADC
1p35.2-p35.1	31181778-33667564	0.283	0.035	0.0003	0.062			ELP6, DHX30, CAMP, NME6, ATRIP, SHISA5, NCKIPSD, SLC25A20, IP6K2, KLHDC8B, TCTA, RHOA, RBM6, RBM5, GNAT1, SLC38A3, GNAI2, IFRD2, NAT6, HYAL1, TUSC2, RASSF1, NPRL2, CYB561D2, TMEM115, HYAL2, BAP1, CACNA2D2, ZMYND10, CISH, MANF, RBM15B, GNL3, RPL29
3p21.31-p21.1	46988350-53030975	0.350	0.093	0.0003	0.062	142	SMARCC1, CDC25A, IMPDH2, HYAL3, HYAL2, BAP1, RPL29	ELP6, DHX30, CAMP, NME6, ATRIP, SHISA5, NCKIPSD, SLC25A20, IP6K2, KLHDC8B, TCTA, RHOA, RBM6, RBM5, GNAT1, SLC38A3, GNAI2, IFRD2, NAT6, HYAL1, TUSC2, RASSF1, NPRL2, CYB561D2, TMEM115, HYAL2, BAP1, CACNA2D2, ZMYND10, CISH, MANF, RBM15B, GNL3, PBRM1, GRM2, ALAS1, TLR9, TWf2, PCBP4, PHF7, ACY1, DUSP7, ITIH3, UPK3B
7q11.23	76461459-76503690	0.133	0.430	0.0003	0.062	1		
7q22.1	102049904-102064180	0.150	0.442	0.0004	0.072	1		
7q11.22	72110621-72172516	0.217	0.512	0.0005	0.086	1		
Losses								
1p36.33	37196-782679	0.450	0.163	0.0002	0.062	3		
1p12-p11.2	120374173-143895512	0.517	0.198	0.0001	0.052	6		NBPF7, NOTCH2, FCGR1B
1p35.2-p35.1	31073820-33664190	0.317	0.058	0.0002	0.057	47	PTP4A2, KHDRBS1	LAPTM5, PUM1, FABP3, SERINC2, PEF1, HCRTR1, BAI2, TXLNA, EIF31, LCK, BSDC1, HDAC1, RBBP4, TRIM62, ADC
1q21.1	142613630-143895512	0.517	0.198	0.0001	0.052	1		
1q21.1	144500777-144849414	0.450	0.151	0.0001	0.054	2		NBPF9, PDE4DIP
1q21.1-q21.2	146107562-147962872	0.433	0.163	0.0005	0.086	11	BCL9	NBPF12, PRKAB2, ACP6, GJA5, GJA8, NBPF24, PPIAL4A
3p21.31-p21.21	47076908-53030975	0.383	0.105	0.0002	0.055	142	SMARCC1, CDC25A, IMPDH2, HYAL3, HYAL2, BAP1, RPL29	ELP6, DHX30, CAMP, NME6, ATRIP, SHISA5, NCKIPSD, SLC25A20, IP6K2, KLHDC8B, TCTA, RHOA, RBM6, RBM5, GNAT1, SLC38A3, GNAI2, IFRD2, NAT6, HYAL1, TUSC2, RASSF1, NPRL2, CYB561D2, TMEM115, HYAL2, BAP1, CACNA2D2, ZMYND10, CISH, MANF, RBM15B, GNL3, PBRM1, GRM2, ALAS1, TLR9, TWf2, PCBP4, PHF7, ACY1, DUSP7, ITIH3, SH2B2, ORAI2
7q22.1	101904918-102096491	0.367	0.105	0.0003	0.062	4		
7q11.21	64221857-64990939	0.350	0.093	0.0003	0.062	5		
7q11.21	73803410-73855403	0.383	0.116	0.0003	0.062	1		ZNF117
7q35	143528314-143704617	0.450	0.174	0.0004	0.083	4		
9p13.1-p11.2	38987240-45095186	0.433	0.151	0.0003	0.062	13		

continued

Table 2. (Continued)

Cytoband	Region	Frequency early-onset ^a	Frequency late-onset ^b	P-value	FDR	No. of genes	Genes of interest ^b	Other genes in the region ^c
9p13.1-q21.12	39054171-70202700	0.350	0.093	0.0003	0.062	13		FOXD4L2, FOXD4L5, FOXD4L6
14q11.1-q11.2	18407780-19124606	0.617	0.302	0.0002	0.062	0		
16p13.11	14778511-15336921	0.367	0.093	0.0002	0.055	6		
16p13.11	16245207-16276991	0.350	0.093	0.0003	0.062	1		
16p13.11	16427307-16435895	0.367	0.093	0.0002	0.055	0		
16p12.3	18038778-18594513	0.317	0.047	0.0001	0.052	2		

FDR: false discovery rate.

^aGlobal frequency of the alteration in each age subset.^bGenes previously reported as important in colorectal cancer.^cGenes previously reported as important in cancer.

to differentiate between both age populations [15]. This can be explained by the sample size, because in studies with a small sample size is more difficult to distinguish between a real effect and random variation. Unlike Berg et al. [15], we found a large part of chromosome 9 (9p13.1-q21.12) and three regions located in chromosomes 14 (14q11.1-q11.2) and 16 (16p13.11, 16p12.3) to be more frequently lost within the early-onset group. The chromosome 9 region encodes several members of the FOX family of transcription factors that play critical roles in the regulation of cell proliferation [29], as well as the *TJP2* gene, whose transcript has been suggested as a marker for CRC [30]. Also different from Berg et al. [15], we did not find statistically significant alterations at chromosomes 2, 4, 10, and 19, although gains at 19q13 (52% in early-onset vs. 42% in late-onset tumors) and losses at 10p11 (47% in early-onset vs. 34% in late-onset) were also present in our series.

We were able to define a small gained region more specifically related to late onset-cases, but in this case our findings were not coincident either with the regions previously published [15]. In our series, three sequence stretches within chromosome 7 located at 7q11.23, 7q22.1, and 7q11.22 were more frequently gained within the late-onset group. Among the genes located at 7q11.23 and 7q22.1 are *UPK3B* and *CUX1*, which have been described as being related to breast cancer and as having tumor suppressor activity, respectively.

Lastly, and also exemplifying the different genomic profiles linked to each group, the unsupervised analysis clearly discriminated between CNAs in early and late-onset cases, showing association between age at onset and specific patterns of copy number changes, which were the same as the differences shown between early and late-onset CRC CIN profiles and GII.

In summary, we performed a comprehensive analysis of the DNA-copy number changes that occur in early-onset and late-onset CRC. On the one hand, we found some quantitative differences between both subsets that did not reach statistical significance. However, we did find some important differences from a qualitative point of view, such as gains or losses of chromosomal regions the genes in which have been previously described as being implicated in colorectal carcinogenesis, mainly for the elderly subset. More interestingly, our findings included a larger amount of regions more frequently associated with the early-onset population; some of these regions have not been related with CRC to date. These results confirm the existence of substantial dissimilarities between the CIN pattern within early-onset CRC and other CRCs. Further studies are required in order to know the significance of these differential gains and losses of genetic material, even when they do not harbor known cancer genes.

