

UNIVERSIDAD DE SALAMANCA

FACULTAD DE MEDICINA

DEPARTAMENTO DE MEDICINA



TESIS DOCTORAL

**CARACTERIZACIÓN DE LA PÉRDIDA DEL GEN *NOMO1*
EN CÁNCER COLORRECTAL**

ABEL JESÚS MARTEL MARTEL

2022



El Prof. Dr. ROGELIO GONZÁLEZ SARMIENTO, Catedrático del Departamento de Medicina de la Universidad de Salamanca.

El Prof. Dr. JUAN JESÚS CRUZ HERNÁNDEZ, Catedrático del Departamento de Medicina de la Universidad de Salamanca.

CERTIFICAN:

Que el trabajo titulado "*CARACTERIZACIÓN DE LA PÉRDIDA DEL GEN NOMO1 EN CÁNCER COLORRECTAL*", que presenta el Graduado en Biología Don Abel Jesús Martel Martel, ha sido realizado bajo su dirección en el Departamento de Medicina y reúne, a su juicio, todos los requisitos necesarios para ser presentado ante el tribunal correspondiente, a fin de optar al Grado de Doctor por la Universidad de Salamanca.

Y para que así conste y a los efectos oportunos, se expide el presente certificado en Salamanca, a 9 de junio de 2022.

Fdo.: Dr. Rogelio González Sarmiento

Fdo.: Dr. Juan Jesús Cruz Hernández

El doctorando ha realizado este trabajo bajo un contrato predoctoral asociado al Servicio de Oncología Médica-IBSAL, financiado a través de los siguientes organismos:

- Fundación Ángel Alonso (*DONAC17-029*): 03/2018 – 03/2019.
- Fundación Doctor Moraza (*FMoraza18/001*): 03/2019 – 07/2022.

El presente trabajo ha sido financiado por el Instituto de Salud Carlos III mediante los proyectos PI16/01920 y PI20/01569.

A mis padres
A mi hermano

Lo que sucede conviene

Agradecimientos

En primer lugar, quisiera agradecer a mis directores de tesis por brindarme la oportunidad de formarme y aprender junto a ellos durante estos últimos 5 años: Al **Dr. Rogelio González Sarmiento**, por su dirección, orientación e inestimable ayuda para sacar este trabajo adelante. Por transmitirme su pasión por la ciencia y la medicina, sin olvidar que primero se forman las personas y luego los profesionales. Ha sido un privilegio aprender junto a mi gran maestro y, sobre todo, hacerlo de esta manera. Al **Dr. Juan Jesús Cruz Hernández**, por depositar su confianza en mi desde el primer día, por abrirme las puertas del Servicio de Oncología Médica para que pudiera sentir desde dentro su pasión por la medicina, y por su permanente disponibilidad para ayudarme a crecer en esta profesión.

Mis palabras de agradecimiento también van dirigidas al **Dr. José Perea** y al **Dr. Juan Luis García**, por ser piezas fundamentales de este proyecto, así como por su esfuerzo e interés en ser partícipes de mi aprendizaje. Al **Dr. Luis Corchete** por su paciencia, su disponibilidad, y su incansable ayuda en este trabajo. Al Servicio de Oncología Médica del Hospital Universitario de Salamanca, por proporcionarme los fármacos empleados en este estudio, así como a todos sus miembros por colaborar en mi formación en este campo que tanto me apasiona. Agradezco al **Dr. Óscar Blanco**, del Servicio de Anatomía Patológica, y al **Dr. Manuel Sánchez**, del Servicio de Experimentación Animal de la USAL, su dedicación y colaboración en este proyecto.

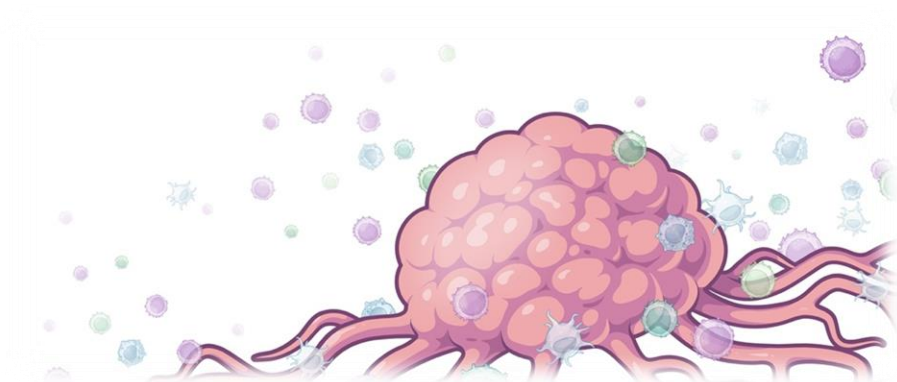
Muchísimas gracias a todos mis compañeros del Departamento de Medicina y del laboratorio 14 del Centro de Investigación del Cáncer, tanto a los que han pasado por el laboratorio como a los que siguen siendo testigos de mi día a día en esta aventura. Es muy complicado nombrar a todas las personas que han pasado por aquí, pero, sin duda, habéis sido muy importantes en esta andadura, siendo mi gran apoyo tanto dentro como fuera del laboratorio. Quisiera hacer una mención especial a la que ha sido mi mentora durante estos 5 años, a la **Dra. Jéssica Pérez**. No solo por compartir proyecto, sino también por empujarme y animarme a luchar contra las dificultades que este ha planteado. No ha sido fácil, pero hubiese sido imposible sin su constante esfuerzo, ilusión y sacrificio. De verdad, muchas gracias. No me quiero olvidar de los integrantes del laboratorio 4 del CIC, que compartieron conmigo dos importantes años, ni de su director, el **Dr. Pedro Lazo**, por ayudarme a dar mis primeros pasos en esta carrera.

A mi familia, por su constante apoyo para que pudiera cumplir mis propósitos, para que no me faltara una sonrisa, en definitiva, para que no me faltara nada. Gracias a mis padres por inculcarme los valores con los que he crecido y que hoy me permiten llegar hasta aquí: el respeto, la educación y el sacrificio. Gracias por empujarme al avión cada vez que ha hecho falta, desde hace más de diez años, sin que faltaran palabras de ánimo, porque sabíais que mi sueño estaba fuera. A mi hermano, mi ejemplo y apoyo constante.

En último lugar, mi agradecimiento y apoyo a todas aquellas personas que padecen esta enfermedad. Sin ellos, nada de lo que hacemos sería posible ni tendría sentido. Ni siquiera la vocación por esta profesión sería la misma sin la confianza que depositan en nosotros. A mi abuelo, por ejemplificar lo que dicen estas líneas, y luchar día tras día contra esta enfermedad sin perder las ganas y la esperanza. ¡Qué lección de vida!