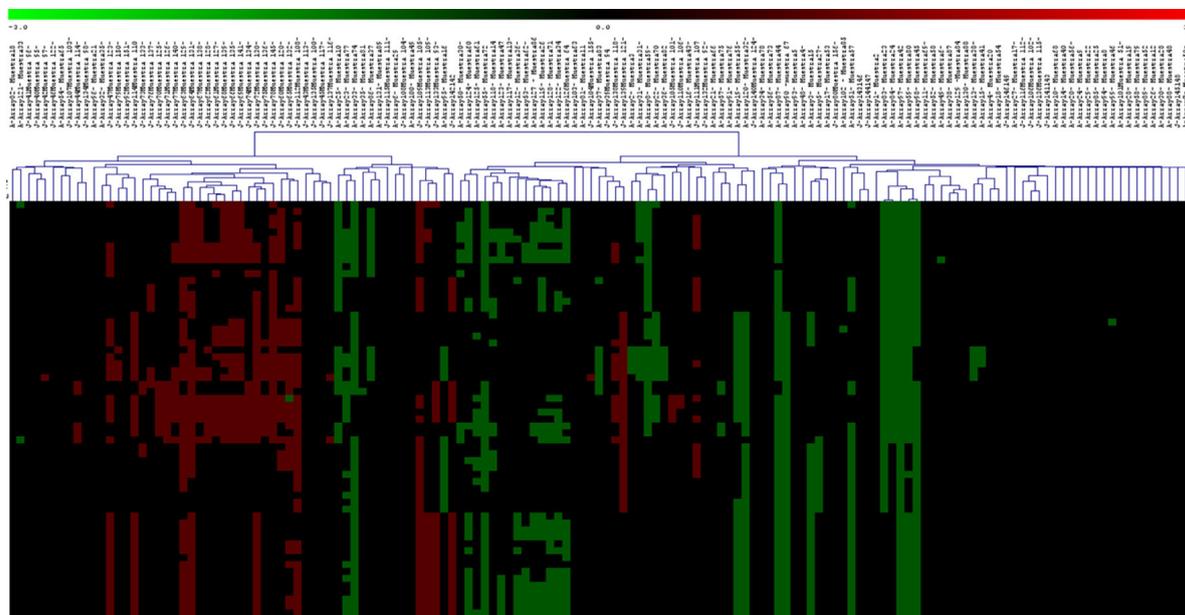


Figure 5. Unsupervised hierarchical clustering based on the genomic instability profile of all the studied tumors.

ACKNOWLEDGMENTS

This work was funded by Project PI10/0683 from the Spanish Ministry of Health and Consumer Affairs. We thank the Tumor Registry of the Pathology Department of the 12 de Octubre University Hospital for providing the paraffin-embedded tissues, and Ron Hartong for his help with the English revision of this article.

REFERENCES

1. <http://www.seom.org/seomcms/images/stories/recursos/sociosyprofs/documentacion/boletinseom/2010/66/09-repor-taje-bol66.pdf>.
2. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. *Br J Surg* 2002;89:845–860.
3. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998;396:643–649.
4. Perucho M. Cancer of the microsatellite mutator phenotype. *Biol Chem* 1996;377:675–684.
5. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 1993;363:558–561.
6. Dunican DS, McWilliam P, Tighe O, Parle-McDermott A, Croke DT. Gene expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA array hybridization. *Oncogene* 2002;21:3253–3257.
7. Goel A, Arnold CN, Niedzwiecki D, et al. Characterization of sporadic colon cancer by patterns of genomic instability. *Cancer Res* 2003;63:1608–1614.
8. Georgiades IB, Curtis LJ, Morris RM, Bird CC, Wyllie AH. Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence of chromosomal or microsatellite instability. *Oncogene* 1999;18:7933–7940.
9. Tang R, Changchien CR, Wu M-C, et al. Colorectal cancer without high microsatellite instability and chromosomal instability: An alternative genetic pathway to human colorectal cancer. *Carcinogenesis* 2004;25:841–846.
10. Banerjee A, Hands RE, Powar MP, Bustin SA, Dorudi S. Microsatellite and chromosomal stable colorectal cancers demonstrate poor immunogenicity and early disease recurrence. *Colorectal Dis* 2009;11:601–608.
11. Perea J, Rueda D, Canal A, et al. Age at onset should be a major criterion for subclassification of colorectal cancer. *J Mol Diagn* 2014;16:116–126.
12. Perea J, Alvaro E, Rodríguez Y, et al. Approach to early-onset colorectal cancer: Clinicopathological, familial, molecular and immunohistochemical characteristics. *World J Gastroenterol* 2010;16:3697–3703.
13. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. *Nat Genet* 2003;34:369–376.
14. Giráldez MD, López-Dóriga A, Bujanda L, et al. Susceptibility genetic variants associated with early-onset colorectal cancer. *Carcinogenesis* 2012;33:613–619.
15. Berg M, Agesen TH, Thiis-Evensen E, et al. Distinct high resolution genome profiles of early onset and late onset colorectal cancer integrated with gene expression data identify candidate susceptibility loci. *Mol Cancer* 2010;6:100.
16. Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004. *Ann Oncol* 2005;16:481–488.
17. Siegel RL, Jermal A, Ward EM. Increase in incidence of colorectal cancer among young men and women in the United States. *Cancer Epidemiol Biomark Prev* 2009;18:1695–1698.
18. Carro A, Rico D, Rueda OM, Díaz-Uriarte R, Pisano DG. WaviCGH: A web application for the analysis and visualization of genomic copy number alterations. *Nucleic Acids Res* 2010;38:W182–W187.
19. Bonnet F, Guedj M, Jones N, et al. An array CGH based genomic instability index (G2I) is predictive of clinical outcome in breast cancer and reveals a subset of tumors without lymph node involvement but with poor prognosis. *BMC Med Genomics* 2012;5:54.
20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Stat Soc B* 1995;57:289–300.

21. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
22. Chen Z, Liu Z, Li W, et al. Chromosomal copy number alterations are associated with tumor response to chemoradiation in locally advanced rectal cancer. *Genes Chromosomes Cancer* 2011;50:689–699.
23. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330–337.
24. Ried T, Knutzen R, Steinbeck R, et al. Genes Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Chromosomes Cancer* 1996;15:234–245.
25. De Angelis PM, Clausen OP, Schjølberg A, Stokke T. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999;80:526–535.
26. Mourra N, Zeitoun G, Buecher B, et al. High frequency of chromosome 14 deletion in early-onset colon cancer. *Dis Colon Rectum* 2007;50:1881–1886.
27. Ried T, Heselmeyer-Haddad K, Blegen H, Schröck E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: A phenotype/genotype correlation. *Genes Chromosomes Cancer* 1999;25:195–204.
28. Beroukhi R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899–905.
29. Freyaldenhoven BS, Fried C, Wielckens K. FOXD4A and FOXD4B, two new winged helix transcription factors, are expressed in human leukemia cell lines. *Gene* 2002;294:131–140.
30. Kim YJ, Jung YD, Kim TO, Kim HS. Alu-related transcript of TJP2 gene as a marker for colorectal cancer. *Gene* 2013;524:268–274.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

NOMO-1 gene is deleted in early-onset colorectal cancer

José Perea^{1,2}, Juan Luis García³, Jessica Pérez³, Daniel Rueda^{2,4}, María Arriba², Yolanda Rodríguez⁵, Miguel Urioste^{6,7}, Rogelio González-Sarmiento³

¹Surgery Department, University Hospital 12 de Octubre, Madrid, Spain

²Digestive Cancer Research Group, 12 de Octubre Research Institute, Madrid, Spain

³Department of Medicine, Molecular Medicine Unit, Biomedical Research Institute of Salamanca (IBSAL), Institute of Molecular and Cellular Biology of Cancer (IBMCC), University of Salamanca, SACYL, CSIC, Salamanca, Spain

⁴Molecular Biology Laboratory, University Hospital 12 de Octubre, Madrid, Spain

⁵Pathology Department, University Hospital 12 de Octubre, Madrid, Spain

⁶Familial Cancer Clinical Unit, Spanish National Cancer Centre (CNIO), Madrid, Spain

⁷Center for Biomedical Network Research on Rare Diseases (CIBERER), Institute of Health Carlos III, Madrid, Spain

Correspondence to: José Perea, **email:** josepereag@hotmail.com

Keywords: early-onset colorectal cancer, NOMO-1, nodal pathway, array comparative genomic hybridization, 16p13.12-p13.11

Received: December 09, 2016

Accepted: February 07, 2017

Published: February 18, 2017

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

ABSTRACT

To characterize clinical features of a recurrent alteration in 16p13.12-p13.11 in Colorectal Cancer (CRC), mainly in Early-onset subgroup (EOCRC), and to assess the status of NOMO1, a gene located in that region, we analyzed differential clinicopathological, familial and molecular features of CRC subsets with and without alterations in the 16p13.12-p13.11, in global and EOCRC groups. We confirmed the region by fluorescence in-situ hybridization, and Quantitative Real-Time PCR analyzed the status of NOMO1 in different age-of-onset and Microsatellite Instability (MSI)-status CRC subsets. Both age-of-onset subsets were subsequently extended to further confirm NOMO1 gene changes. 16p13.12-p13.11 alterations were observed in 23.3% of CRCs, and was detected more frequently in EOCRC (33.3%) than in late-onset CRC (16.3%). The group with deletion in 16p showed a higher frequency of females and left-colon locations; a better prognosis; and higher Chromosomal Instability. Within the primary EOCRC population, 34 out of 34 of tumours showed a homozygous deletion in NOMO1, while in the late-onset population only 2 of the 17 tumours (11.7%) showed it. In the extended group, we found 61 out of 75 EOCRC patients (81.3%) with homozygous deletion and 7 patients (9.3%) with heterozygous deletion of NOMO1; moreover, in the new 50 late-onset patients, the proportions of deletions decreased. Microsatellite-Stable (MSS) EOCRC showed a very high proportion of homozygous loss of NOMO1 (54 of 59 cases, 91.5%), while the deletion was observed in only 7 out of 16 MSI cases. Deletion of NOMO1 is a molecular marker predominantly associated with EOCRC, particularly MSS subtypes.

INTRODUCTION

Early-onset Colorectal Cancer (EOCRC) has an incidence of 2–8% of all Colorectal Cancers (CRCs) and it has increased in the past decades to reach 11% of colon cancers and 18% of rectal cancers [1]. The impact of EOCRC on the population is undeniably important, and until recently the idea prevailed that this subset of CRC occurred mainly in hereditary CRC forms. However, recent studies challenge this statement, since they are

primarily Microsatellite Stable (MSS) cases [2–6]. Moreover, EOCRC (except cases with an already known hereditary component) may be a specific subgroup of CRC [3–5], so that a deeper understanding of the underlying molecular mechanisms is essential.

EOCRC has evolved from a controversy on its natural history and prognosis to the characterization of an important heterogeneity within this group [6]. Moreover, it has been proposed that age of onset is a major criterion for subclassifying CRC [3]. Most studies conclude that there

are differential features within this age group not only from a clinical point of view but also, and more significantly, according to a number of differential molecular features: a high degree of LINE-1 hypomethylation [7], more frequent chromosomal and genetic alterations (including some susceptibility variants), underlying an inherited or familial predisposition [8–10], and unique characteristics, including clinical and molecular features and the type of telomere maintenance mechanism [4, 5]. In spite of all these differential molecular features, to date there is no known molecular or genetic target associated with EOCRC.

CRC results from the accumulation of genetic alterations, and somatic copy number alterations (CNAs) play an important role in its development. Genome-wide survey of CNAs provides opportunities for identifying cancer driver genes in an unbiased manner [11]. In an attempt to better characterize EOCRC, we have previously reported results from an array-Comparative Genomic Hybridization (aCGH) study in which we compared early and late-onset CRC [10]. Analysis of the data have revealed a recurrent deletion in chromosome 16p13.12-p13.11, either alone or associated with other changes (Supplementary Figure 1). The aims of our study were to define the possible clinical phenotype of the cases showing this chromosomal alteration as well as to find out genes in this region that could be altered.

RESULTS

16p deletion is more frequent in EOCRC than in late-onset CRC

Sixty cases from the EOCRC subset and 86 from the late-onset population were studied by aCGH (the others could not be studied because of lack of material); analysis of the data revealed a recurrent focal deletion in chromosome 16p13.12-p13.11, either as the unique change or associated with other changes (Supplementary Figure 1). This focal alteration was observed in 23.3% (34/146) cases in our series, and was detected more frequently in early-onset CRC (33.3%, 20/60) than in late-onset CRC (16.3%, 14/86) ($p = 0.028$) (Figure 1).

In Table 1 we compare clinico-pathological, familial, molecular and follow-up features of the groups according to 16p status as determined by aCGH. In the global population (without defining any age-of onset criteria), the 16p deletion appeared to be more frequent in females (58.8% vs 39.3%). Moreover, in the group with 16p deletion we observed a very low rate of rectal (17.6% vs 42%) and a high rate of left-colon locations (47.1% vs 27.7%); a better prognosis, and a higher Chromosomal Instability. Kaplan-Meier curves are shown in Figure 2, confirming the better prognosis of the group with 16p13.12-p13.11 deletion, mainly for OS. When we analysed the EOCRC group separately, only tumor

location and Chromosomal Instability (number of CNAs) remained statistically significantly different between the patients with and without 16p deletion (Table 1).