A todos ellos, muchas gracias.

Índices



Índice

Índice	1
Índice de figuras	5
Índice de tablas	8
Glosario de abreviaturas	13
Introducción	19
1. El Cáncer	19
2. El cáncer colorrectal (CCR)	21
2.1. Epidemiología del CCR	23
2.2. Clasificación del CCR	24
2.2.1. Clasificación TNM	24
2.2.2. Clasificación histológica	25
2.3. Tratamiento del CCR	26
2.3.1. Quimioterapia	27
2.3.2. Terapia dirigida	27
2.3.3. Inmunoterapia	28
3. Etiología del CCR	28
3.1. Factores de riesgo asociados a la historia médica personal	28
3.2. Factores de riesgo hereditario	29
3.3. Factores de riesgo biológicos	29
3.3.1. Microbiota intestinal	29
3.3.2. Edad	29
3.3.3. Género y raza	29
3.4. Factores de riesgo asociados al estilo de vida	30
3.4.1. Dieta	30
3.4.2. Sobrepeso y obesidad	30
3.4.3. Inactividad física	30
3.4.4. Tabaco y alcohol	31
4. Síndromes de predisposición hereditaria al CCR	31
4.1. CCR hereditario no polipósico (CCHNP) o síndrome de Lynch (SL)	32
4.2. Síndromes polipósicos	33
4.2.1. Poliposis adenomatosa familiar	33
4.2.2. Poliposis adenomatosa asociada a <i>MUTYH</i>	33
4.2.3. Poliposis asociada a la actividad correctora de errores de la polimerasa	34
4.2.4. Síndrome de Peutz-Jeghers	34
4.2.5. Síndrome de poliposis juvenil	34
4.2.6. Síndrome PTEN-Hamman-Richards-Tumor	34
4.2.7. Síndrome de poliposis mixta hereditaria	35
5. Carcinogénesis del CCR	35
5.1. Vía de inestabilidad cromosómica (CIN)	36
5.2. Vía de inestabilidad de microsatélites (MSI)	36
5.3. Vía serrada o vía del fenotipo metilador de Islas CpG (CIMP)	37
5.4. Clasificación molecular del CCR	38
6. Cáncer colorrectal de aparición temprana (EOCRC)	39
6.1. Epidemiología del EOCRC	39
6.2. Características clínico-patológicas del EOCRC	39
6.3. Perfil molecular del EOCRC	40
6.4. Perfil cromosómico del EOCRC	42

6.5. Región cromosómica 16p13.12-p13.11 en EOCRC	43
7. Papel de <i>NOMO1</i> en la carcinogénesis del EOCRC.....	44
7.1. Aspectos generales de <i>NOMO1</i>	45
7.2. Complejo NOMO/NCLN/TMEM147	45
8. Vía de señalización Nodal	46
8.1. Papel de la vía Nodal en CCR	48
Hipótesis	53
Objetivos.....	57
Material y métodos	61
1. Selección de pacientes.....	61
2. Técnicas de manipulación de ácidos nucleicos	61
2.1. Extracción de DNA a partir de sangre periférica	61
2.2. Extracción de DNA a partir de líneas celulares y MEFs.....	62
2.3. Extracción de DNA a partir de tejido fresco de ratón.....	62
2.4. Extracción de RNA a partir de líneas celulares y MEFs.....	62
2.5. Extracción de RNA a partir de tejido fresco de ratón	63
2.6. Cuantificación de ácidos nucleicos	63
2.7. Amplificación de fragmentos de DNA mediante PCR	64
2.8. Electroforesis en gel de agarosa	65
2.9. Purificación de fragmentos de DNA amplificados mediante PCR.....	65
2.10. Secuenciación mediante el método de Sanger	65
2.11. Síntesis de DNA complementario (cDNA) mediante transcripción reversa (RT)	66
2.12. PCR cuantitativa-comparativa en tiempo real (RT-qPCR).....	67
2.13. Estudio de longitud telomérica mediante RT-qPCR.....	71
2.14. Discriminación alélica mediante PCR con sondas TaqMan®	72
2.15. Estudio del transcriptoma	72
2.15.1. Microarrays de expresión (WT PLUS Affymetrix)	73
2.15.2. Secuenciación de RNA (RNA-seq)	73
3. Técnicas de manipulación de proteínas	75
3.1. Extracción de proteínas a partir de líneas celulares	75
3.2. Cuantificación de proteínas por el método Bradford	75
3.3. Electroforesis en SDS-PAGE	75
3.4. Transferencia semi-seca y western blot (WB)	76
3.5. Anticuerpos	77
3.6. Estudio del proteoma mediante cromatografía líquida acoplada a espectrometría de masas (LC-IMS/MS).....	78
4. Análisis de enriquecimiento funcionales	79
5. Cultivo de líneas celulares	80
5.1. Contaje de células.....	80
5.2. Fármacos	81
5.3. Ensayo de viabilidad celular	82
5.4. Análisis de muerte celular mediante citometría de flujo	83
5.5. Análisis del estado <i>stem cell</i> mediante citometría de flujo	84
6. Estudio de migración celular	85
6.1. Ensayo de <i>scratch</i> o <i>cierre de herida</i>	85
6.2. Ensayo de migración celular mediante <i>transwell</i>	86
7. Edición génica mediante la tecnología CRISPR/cas9.....	87
7.1. Diseño de sgRNAs	89