16p13.12-p13.11 deletion causes loss of *NOMO1* gene

Analysis of 16p13.12-p13.11 region by Fluorescence in-situ hybridization analysis (FISH) showed a minimal common region between bases 14738223 and 15353060 with a size of 614838 bp (hg19) (<http://genome.ucsc.edu>). This region contains the genes *BFAR* (bifunctional apoptosis regulator), *PLA2G10* (phospholipase A2), *NPIPA2* (nuclear pore complex interacting protein) *ABCC6P2* (ATP-binding cassette, sub-family C, member 6 pseudogene 2), *NOMO1* (NODAL modulator 1), *PDXDC1* (pyridoxal-dependent decarboxylase domain containing 1), *NTANI* (N-terminal asparagine amidase), *RRN3* (RRN3 RNA polymerase I transcription factor homolog), and *PKDIP6* (polycystic kidney disease 1 (autosomal dominant) pseudogene 6).

We analysed the status of the *NOMO1* gene in both age-at-onset subgroups. Within EOCRC, all 20 cases with a cytogenetic 16p13.12-p13.11 deletion showed homozygous deletion of *NOMO1*. Unexpectedly, we also observed 14 cases without cytogenetic evidence of a 16p13.12-p13.11 deletion that presented also homozygous deletion of *NOMO1* (Figure 1). Thus, within the EOCRC population, 34 of 34 studied tumors showed a homozygous deletion in *NOMO1* (100%). Lack of material prevented us to further analyse the remaining cases.

Within late-onset CRC, we were able to study 9 of the 14 cases showing cytogenetic 16p13.12-p13.11 deletion: only two of these cases showed homozygous loss of *NOMO1*, five showed a heterozygous deletion and the other two showed *NOMO1* wild-type (Figure 1). We also studied 8 late-onset CRC without 16p deletion, and all of them had *NOMO1* wild-type (Figure 1). In summary, only 2 out of 17 late-onset CRCs (11.7%) showed a homozygous deletion in *NOMO1*.

To further investigate the high rate of *NOMO1* loss observed in the preliminary series of EOCRC (34 out of 34), we increased the number to 75 early-onset cases as mentioned in the Methods section. Results are shown in Table 2. Of the 41 new cases, 27 showed *NOMO1* homozygous deletion, 7 showed heterozygous deletion, and 7 showed *NOMO1* wild-type. Taken together, from a total of 75 EOCRC cases we found 61 patients (81.3%) with a homozygous loss; 7 patients (9.3%) with a heterozygous loss; and 7 patients (9.3%) without loss of *NOMO1*. Of the other 50 late-onset CRC added, only one showed *NOMO1* homozygous deletion, 4 showed heterozygous deletion, and 45 showed *NOMO1* wild-type. Taken together, from a total of 67 late-onset cases, we found 3 patients (4.5%) with a homozygous loss; 9 patients (13.4%) with a heterozygous loss; and the rest without loss of *NOMO1*.

We divided our EOCRC cases according to Microsatellite Instability (MSI) phenotype. MSS-EOCRC showed a very high proportion of homozygous *NOMO1* deletion (54 of 59, 91.5%), while it was present only in 7 of 16 MSI-EOCRC cases (43.74%). Heterozygous deletion of *NOMO1* appears to be rare in MSS-EOCRC (3.3%, 2/59) and relatively common in MSI-EOCRC cases (31.2%, 5/16) (Differences between MSI groups according to the *NOMO1* status in the Global group: $p < 0.001$). Fourteen out of the 16 MSI cases showed germline mutations in the Mismatch Repair (MMR) genes, so that they were defined as Lynch syndrome cases. Only one of the others showed *MLH1* gene promoter hypermethylation. These last two, without MMR germline mutation, exhibited homozygous *NOMO1* deletion, showing the Lynch syndrome cases all the three possible *NOMO1* mutation status.

NOMO1 deletion is somatic

We analysed *NOMO1* in DNA obtained from peripheral blood samples of 13 individuals with EOCRC and homozygous deletion in their tumours. None of these cases showed *NOMO1* deletion in their peripheral blood, which strongly suggests that *NOMO1* deletion is somatic.

DISCUSSION

Deletions of 16p13 observed by aCGH have been associated with multiple congenital anomalies [12]. In cancer, these changes were previously reported in Perivascular Epithelioid Cell tumors (PEComa) and prostate cancer [13, 14]. A small deletion in chromosome 16p13.2 affecting the alternative splicing factor *RBFOX1* was found at a significantly higher rate in the CRC British Bangladeshi patients (50%) than in CRC British Caucasians patients (15%), with the Bangladeshi CRC patients being considerably younger [15]. However, this gene is located at a more telomeric position than the deletion we have pointed out, and we therefore decided to further study the involvement of the 16p deletion in our series of colorectal tumors. Interestingly, the 16p13.12-p13.11 region we observed was significantly more frequently deleted in EOCRC than in late-onset CRC and showed some important clinical and prognostic implications. Apart from a better prognosis, we observed the almost complete absence of rectal tumors within the EOCRC subset with 16p deletion. The predominance of left-colon cancer and a high chromosomal instability in these cases are also remarkable. Different nearby regions