7.2. Clonaje de sgRNAs	91
7.3. Transformación de bacterias competentes <i>Escherichia coli</i> DH5- α	94
7.4. Extracción y purificación de DNA plasmídico	95
7.5. Confirmación de la inserción del sgRNA mediante secuenciación Sanger	95
7.6. Transfección y selección de las líneas celulares HCT-116 y HS-5	96
7.7. <i>Cell sorting</i> mediante citometría de flujo	98
8. Modelo murino para el estudio del cáncer colorrectal de aparición temprana.....	100
8.1. Bienestar animal	100
8.2. Generación de un ratón condicional para <i>Nomo1</i> mediante la tecnología CRISPR/cas9	100
8.3. Extracción de fibroblastos embrionarios de ratón (MEFs)	103
8.4. Deleción de <i>Nomo1</i> en modelos <i>in vitro</i> e <i>in vivo</i>	104
8.5. Generación de un ratón condicional <i>Nomo1</i> Flox / Msi1-Cre	105
8.6. Inducción de carcinogénesis colorrectal en ratones <i>Nomo1</i> Flox / Msi1-Cre	105
8.7. Estudio anatomopatológico	106
9. Análisis estadístico	106
Resultados	109
CAPÍTULO I. CARACTERIZACIÓN FUNCIONAL DE LA PÉRDIDA DE <i>NOMO1</i> EN CÁNCER COLORRECTAL DE APARICIÓN TEMPRANA	109
1. Generación de líneas celulares <i>knockout</i> para <i>NOMO1</i> mediante el sistema de edición génica CRISPR/cas9.....	109
1.1. Validación de la edición génica de las líneas celulares HCT-116 y HS-5	109
2. Caracterización fenotípica de las líneas celulares <i>NOMO1 knockout</i>	112
2.1. Análisis de expresión del complejo NOMO/NCLN/TMEM147	112
2.2. Análisis de expresión proteica de los receptores de la vía de señalización Nodal	114
2.3. Estudio de la actividad de la vía Nodal	115
2.4. Estudio de la viabilidad celular	118
2.5. Estudio del estado <i>stem cell</i>	119
3. Estudio del perfil transcriptómico y proteómico de las líneas celulares carentes de <i>NOMO1</i>	121
3.1. Estudio del transcriptoma	121
3.2. Estudio del proteoma	128
4. Análisis de la capacidad migratoria	131
CAPÍTULO II. CARACTERIZACIÓN DE LA PÉRDIDA DE <i>NOMO1</i> EN UN MODELO MURINO CONDICIONAL	135
1. Generación de un modelo murino condicional para <i>Nomo1</i> mediante el sistema de edición génica CRISPR/cas9.....	135
1.1. Validación del sistema cre-loxp <i>in vitro</i>	136
1.2. Validación del sistema cre-loxp <i>in vivo</i>	138
2. Análisis anatomopatológico de los modelos murinos con inactivación de <i>Nomo1</i> ...	139
2.1. Inactivación de <i>Nomo1</i> mediante inyección con adenovirus portador de la recombinasa Cre.....	139
2.2. Inactivación de <i>Nomo1</i> en ratones Msi1-Cre	140
2.3. Modelo de carcinogénesis colorrectal mediante administración de 1,2-dimetilhidrazina (DMH)..	142
CAPÍTULO III. ESTUDIO DE LA RESPUESTA AL TRATAMIENTO CON AGENTES QUIMIOTERÁPICOS	144

1. Estudio de la respuesta al tratamiento con 5-fluorouracilo, irinotecán, oxaliplatino y cisplatino de las líneas celulares carentes de <i>NOMO1</i>	144
1.1. Análisis de viabilidad celular mediante MTT	144
1.2. Análisis de viabilidad celular mediante el ensayo Anexina V/Yoduro de Propidio.....	147
CAPÍTULO IV. ESTUDIO DEL ACORTAMIENTO TELOMÉRICO EN PACIENTES CON CÁNCER COLORRECTAL DE APARICIÓN TEMPRANA	150
<i>Discusión</i>.....	155
1. CRISPR/Cas9 genera eficientemente la inactivación de <i>NOMO1</i> en líneas celulares	156
2. La desestabilización del complejo NOMO/NCLN no afecta a los niveles de expresión de TMEM147	158
3. La inactivación de NOMO no modifica la actividad de la vía Nodal	159
4. La inactivación de NOMO no modifica el estado <i>stem cell</i> ni la viabilidad celular en respuesta al tratamiento	161
5. La inactivación del gen <i>NOMO1</i> genera un perfil transcripcional y proteómico diferencial en las líneas celulares HCT-116 y HS-5	162
6. La inactivación de NOMO promueve la migración celular en EOCRC	164
7. La pérdida del gen <i>NOMO1</i> es una mutación <i>passenger</i> en EOCRC	165
8. La longitud telomérica como un marcador predictivo del CCR.....	167
<i>Conclusiones</i>	171
<i>Referencias</i>	175
<i>Anexos</i>	191
Anexo 1. Rutas afectadas en las líneas celulares HCT-116 y HS-5 tras la inactivación de <i>NOMO1</i> . Set de genes identificados en el microarray de expresión comparados con las bases de datos “KEGG” y “Reactome”.....	191
Anexo 2. Rutas afectadas en las líneas celulares HCT-116 y HS-5 tras la inactivación de <i>NOMO1</i> . Set de genes identificados en la secuenciación de RNA comparados con las bases de datos “KEGG” y “Reactome”.....	192
Anexo 3. Cuantificación del porcentaje de área ocupada de los clones NOMO-WT y NOMO-KO de cada línea celular. Se representan todas las réplicas a todos los tiempos estudiados. ...	194
Anexo 4. Cuantificación del número de partículas positivas (células que han migrado) de los clones NOMO-WT y NOMO-KO de cada línea celular. Se representan todas las réplicas a todos los tiempos estudiados.....	196
<i>Artículo 1</i>	201
<i>Artículo 2</i>	223

Índice de figuras

Figura 1. Características del cáncer.....	19
Figura 2. Incidencia estimada de tumores en la población mundial para los años 2020 y 2040.	20
Figura 3. Anatomía del intestino grueso.	21
Figura 4. Histología de los pólipos adenomatosos.....	22
Figura 5. Histología del cáncer colorrectal.....	26
Figura 6. Clasificación fenotípica de los síndromes de predisposición hereditaria al CCR con y sin poliposis, modo de herencia, genes causales y vías moleculares afectadas.	31
Figura 7. Modelo de secuencia adenoma-carcinoma propuesto por Fearon y Vogelstein.	35
Figura 8. Vías de carcinogénesis del CCR.	38
Figura 9. Esquema de la vía de señalización Nodal y complejo antagonista NOMO/NCLN/TMEM147.....	47
Figura 10. Programa de amplificación estándar empleado para la amplificación del fragmento de interés, donde X es la temperatura de anillamiento específica de cada pareja de oligonucleótidos.....	64
Figura 11. Secuencia correspondiente al exón 3 del gen <i>NOMO1</i> (46 pb).	66
Figura 12. Esquema de síntesis de DNA complementario (cDNA) a partir de RNA mensajero (mRNA).	67
Figura 13. Curva de amplificación de DNA medida por el aumento de fluorescencia del agente intercalante.	68
Figura 14. Representación gráfica de las curvas de fusión de los productos de amplificación para el control endógeno (<i>LEMD3</i>) y el fragmento de interés del gen <i>NOMO1</i>	69
Figura 15. Representación gráfica del programa establecido en el termociclador para la cuantificación relativa del número de copias del fragmento de interés del gen <i>NOMO1</i> mediante RT-qPCR.	70
Figura 16. Representación gráfica del programa establecido en el termociclador para la cuantificación absoluta de la longitud telomérica de pacientes con EOCRC.	72
Figura 17. Esquema de trabajo para la secuenciación de RNA.	74
Figura 18. Representación gráfica del procedimiento para medir la viabilidad celular mediante el ensayo de MTT.	82
Figura 19. Representación gráfica de las poblaciones detectadas mediante citometría de flujo en el análisis de muerte celular (Anexina V/Yoduro de propidio).	83
Figura 20. Funcionamiento básico de los principales métodos de edición del genoma.....	87
Figura 21. Arquitectura bilobular de la enzima Cas9.	88
Figura 22. Mecanismos de reparación del daño en el DNA inducido por la acción de la enzima Cas9.	89
Figura 23. Representación gráfica de la secuencia genómica del gen <i>NOMO1</i>	90
Figura 24. Esquema del vector PX458.....	92