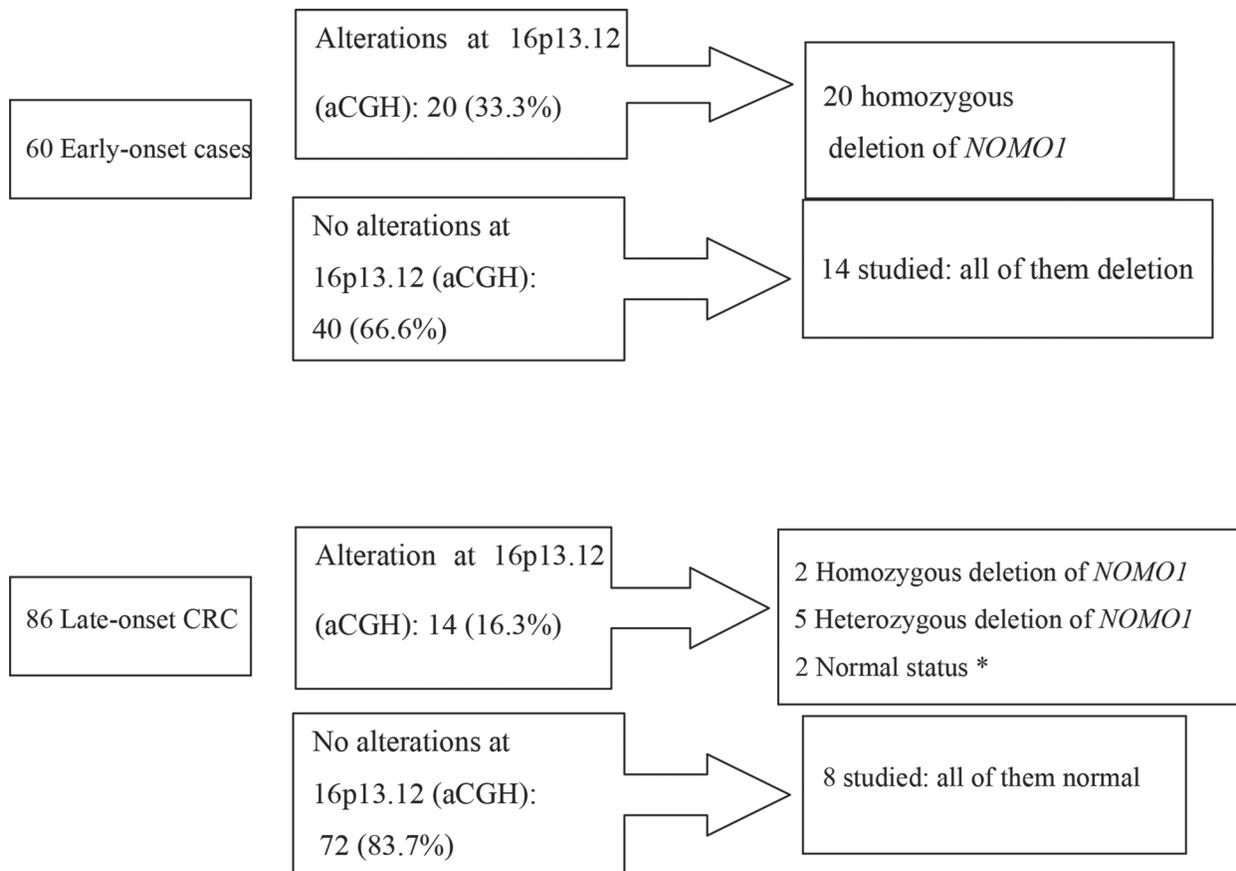


Figure 1: Losses at 16p13.12-p13.11 and *NOMO-1* alterations within the early-onset CRC and late-onset CRC subgroups. *Only 9 cases could be analyzed, because of lack of material.

Table 1: Clinical, pathological and familial features of all CRC cases (global group), and comparison of the subgroups with and without 16p alterations within the global and the early-onset groups

	16p altered Global n (%)	16p normal Global n (%)	<i>p</i> (χ^2)	16p altered Early-onset n (%)	16p normal Early-onset n (%)	<i>p</i> (χ^2)
Patients	34 (23.3)	112 (76.7)		20 (33.3)	40 (66.7)	
Mean age of onset (SD) ¹				38.9 (5.6)	39.4 (4.9)	NS
Sex:						
Male	14 (41.2)	68 (60.7)	0.05	9 (45)	27 (67.5)	NS
Female	20 (58.8)	44 (39.3)		11 (55)	13 (32.5)	
Location:						
Right colon	12 (35.3)	34 (30.4)	0.024	7 (35)	6 (15)	0.001
Left colon	16 (47.1)	31 (27.7)		13 (65)	15 (37.5)	
Rectum	6 (17.6)	47 (42)		0 (0)	19 (47.5)	
Tumor differentiation ² :						
Poor	2/30 (6.7)	5/99 (5.1)	NS	2/16 (12.5)	1/33 (3)	NS
Mucin production ² .	6/30 (20)	23/99 (23.2)	NS	5/16 (31.3)	9/33 (27.3)	NS
“Signet ring” cells ² .	0 (0)	4/99 (4)	NS	0/16 (0)	2/33 (6.1)	NS
Modified Astler Collier stage:						
A	6 (17.6)	13 (11.6)	NS	4 (20)	11 (27.5)	NS
B	18 (52.9)	48 (42.9)		11 (55)	12 (30)	
C	4 (11.8)	29 (25.9)		1 (5)	9 (22.5)	
D	6 (17.6)	22 (19.6)		4 (20)	8 (20)	
Associated polyps	25 (73.5)	66 (58.9)	NS	13 (65)	22 (55)	NS
Mean number of polyps (SD) ⁴	2.3 (2.8)	2.8 (6.3)	NS	2.2 (2.6)	2.8 (8)	NS
Type:						
Adenomatous	11 (44)	39 (59)	NS	5 (38.5)	9 (40.9)	NS
Hyperplastic	3 (22)	9 (14)		2 (15.4)	5 (22.7)	
Mixed	11 (44)	18 (27)		6 (46.1)	8 (36.4)	
Synchronous or metachronous CRCs.	4 (11.8)	19 (17)	NS	3 (15)	2 (5)	NS
Recurrence ³	2 (7.7)	15 (16.7)	NS	2 (12.5)	5 (15.6)	NS
Related mortality	6 (17.6)	36 (32.1)	NS	3 (15)	11 (27.5)	NS
Disease-free survival (SD) ¹	55.8 (46.4)	32.9 (31.9)	0.02	65.4 (47)	51.7 (36.5)	0.2
Overall survival (SD) ¹	61.3 (43.7)	38.7 (30.3)	0.15	72 (42.8)	61.1 (29.8)	0.3
MSI	4 (11.8)	12 (10.7)	NS	3 (15)	6 (15)	NS
MMR genes mutations	3 (8.8)	5 (4.5)	NS	3 (15)	4 (10)	NS
CIMP-High	9 (26.5)	22 (19.6)	NS	6 (30)	6 (15)	0.1
CNA (SD) ¹	175 (90.5)	110 (75.5)	0.01	160 (128.5)	63.7 (67.3)	0.004
Familial history of cancer			NS			NS
Amsterdam II Positive families.	4 (11.8)	7 (6.3)	NS	4 (20)	7 (17.5)	NS
Aggregation for Lynch neoplasm.	13 (38.2)	27 (24.1)		10 (50)	19 (47.5)	
Aggregation for Lynch unrelated neoplasm.	8 (23.5)	22 (19.6)		8 (40)	14 (36)	
Sporadic cases.	17 (50)	75 (67)		32 (46.4)	17 (42.5)	

¹ Statistical analysis was carried out using Student’s *t* test. ² Percentages shown are based on varying total numbers as some cases were excluded because only one biopsy was taken (stage D), or because tumours were severely dysplastic with “*in situ*” carcinoma, and it was not possible to study any other characteristic. ³ Cases showing recurrence are those with stage C or less at diagnosis. SD: Standard Deviation. NS: Not significant. CRC: Colorectal Cancer CIMP: CpG Island Methylator Phenotype. MSI: Microsatellite instability. MMR: Mismatch Repair system. CNA: Copy Number Alterations per case.

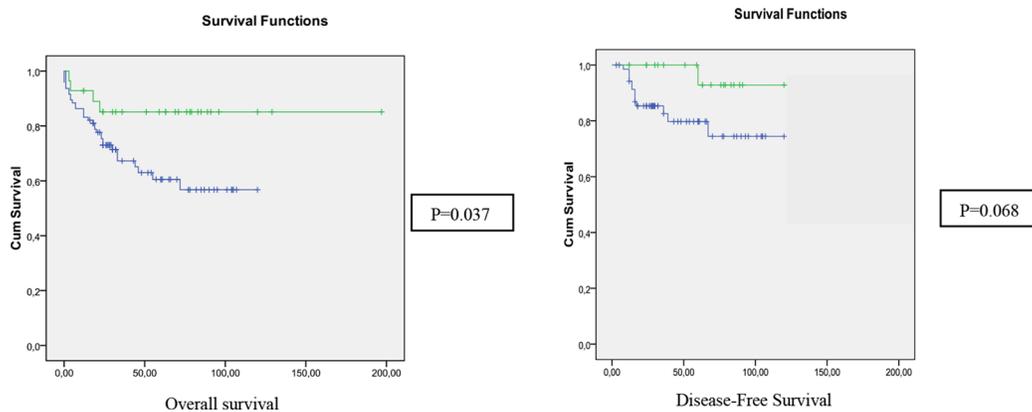


Figure 2: Kaplan-meier curves for overall survival and disease-free survival of the global subset, according to the status of 16p13.12 in aCGH (Green: altered; Blue: normal).

Table 2: Alterations in *NOMO1* in early-onset CRC in the primary group, the subset extension group, and the global CRC group

	Primary group		Subset extension group		GLOBAL	
	MSS <i>n</i> (%)	MSI <i>n</i> (%)	MSS <i>n</i> (%)	MSI <i>n</i> (%)	MSS <i>n</i> (%)	MSI <i>n</i> (%)
<i>NOMO1</i> Homozygous deletion	31 (100)	3 (100)	23 (82)	4 (31)	54 (92)	7 (44)
<i>NOMO1</i> Heterozygous deletion	0	0	2 (7)	5 (38)	2 (3)	5 (31)
<i>NOMO1</i> Normal	0	0	3 (11)	4 (31)	3 (5)	4 (25)

Primary group: Preliminary series of early-onset CRC, from the aCGH study. Subset extension group: Expansion of the series with Early-onset CRC patients from three different institutions, in order to confirm the preliminary results. GLOBAL: The sum of the two previous. Differences between MSI groups according to the *NOMO1* status in the Global group: $p < 0.001$. MSI: Microsatellite Instability. MSS: Microsatellite Stability. Ratios are calculated for subgroups of MSI-status.

have been related with prognostic aspects. For example, losses at the 16p13.3 region have been described to lead to poor prognosis in CRC; this region bears a total of 167 known genes among which the tumor suppressor gene *AXIN1* [11, 16]. The apparent contradiction between better prognosis and high chromosomal instability of cases with 16p deletion observed in the present study should be explained by the greater proportion of Microsatellite And Chromosomal Stable tumours within EOCRC, which are the ones showing a worse prognosis in this particular subset of CRC.

One of the genes located in 16p13.12-p13.11 is *NOMO1*. Unexpectedly, we found *NOMO1* deletion not only in cases with cytogenetic deletion in 16p, mainly in the EOCRC, but also in most early-onset cases without 16p deletion (34 out of 34 EOCRC cases studied by aCGH and FISH), while it barely reached 12% in the late-onset CRC subset. The identification of deletions of *NOMO1* gene detected by qRT-PCR highlight the role of this gene in colorectal carcinogenesis, especially in early-onset tumors. However, to date there is no evidence that *NOMO1* is related with carcinogenesis. Its main function currently known is to form part of a protein complex that antagonizes Nodal signaling, a pathway essential for patterning of the early embryo during mesoderm and endoderm induction as well as for the specification of left-right asymmetry [17]. According to the Cancer Genome Atlas data alterations involving 16p13 appears to have a limited role in CRC that is confirmed in our analysis of other tumours [18].

When we extended the analysis, we confirmed the important proportion of EOCRC cases showing *NOMO1* loss, and preliminary findings also indicate that this loss is specific of CRC and, more significantly, that it appears to be an important clinical marker of MSS EOCRC as we observed the deletion in more than 90% of the cases. Its possible carcinogenetic role remains uncertain but there is recent evidence supporting the possibility that *NOMO1*

could act as a tumor suppressor gene: Nodal pathway activity is upregulated in human cancers such as malignant melanoma [19]; the upregulation of Cripto-1, a protein involved downstream of Nodal signaling is observed in many epithelial cancers like CRC [20], and Cripto-1 overexpression promotes tumorigenesis in xenografts and transgenic mice [21]. Moreover, collectively, Nodal signaling pathway promotes the self-renewal of human Colon Cancer Stem cells and mediates carcinogenesis of human CRC in an autocrine manner through Smad2/3 pathway [22]. As *Nomo1* antagonizes the Nodal signaling pathway [23], the deletion of *NOMO1* could lead in a downregulation of the protein and consequently, an upregulation of the Nodal signaling pathway.

In summary, apart from the clinical value of loss of 16p13.12-p13.11, we identify loss of *NOMO1* as a molecular marker mainly associated with EOCRC, and particularly with MSS subtypes. Our findings may serve as a starting point for further studies to confirm the potential carcinogenetic value of this deletion, which would place *NOMO1* in a suitable position as a potential therapeutic target for EOCRC treatment.

MATERIALS AND METHODS

Families, samples and data collection

A total of 82 consecutive individuals with CRC diagnosed at an age of 45 years or younger were collected from our institution. We also collected 97 consecutive individuals who were diagnosed during the same period with CRC but at an age of 70 years or older, to compare with the EOCRC group. These groups have been described previously [4, 10]. We collected clinicopathological data and analyzed MSI status, the mutational state of MMR genes, and the CpG methylation phenotype of all cases. Follow-up was at least 5 years from surgery, and Disease-Free Survival (DFS) and Overall Survival (OS),

recurrence and cancer-related death for each case were determined. Details of these studies have been previously reported [4, 10].

In order to confirm our findings, we expanded our series with EOCRC patients from three different institutions (University Hospital of Salamanca, Familial Cancer Clinical Unit of the Spanish National Cancer Research Centre, and 12 de Octubre University Hospital) obtaining 41 additional cases. The main clinical features of these patients were also recorded. We collected as well another 50 late-onset CRC from University Hospital of Salamanca, all of them showing MSS.

Chromosomal instability: array comparative genomic hybridization (aCGH)

Sixty early-onset CRC cases and 86 late-onset CRCs were studied by array-CGH using oligonucleotide microarrays (Roche NimbleGen, Inc., Reykjavik, Iceland) in order to identify CNAs as previously reported (10). Each genomic region exhibiting a copy number change was examined using the UCSC genome browser (<http://genome.ucsc.edu/>) to determine the location and significance of the change.