Figura 25. Secuencias correspondientes a los sgRNAs insertados en el vector PX458. Se representa con sombreado azul la secuencia específica de cada sgRNA.	96
Figura 26. Imágenes de microscopía de fluorescencia de las líneas celulares HCT-116 y HS-5 transfectadas con el plásmido vacío (PX458) y con la combinación de sgRNAs (G1+G3). Se detectan en verde las células transfectadas eficientemente (GFP+).	98
Figura 27. Funcionamiento del sistema Cre-loxp.	101
Figura 28. Representación esquemática de la generación de un alelo mutado de <i>Nomo1</i> mediante CRISPR/cas9 para el funcionamiento del sistema Cre-loxp.	102
Figura 29. Proceso de edición génica en modelos murinos mediante la tecnología CRISPR/cas9.	103
Figura 30. Secuencias de nucleótidos correspondientes al exón 3 de <i>NOMO1</i> y sus regiones intrónicas adyacentes.	110
Figura 31. Normalización de la cantidad de DNA amplificado para el fragmento de interés del gen <i>NOMO1</i> mediante PCR cuantitativa en tiempo real.	111
Figura 32. Análisis de la expresión proteica de NOMO en las líneas celulares HCT-116 y HS-5.	112
Figura 33. Niveles de expresión de los miembros del complejo NOMO/NCLN/TMEM147 en las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO.	113
Figura 34. Niveles de expresión proteica de los receptores (ACTRII-B y ALK4) y del correceptor (Cripto-1) de la vía de señalización Nodal en las líneas celulares HCT-116 y HS-5.	115
Figura 35. Niveles de expresión de las proteínas SMAD2, SMAD3, pSMAD2 y pSDMAD3 en los clones NOMO-WT y NOMO-KO de las líneas celulares HCT-116 y HS-5.	117
Figura 36. Niveles de expresión proteica de SMAD4 en los clones NOMO-WT y NOMO-KO de las líneas celulares HCT-116 y HS-5.	118
Figura 37. Efecto de la inactivación de NOMO sobre la viabilidad celular.	119
Figura 38. Expresión de los marcadores de membrana CD44 y CD24 en las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO.	120
Figura 39. Perfil de expresión diferencial de las líneas celulares HCT-116 y HS-5 con inactivación de <i>NOMO1</i> . Representación mediante heatmap de los genes comunes desregulados en el microarray de expresión	122
Figura 40. Procesos biológicos desregulados asociados a la pérdida de <i>NOMO1</i> en las líneas celulares HCT-116 y HS-5. Se representan los procesos alterados al introducir el set de genes desregulados en el microarray de expresión en el análisis GSEA-hallmarks	123
Figura 41. Perfil de expresión diferencial de las líneas celulares HCT-116 y HS-5 con inactivación de <i>NOMO1</i> . Representación mediante heatmap de los genes comunes desregulados en la secuenciación de RNA	124
Figura 42. Procesos biológicos desregulados asociados a la pérdida de <i>NOMO1</i> en las líneas celulares HCT-116 y HS-5. Se representan los procesos alterados al introducir el set de genes desregulados en la secuenciación de RNA en el análisis GSEA-hallmarks	125
Figura 43. Procesos biológicos desregulados en las líneas celulares carentes de <i>NOMO1</i> , al incluir los genes diferencialmente expresados que mostraban asociación con la EMT en el análisis de sobre-representación de rutas.	126

Figura 44. Expresión proteica de los marcadores de EMT en las líneas celulares HCT-116 y HS-5 con inactivación de NOMO.	127
Figura 45. Representación mediante diagramas de Venn de las proteínas infra-expresadas y sobre-expresadas identificadas mediante LC-IMS/MS de forma común en las líneas celulares HCT-116 y HS-5 con inactivación de NOMO.	128
Figura 46. Expresión de las proteínas asociadas a migración y EMT que se muestran desreguladas en el proteoma de las líneas celulares HCT-116 y HS-5 carentes de NOMO.	130
Figura 47. Análisis mediante ensayo de cierre de herida de la capacidad migratoria de dos clones NOMO-KO y dos clones <i>wild type</i> de las líneas celulares HCT-116 y HS-5.....	132
Figura 48. Análisis de la capacidad migratoria de las líneas HCT-116 y HS-5 mediante ensayo de <i>transwell</i> . Se han comparado dos clones NOMO-KO y dos clones NOMO-WT de cada línea celular.....	134
Figura 49. Extracción de fibroblastos embrionarios de ratón (MEFs) para la validación <i>in vitro</i> del sistema cre-loxp.	136
Figura 50. Validación del sistema cre-loxp <i>in vitro</i> . Escisión del exón 3 de <i>Nomo1</i> en fibroblastos embrionarios de ratón (MEFs) mediante infección con AD5CMVCre-GFP.....	137
Figura 51. Validación del sistema cre-loxp <i>in vivo</i>	138
Figura 52. Histología del intestino delgado y grueso normal.....	140
Figura 53. Angiosarcoma rectal de un animal <i>Nomo1^{flox/wt}; Msi1-Cre+</i>	141
Figura 54. Histología del tejido pulmonar e intestinal de animales con inactivación de <i>Nomo1</i> en heterocigosis tratados con DMH. Se representa un adenocarcinoma de pulmón y un adenocarcinoma de colon.....	143
Figura 55. Análisis de viabilidad celular mediante MTT de las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO, tras el tratamiento con agentes quimioterápicos.	147
Figura 56. Análisis de viabilidad celular mediante citometría de flujo de las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO, tras el tratamiento con agentes quimioterápicos.	149
Figura 57. Cuantificación mediante RT-qPCR de la longitud telomérica absoluta en DNA procedente de sangre periférica de pacientes con EOCRC e individuos sanos. Se representan las longitudes teloméricas absolutas por célula diploide.....	150

Índice de tablas

Tabla 1. Sistema de clasificación internacional TNM para el cáncer colorrectal.	25
Tabla 2. Características clínico-patológicas y moleculares del CCR de aparición temprana y del CCR de aparición tardía.	40
Tabla 3. Genes localizados en la región cromosómica 16p13.12-p13.11.	44
Tabla 4. Oligonucleótidos empleados para la amplificación de la región de interés del gen <i>NOMO1</i> mediante PCR: <i>NOMO1</i> -humano (<i>h-NOMO1</i>) y <i>Nomo1</i> -ratón (<i>m-Nomo1</i>).	64
Tabla 5. Condiciones para la síntesis de cDNA mediante transcripción reversa (RT).	67
Tabla 6. Oligonucleótidos utilizados en los ensayos de pérdida de heterocigosidad (LOH) mediante PCR cuantitativa-comparativa en tiempo real (RT-qPCR): <i>NOMO1</i> G1+G2 y <i>LEMD3</i>	69
Tabla 7. Oligonucleótidos utilizados en los ensayos de expresión génica mediante PCR cuantitativa-comparativa en tiempo real (RT-qPCR).	70
Tabla 8. Anticuerpos primarios y secundarios utilizados en este trabajo para la determinación de expresión proteica mediante western blot (WB).	77
Tabla 9. Principales características de las líneas celulares utilizadas.	80
Tabla 10: Representación de las dosis utilizadas en los ensayos de viabilidad celular en respuesta al tratamiento con 5-FU, irinotecán, oxaliplatino y cisplatino en líneas celulares.	81
Tabla 11. Anticuerpos utilizados para la determinación del estado de diferenciación mediante citometría de flujo.	85
Tabla 12. Secuencias dianas y oligonucleótidos necesarios para la construcción de los sgRNAs de <i>NOMO1</i> . En minúscula se indican las secuencias que facilitan la creación de los extremos cohesivos y su inserción en el plásmido.	91
Tabla 13. Etapas y reactivos necesarios para el clonaje de los sgRNAs en el vector PX458.	93
Tabla 14. Condiciones de temperatura y tiempo para cada etapa en el clonaje de sgRNAs en el vector PX458.	94
Tabla 15. Análisis de la eficiencia de transfección en las líneas celulares HCT-116 y HS-5 mediante citometría de flujo.	99
Tabla 16. Secuencias dianas y oligonucleótidos necesarios para la construcción de los sgRNAs de <i>Nomo1</i> en ratón.	102
Tabla 17. Genes asociados a la transición epitelio-mesénquima (EMT) que se encuentran desregulados en las líneas celulares HCT-116 y HS-5 tras la inactivación de <i>NOMO1</i>	126
Tabla 18. Proteínas desreguladas tras la pérdida de <i>NOMO</i> en las líneas HCT-116 y HS-5, identificadas mediante LC-IMS/MS.	129
Tabla 19. Análisis anatomopatológico de los ratones condicionales para <i>Nomo1</i> infectados con el adenovirus portador de la recombinasa Cre.	139
Tabla 20. Análisis anatomopatológico de los ratones <i>Nomo1</i> ^{flox/wt} con expresión constitutiva de la recombinasa Cre.	141
Tabla 21. Análisis anatomopatológico de los ratones <i>Nomo1</i> ^{flox/wt} con expresión constitutiva de la recombinasa Cre y de los ratones control en un modelo de carcinogénesis colorrectal inducida por DMH.	142

Tabla 22. Distribución genotípica del polimorfismo rs2853669 del gen <i>hTERT</i> en EOCRC y en la población control.	151
--	-----

Abreviaturas



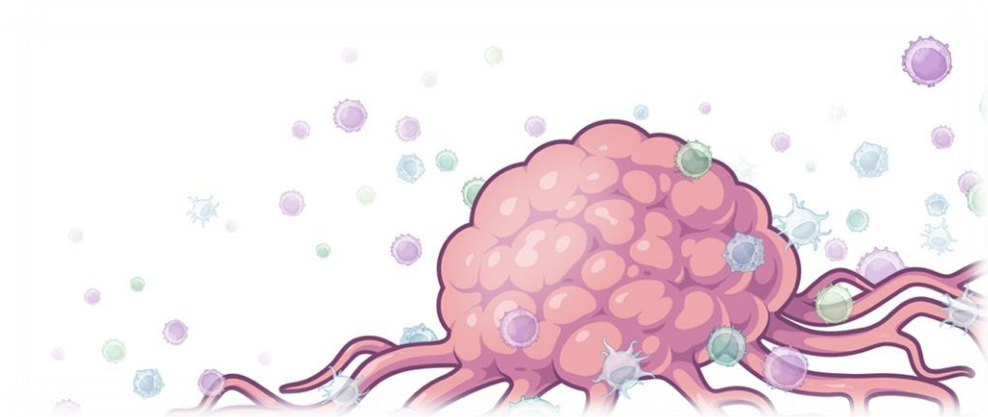
Glosario de abreviaturas

- **5 FU**: 5-fluorouracilo
- **ADC**: Adenocarcinoma
- **AJCC**: *International American Joint Committee on Cancer*
- **APS**: Persulfato de amonio
- **BLS**: Buffer de lisis suave
- **BSA**: Albúmina sérica bovina
- **CAP**: Capecitabina
- **CCHNP**: Cáncer colorrectal hereditario no polipósico
- **CCR**: Cáncer colorrectal
- **CCSC**: *Colon cancer stem cell*
- **cDNA**: DNA complementario
- **CIAA**: Cloroformo-alcohol isoamílico
- **CIMP**: Fenotipo metilador de islas CpG
- **CIN**: Inestabilidad cromosómica
- **CNA**: Alteración en el número de copias
- **CRISPR**: *Clustered Regularly Interspaced Short Palindromic Repeat*
- **CSC**: *Cancer stem cell*
- **CT**: Ciclo umbral
- **CTND1**: Delta-catenina
- **DMEM**: *Dulbecco's Modified Eagle Medium*
- **DMH**: 1,2-Dimetilhidrazina
- **DMSO**: Dimetilsulfóxido
- **DNA**: Ácido desoxirribonucleico
- **DSB**: Rotura de doble hebra
- **EDTA**: Ácido etilendiaminotetraacético
- **EGFR**: Receptor del factor de crecimiento epidérmico
- **EII**: Enfermedad inflamatoria intestinal
- **EMT**: Transición epitelio-mesénquima
- **EOCRC**: Cáncer colorrectal de aparición temprana
- **ESGE**: *European Society of Gastrointestinal Endoscopy*
- **FBS**: Suero fetal bovino
- **FC**: *Fold change*
- **FDR**: *False Discovery Rate*
- **FITC**: Fluoresceína-5-isotiocianato
- **GEO**: *Gene Expression Omnibus*
- **GFP**: Proteína verde fluorescente
- **GSEA**: *Gene Set Enrichment Analysis*
- **HDR**: Reparación directa por homología