Fluorescence in-situ hybridization analysis (FISH)

To confirm the gains and losses detected by array-CGH, FISH analysis was performed using BAC clones 354N7 mapped to 16q22.1 (bases 68,727,161-68,887,391) and CTD 2504F3 mapped to 16p13.1 (bases 15,982,491-16,190,907), as previously described (NCBI16/hg18) [24]. These clones were selected from the same BAC clone library used for the BAC-array studies (Wellcome Trust Sanger Institute, Cambridge, UK). DNA from the BAC clones was isolated and directly labelled with either Spectrum Green-dUTP or Spectrum Orange-dUTP (Vysis, Downers Grove, IL), by nick translation and hybridized as previously described [24]. All BAC clones were first hybridized to normal human metaphase chromosomes in order to verify their location.

Paraffin-embedded tissue sections (4- μ m thick) were deparaffinized, dehydrated and air-dried. The slides were placed in 2 mM EDTA (pH 9) for 15 minutes. After cooling, the slides were transferred to a Coplin jar containing 40 ml of 0.9% NaCl (pH 1.5) and 160 mg of pepsin (Sigma) preheated to 37°C. Following incubation for 15 minutes, the sections were dehydrated in an alcohol series and fixed in 3:1 methanol:acetic acid for 10 minutes. The slides were then incubated in a Hybrite hybridization chamber (Vysis) for 12 minutes at 72°C, followed by 15 to 20 hours at 37°C. After hybridization, the slides were washed in standard solutions of saline citrate. Nuclei were counterstained with DAPI (Vector Laboratories Inc.). The images were captured with an Olympus BX60 epifluorescence microscope coupled to a CCD camera and

evaluated with Cytovision software (Applied Imaging). Approximately 400 non-overlapping tumor cells were evaluated (Supplementary Figure 2).

Quantitative real-time PCR

For real-time quantification of target gene expression, one-step real-time Polymerase Chain Reaction (RT-PCR) was performed using FastStart Universal SYBR Green Master (ROX) in a StepOnePlus™ Real-Time PCR System (Life Technologies-Invitrogen, California, U.S.A.). A fragment of the *NOMO1* gene was amplified from the DNA of patients and controls using the following primers: F: 5'-agctccatgtggatggagtc-3' and R: 5'-acggatgaagtacagatcc-3'. As internal control, the 36b4 gene was amplified from the same DNA using the primers: F: 5'-cagcaagtgggaaggtgtaacc-3' and R: 5'-cccattctatcatcaacgggtacaa-3'.

Ten μ l RT-PCR of a mix containing 15 ng of total DNA, 1 μ l of the primer dilution, 4 μ l FastStart Universal SYBR Green Master and 4 μ l H₂O were used for amplification. One-step RT-PCR reactions were carried out in 96-well optical reaction plates, covered with MicroAmp® Optical Adhesive Film (Life Technologies-Invitrogen, California, U.S.A.). Cycling was as follows: 10 minutes at 95°C followed by 40 cycles of 95°C for 15 seconds, 58°C for 45 seconds and 72°C for 15 seconds. RQ Manager software was used to analyse the values.

The comparative Ct method ($2^{-\Delta\Delta C_t}$) was used to calculate the relative expression levels of each amplicon. RT-PCR specificity of each PCR reaction was verified by melting curve analysis.

Statistical analyses

Continuous variables were expressed as mean values plus/minus standard deviation (SD), and categorical variables were expressed as number of cases and their percentage. Differences were considered significant when $p < 0.05$. For associations between discrete variables, statistical analyses were performed using Pearson's Chi Square (χ^2) Test for parametric variables, and Fisher's Exact Test for non-parametric variables. For continuous variables, Student's *t* test was used. The SPSS v.11.5 for Windows (SPSS, Inc., Chicago, IL) statistical package was used. The Kaplan–Meier method was used to describe the distribution of survival time, and the log-rank test was applied. For analyses of colorectal cancer-specific mortality, death as a result of CRC was the primary end point, and deaths from other causes were censored. The same analysis was carried out for disease-free survival time, using recurrence as the primary endpoint. For aCGH analysis, statistics were as published before [10].

Abbreviations

aCGH: array-Comparative Genomic Hybridization.
CNAs: Copy Number Alterations. CRC: Colorectal

Cancer. DFS: Disease-Free Survival. EOCRC: Early-onset Colorectal Cancer. FISH: Fluorescence in-situ hybridization. MMR genes: Mismatch Repair genes. MSI: Microsatellite Instability. MSS: Microsatellite Stability-Microsatellite Stable. PCR: Polymerase Chain Reaction. SD: Standard Deviation.

ACKNOWLEDGMENTS

We thank the Tumor Registry of the Pathology Department of the 12 de Octubre University Hospital, the Pathology Department of the University Hospital of Salamanca, and the Tumor Bank of the Spanish National Cancer Research Centre, for providing us with paraffin-embedded tissues, and Ron Hartong for his help with the English revision of this article.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

GRANT SUPPORT

This work was funded by Projects PI10/0683, PI13/01741, PI13/0127, PI16/01920 and PI16/01650, from the Spanish Ministry of Health and Consumer Affairs and FEDER, and was approved by the Ethics Committee of our Institution.

Authors' contributions

Conception and design: José Perea, Juan L. García, Miguel Urioste, Rogelio González-Sarmiento. Provision of study materials or patients: José Perea, Jessica Pérez, Daniel Rueda, María Arriba, Yolanda Rodríguez, Miguel Urioste, Rogelio González-Sarmiento. Collection and assembly of data: José Perea, Juan L. García, Jessica Pérez, Rogelio González-Sarmiento. Data analysis and interpretation: José Perea, Juan L. García, Rogelio González-Sarmiento. Manuscript writing: José Perea, Juan L. García, Miguel Urioste, Rogelio González-Sarmiento. Final approval of manuscript: All authors.