- **HNSCC**: Carcinoma escamoso de cabeza y cuello
- **HR**: Recombinación homóloga
- **IARC**: Agencia Internacional para la Investigación del Cáncer
- **IMC**: Índice de masa corporal
- **IRI**: Irinotecán
- **KI**: *Knockin*
- **KO**: *Knockout*
- **LC-IMS/MS**: Cromatografía líquida acoplada a espectrometría de masas
- **LCRs**: *Low copy repeats*
- **LMNB1**: Lamina B1
- **LOCRC**: Cáncer colorrectal de aparición tardía
- **LOH**: Pérdida de heterocigosidad
- **LT**: Longitud telomérica
- **MEFs**: Fibroblastos embrionarios de ratón
- **MSI**: Inestabilidad de microsatélites
- **MMR**: *Mismatch repair deficiency*
- **mRNA**: RNA mensajero
- **Msi1-Cre**: Musashi-1-Cre
- **MSS**: Estabilidad de microsatélites
- **NCLN**: Nicalina
- **NGS**: *Next Generation Sequencing*
- **NHEJ**: Unión de extremos no homólogos
- **NOMO1**: *Nodal Modulator 1*
- **OMS**: Organización Mundial de la Salud
- **ORA**: *Over Represent Analysis*
- **OSCC**: Carcinoma oral de células escamosas
- **OX**: Oxaliplatino
- **PAF**: Poliposis adenomatosa familiar
- **PAFA**: Poliposis adenomatosa familiar atenuada
- **PAM**: Poliposis adenomatosa asociada a MUTYH
- **PAPP**: Poliposis asociada a la actividad correctora de errores de la polimerasa
- **PBS**: Tampón fosfato salino
- **PCR**: Reacción en cadena de la polimerasa
- **PI**: Yoduro de propidio
- **PS**: Fosfatidilserina
- **RAR**: Regiones recurrentemente alteradas
- **RE**: Retículo endoplasmático
- **REDECAN**: Red Española de Registros de Cáncer
- **RNA-seq**: Secuenciación de RNA
- **RNA**: Ácido ribonucleico

- **rRNA**: RNA ribosómico
- **RT-qPCR**: PCR cuantitativa en tiempo real
- **RT**: Transcripción reversa
- **RTK**: Receptor tirosina-quinasa
- **SDH**: Succinato deshidrogenasa
- **SDS**: Dodecilsulfato sódico
- **SEOM**: Sociedad Española de Oncología Médica
- **SECOC**: *Spanish Early-Onset Colorectal Cancer Consortium*
- **sgRNA**: *Single-guide RNA*
- **SL**: Síndrome de Lynch
- **SMT**: Síndrome de Muir-Torre
- **SNP**: Polimorfismo de nucleótido único
- **SPHT**: Síndrome PTEN-Hamartoma-Tumor
- **SPJ**: Síndrome de poliposis juvenil
- **ssDNA**: *Single strand DNA*
- **STR**: *Short tandem repeat*
- **TGF- β** : Factor de crecimiento transformador beta
- **THH**: Telangiectasia hemorrágica hereditaria
- **Tm**: Temperatura de fusión
- **TNM**: *Tumor node metastasis*
- **tracrRNA**: RNA-CRISPR trans-activador
- **VEGF**: Factor de crecimiento del endotelio vascular
- **WB**: *Western blot*
- **WebGestalt**: *WEB-based Gene Set Analysis Toolkit*

Introducción



Introducción

1. El Cáncer

El término cáncer engloba una serie de enfermedades caracterizadas por un crecimiento y división descontrolado de las células, que origina una proliferación anormal de las mismas, dando lugar a un tumor y provocando graves alteraciones en la forma y función de los tejidos afectados. Además, las células tumorales pueden diseminarse a otros tejidos a través del torrente sanguíneo y sistema linfático, proceso que se conoce como metástasis^{1,2}.

El cáncer se produce como consecuencia de la acumulación de alteraciones genéticas y epigenéticas en la célula a lo largo del tiempo y afecta, principalmente, a los procesos de proliferación, diferenciación y muerte celular. En este sentido, en los últimos años se han descrito diversos mecanismos o capacidades que poseen las células tumorales: evasión de supresores de crecimiento, evasión del sistema inmune, permitir la inmortalidad replicativa, inflamación promovida por el tumor, activación de la invasión y metástasis, inducción de la angiogénesis, inestabilidad y mutación del genoma, resistencia a la muerte celular, disregulación del metabolismo celular y mantenimiento de la señal proliferativa^{3,4}. En este último año, se han añadido 4 características nuevas que corroboran la complejidad de esta enfermedad: desbloqueo de la plasticidad fenotípica, reprogramación epigenética, microbioma polimórfico y senescencia celular⁵ (Figura 1).

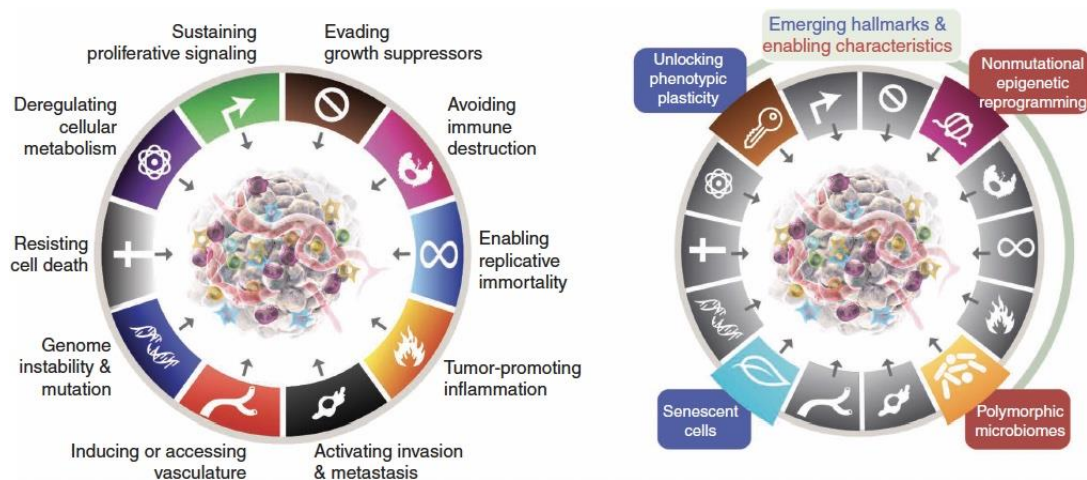


Figura 1. Características del cáncer (Hanahan *et al.* 2022).

Según el último informe publicado por la Sociedad Española de Oncología Médica (SEOM), el cáncer sigue constituyendo una de las principales causas de morbi-mortalidad a

nivel mundial. Se estima que en el año 2020 se diagnosticaron aproximadamente 18,1 millones de nuevos casos y murieron alrededor de 9,9 millones de personas. Sin embargo, la pandemia de la COVID-19 ha afectado tanto al número de diagnósticos como al registro de muertes por cáncer en muchos países, por lo que probablemente el número real habrá sido mayor. Además, para el año 2040 se prevé que el número de nuevos casos se situará en 27 millones (Figura 2) y la mortalidad alcanzará los 16 millones de fallecimientos⁶.

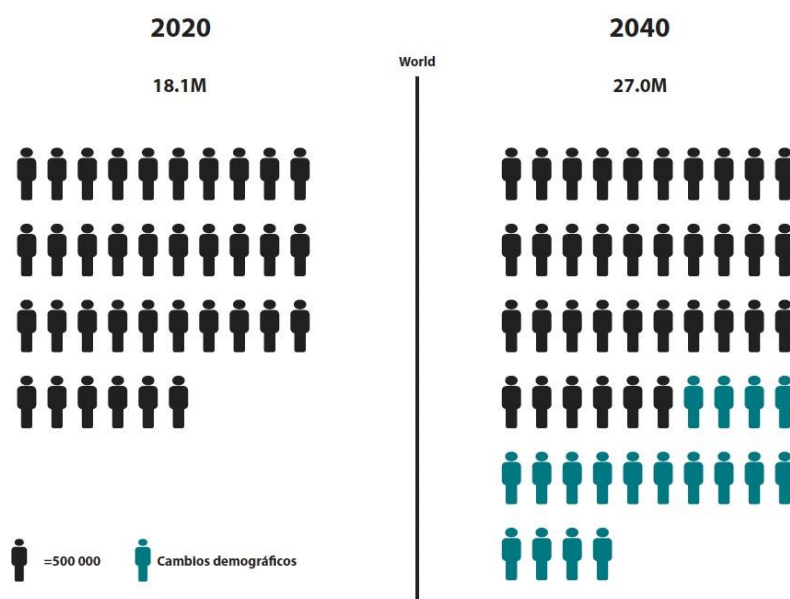


Figura 2. Incidencia estimada de tumores en la población mundial para los años 2020 y 2040. (Gráfico: GLOBOCAN, 2022).

En España, el número absoluto de cánceres diagnosticados ha aumentado en las últimas décadas debido al aumento poblacional. Según los cálculos de la Red Española de Registros de Cáncer (REDECAN), se estima que en 2022 el número de nuevos diagnósticos alcanzará los 280.100 casos. El tumor más frecuentemente diagnosticado será el colorrectal (43.370 nuevos casos), seguido del cáncer de mama (34.750), pulmón (30.948) y próstata (30.884). En cuanto a la mortalidad, se estima que en 2020 murieron por cáncer en España alrededor de 112.741 personas y que en 2040 se alcanzará la cifra de 160.000 defunciones. Actualmente, el tumor responsable de mayor mortalidad es el cáncer de pulmón (20,3%), seguido del cáncer colorrectal (14,6%), páncreas (6,7%), mama (5,8%) y próstata (5,1%)⁷.

2. El cáncer colorrectal (CCR)

El intestino grueso es un órgano del aparato digestivo inferior que conecta el intestino delgado y el ano. El colon es su parte mayoritaria y abarca desde el íleon hasta el recto con una longitud de aproximadamente 1,5 metros. Su función principal es la de absorber agua, nutrientes y electrolitos de los alimentos parcialmente digeridos. El material restante son los residuos sólidos o heces que se mueven a través del colon y se almacenan en el recto⁸. Anatómicamente, se divide en colon ascendente o proximal, colon transverso, colon descendente o distal, colon sigmoide y recto. La pared del colon está formada por las siguientes capas: mucosa, submucosa, muscular y serosa (Figura 3). Debido a las funciones que posee, el colon presenta una alta tasa de regeneración y está expuesto a diversos agentes fisiológicos, biológicos y químicos que incrementan el riesgo de padecer ciertas enfermedades, incluyendo el cáncer⁹.

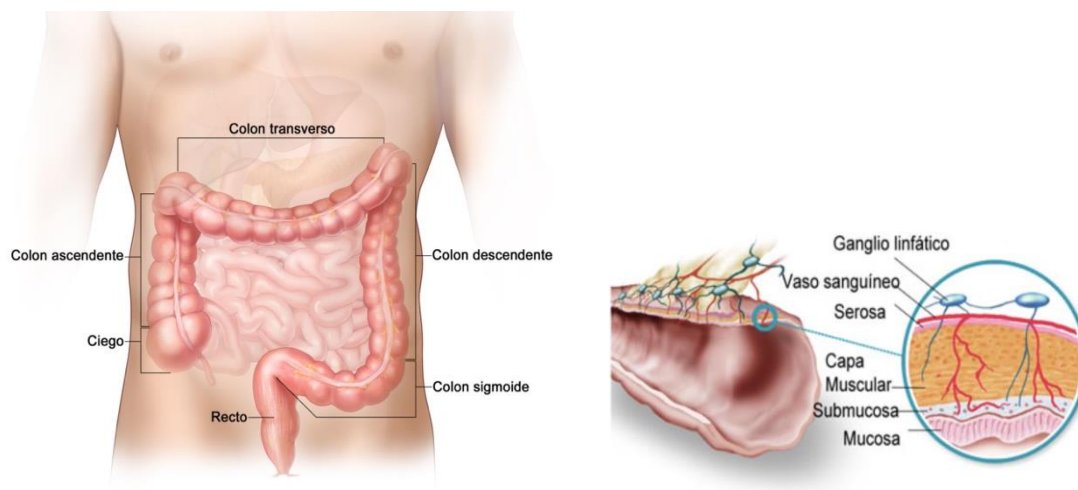


Figura 3. Anatomía del intestino grueso.

El cáncer colorrectal (CCR) constituye un conjunto heterogéneo de neoplasias localizadas en el colon o en el recto. En más del 90% de los casos, se presenta con una histología típica en forma de adenocarcinoma (ADC) que se desarrolla a partir de una lesión precursora benigna, denominada pólipo. Un pólipo se define como un crecimiento del revestimiento interno del colon o del recto que sobresale en el lumen^{10,11}. El periodo de malignización de estas lesiones abarca al menos 10 años, aunque pueden ser observadas y extirpadas con anterioridad durante las endoscopias en los procesos de cribado¹². Sin embargo, no todos los pólipos progresan a adenocarcinoma. Los tres tipos principales de pólipos son:

- **Pólipos inflamatorios o pólipos hiperplásicos:** los pólipos hiperplásicos no presentan displasia, carecen de riesgo de malignización, son asintomáticos y no necesitan seguimiento. Se trata de lesiones pequeñas (menos de 5 mm) y sésiles que se originan por defectos en el proceso de apoptosis de las células epiteliales que se localizan en las crestas de los pliegues¹³.