REFERENCES

1. Ahnen DJ, Wade SW, Jones WF, Sifri R, Mendoza-Silveiras J, Greenamyre J, Guiffre S, Axilbund J, Spiegel A, You YN. The increasing incidence of young-onset colorectal cancer: A call to action. *Mayo Clinic Proceedings*. 2014.
2. Perea J, Álvaro E, Rodríguez Y, Rivera B, Sánchez E, Carbonell P, Mercadillo F, Grávalos C, González-Sarmiento R, Hidalgo M, Urioste M. Approach to the early-onset colorectal cancer: Clinicopathological, familial, molecular and immunohistochemical characteristics. *World J Gastroenterol*. 2010; 16:3697–3703.
3. Perea J, Rueda D, Canal A, Rodríguez Y, Álvaro E, Osorio I, Alegre C, Rivera B, Martínez J, Benítez J, Urioste M. Age at onset should be a major criterion for subclassification Colorectal Cancer. *J Mol Diagn*. 2014; 16:116–26.
4. Boardman LA, Johnson RA, Viker KB, Hafner KA, Jenkins RB, Riegert-Johnson DL, Smyrk TC, Litzelman K, Seo S, Gangnon RE, Engelman CD, Rider DN, Vanderboom RJ, et al. Correlation of chromosomal instability, telomere length and telomere maintenance in Microsatellite stable rectal cancer: A molecular subclass of rectal cancer. *PLoS ONE*. 8:e80015. doi: 10.1371/journal.pone.0080015.
5. Kirzin S, Marisa L, Guimbaud R, De Reynies A, Legrain M, Laurent-Puig P, Cordelier P, Pradère B, Bonnet D, Meggetto F, Portier G, Brousset P, Selves J, et al. Sporadic Early-onset colorectal cancer is a specific subtype of cancer: a morphological, molecular and genetics study. *PLoS ONE*. 9:e103159. doi:10.1371/journal.pone.0103159.
6. Losi L, Di Gregorio C, Pedroni M, Ponti G, Roncucci L, Scarselli A, Genuardi M, Baglioni S, Marino M, Rossi G, Benatti P, Maffei S, Menigatti M, et al. Molecular genetic alterations and clinical features in early-onset colorectal carcinomas and their role for the recognition of hereditary cancer syndromes. *Am J Gastroenterol*. 2005; 100:2280–2287.
7. Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, Cuatrecasas M, Bujanda L, Giraldez MD, Takahashi M, Cabanne A, Barugel ME, Arnold M, et al. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PLoS One*. 2012; 7:e45357. doi: 10.1371/journal.pone.0045357.
8. Giráldez MD, López-Dóriga A, Bujanda L, Abulí A, Bessa X, Fernández-Rozadilla C, Muñoz J, Cuatrecasas M, Jover R, Xicola RM, Llor X, Piqué JM, Carracedo A, et al. Susceptibility genetic variants associated with early-onset colorectal cancer. *Carcinogenesis*. 2012; 33:613–19.
9. Giráldez MD, Balaguer F, Bujanda L, Cuatrecasas M, Muñoz J, Alonso-Espinaco V, Larzabal M, Petit A, Gonzalo V, Ocaña T, Moreira L, Enríquez-Navascués JM, Boland CR, et al. MSH6 and MUTYH deficiency is a frequent event in early-onset colorectal cancer. *Clin Cancer Res*. 2010; 16:5402–13.
10. Arriba M, García JL, Inglada-Pérez L, Rueda D, Osorio I, Rodríguez Y, Álvaro E, Sánchez R, Fernández T, Pérez J, Hernández JM, Benítez J, González-Sarmiento R, et al. DNA copy number profiling reveals different patterns of chromosomal instability within colorectal cancer according to the age of onset. *Mol Carcinog*. 2016; 55:705–16.
11. Wang H, Liang L, Fang JY, Xu J. Somatic gene copy number alterations in colorectal cancer: new quest for cancer drivers and biomarkers. *Oncogene*. 2016; 35:2011–9.
12. Nagamani SC, Erez A, Bader P, Lalani SR, Scott DA, Scaglia F, Plon SE, Tsai CH, Reimschisel T, Roeder E, Malphrus AD, Eng PA, Hixson PM, et al. Phenotypic manifestations of copy number variation in chromosome 16p13.11. *Eur J Hum Genet*. 2011; 19:280–6.

13. Pan CC1, Chung MY, Ng KF, Liu CY, Wang JS, Chai CY, Huang SH, Chen PC, Ho DM. Constant allelic alteration on chromosome 16p (TSC2 gene) in perivascular epithelioid cell tumour (PEComa): genetic evidence for the relationship of PEComa with angiomyolipoma. *J Pathol.* 2008; 214:387–393.
14. Choucair KA, Guérard KP, Ejdelman J, Chevalier S, Yoshimoto M, Scarlata E, Fazli L, Sircar K, Squire JA, Brimo F, Cunha IW, Aprikian A, Gleave M, et al. The 16p13.3 (PDPK1) Genomic Gain in Prostate Cancer: A Potential Role in Disease Progression. *Transl Oncol.* 2012; 5:453–60.
15. Sengupta N, Yau C, Sakthianandeswaren A, Mouradov D, Gibbs P, Suraweera N, Cazier JB, Polanco-Echeverry G, Ghosh A, Thaha M, Ahmed S, Feakins R, Propper D, et al. Analysis of colorectal cancers in British Bangladeshi identifies early onset, frequent mucinous histotype and a high prevalence of RBFOX1 deletion. *Mol Cancer.* 2013; 12:1.
16. Kurashina K, Yamashita Y, Ueno T, Koinuma K, Ohashi J, Horie H, Miyakura Y, Hamada T, Haruta H, Hatanaka H, Soda M, Choi YL, Takada S, et al. Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. *Cancer Sci.* 2008; 99:1835–1840.
17. Schier AF. Nodal signaling in vertebrate development. *Annu Rev Cell Dev Biol.* 2003; 19:589–621.
18. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012; 487:330–7.
19. Topczewska JM, Postovit LM, Margaryan NV, Sam A, Hess AR, Wheaton WW, Nickoloff BJ, Topczewski J, Hendrix MJ. Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat Med.* 2006; 12:925–932.
20. Strizzi L, Bianco C, Normanno N, Salomon D. Cripto-1: a multifunctional modulator during embryogenesis and oncogenesis. *Oncogene.* 2005; 24:5731–5741.
21. Sun Y, Strizzi L, Raafat A, Hirota M, Bianco C, Feigenbaum L, Kenney N, Wechselberger C, Callahan R, Salomon DS. Overexpression of human Cripto-1 in transgenic mice delays mammary gland development and differentiation and induces mammary tumorigenesis. *Am J Pathol.* 2005; 167:585–597.
22. Gong Y, Guo Y, Hai Y, Yang H, Liu Y, Yang S, Zhang Z, Ma M, Liu L, Li Z, He Z. Nodal promotes the self-renewal of human colon cancer stem cells via an autocrine manner through Smad2/3 signalling pathway. *Biomed Res Int.* 2014; 2014:364134.
23. Haffner C, Frauli M, Topp S, Irmeler M, Hofmann K, Regula JT, Bally-Cuif L, Haass C. Nicalin and its binding partner Nomo are novel Nodal signalling antagonists. *EMBO J.* 2004; 23:3041–50.
24. Gonzalez MB, Hernandez JM, Garcia JL, Lumbreras E, Castellanos M, Hernández JM, Fernández-Calvo J, Gutiérrez NC, San Miguel JF. The value of fluorescence *in situ* hybridization for the detection of 11q in multiple myeloma. *Haematologica.* 2004; 89:1213–1218.