- **Pólipos serrados:** los pólipos serrados representan un grupo heterogéneo de lesiones (hiperplásicos, adenomas serrados tradicionales, adenomas sésiles, mixtos) que tienen una apariencia morfológica dentada. Son precursores de aproximadamente el 10-15% del CCR esporádico. El subtipo de pólipos serrados hiperplásicos es el más frecuente (80-90%) y pueden presentar la capacidad de malignizar^{14,15}.

- **Pólipos adenomatosos (adenomas):** representan el clásico precursor del CRC. Aproximadamente el 85-90% de los tumores colorrectales esporádicos se generan a partir de adenomas¹¹. Pueden presentar alto o bajo grado de displasia y se clasifican como tubulares, vellosos o túbulovellosos (Figura 4). Estos pólipos se inician en las glándulas productoras de moco que recubren el colon y el recto, pudiendo invadir otros tejidos si adquieren la capacidad de metastatizar¹². La agresividad de este tipo de tumor está estrechamente relacionada con el número de pólipos, el tamaño, el grado de displasia y la proporción de componen vellosos^{16,17}.

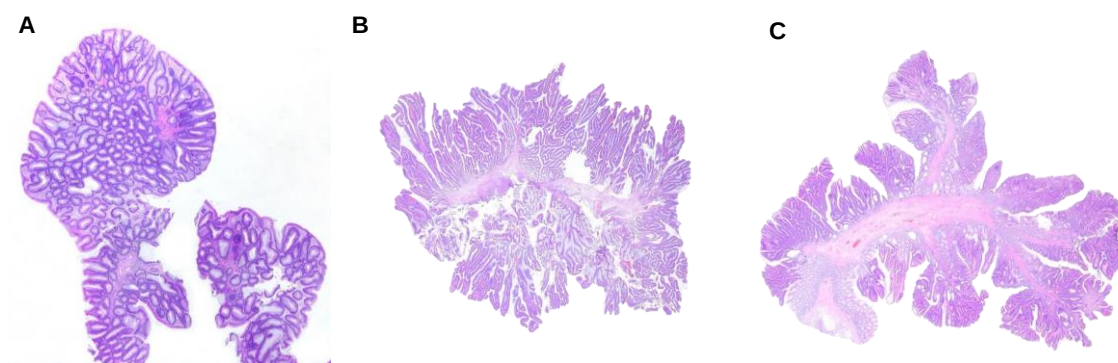


Figura 4. Histología de los pólipos adenomatosos. A- Adenoma tubular. B- Adenoma vellosos. C- Adenoma túbulovellosos.

Además, existen otras formas menos frecuentes de presentación del CCR como son: los carcinomas de células escamosas, los carcinomas adenoescamosos, los carcinomas de células fusiformes y los carcinomas indiferenciados^{18,19}. Por otro lado, existen otros tumores que se desarrollan en la región colorrectal como son los tumores del estroma gastrointestinal, los tumores carcinoides, linfomas y sarcomas²⁰.

2.1. Epidemiología del CCR

El CCR representa aproximadamente el 10% de todos los cánceres diagnosticados anualmente y de las muertes asociadas al cáncer en todo el mundo²¹. Según las estimaciones de la Agencia Internacional para la Investigación del Cáncer (IARC), en 2018 el CCR constituía 1,8 millones de nuevos casos y 900.000 muertes anuales, lo que le convierte en la tercera neoplasia maligna más diagnosticada y la segunda causa de muerte por cáncer²². Globalmente, el CCR es el segundo más común en mujeres y el tercero en hombres, siendo la incidencia y mortalidad un 25% más baja en el género femenino. Desde el punto de vista geográfico, las tasas más elevadas se registran en los países desarrollados por lo que se prevé que la incidencia del CCR en todo el mundo aumente hasta los 2,5 millones de casos nuevos en 2035^{21,23}.

La industrialización y el crecimiento económico han contribuido a la adquisición de un patrón dietético occidental, un estilo de vida sedentario y un aumento de la obesidad, que son los principales factores de riesgo del CCR asociados al estilo de vida²⁴. Por lo tanto, la occidentalización de los países unida a los altos niveles de ingresos y educación tiende a ir seguida de un aumento de la incidencia^{25,26}. En este sentido, las tasas de incidencia estandarizadas por edad varían a lo largo de la geografía. Por ejemplo, la incidencia registrada en Hungría (51,2 casos por 100.000 habitantes) es 45 veces más alta que la registrada en Gambia (1,1 casos por 100.000 habitantes) en el mismo año. Esta variación a nivel mundial representa un efecto combinado de múltiples factores, como el estilo de vida, la genética y la esperanza de vida. Por ejemplo, los países subdesarrollados pueden tener tasas de incidencia más bajas porque menos personas viven hasta los 65-79 años, momento en el que se diagnostica la mayor parte del CCR²².

Los patrones temporales de las tasas de incidencia del CCR también varían a nivel mundial. En los países que se encuentran en transición económica de América del Sur, Europa del Este y Asia hay un incremento de la incidencia (Brasil, Eslovaquia y China). En los países consolidados con rentas altas de América del Norte, Europa y Oceanía la incidencia está disminuyendo (EE. UU., Francia y Nueva Zelanda), se ha mantenido relativamente estable (Reino Unido y Australia) o está aumentando (Italia, Noruega y España)²³.

Estas tendencias de estabilización y disminución observadas en los países desarrollados se atribuyen principalmente a la implementación y mejora de los programas de cribado a los que se somete la población mayor de 50 años. De esta manera se permite un diagnóstico precoz de la enfermedad^{27,28}. Además, los avances en el tratamiento del CCR

también han permitido un descenso de la mortalidad en países como Eslovenia o Italia²³. Sin embargo, en los últimos años se ha observado un preocupante aumento de la incidencia del CCR en pacientes menores de 50 años, sobre todo de cáncer de recto y colon izquierdo²⁹⁻³².

En España, la estimación para el año 2022 realizada por REDECAN sitúa al CCR como el tumor más diagnosticado en la población general (43.370 casos). Se prevé que sea el segundo tumor más incidente en varones (26.862 casos), después del cáncer de próstata (30.884) y por delante del cáncer de pulmón (22.316). En mujeres, se estima que el CCR será el segundo tumor más frecuente (16.508 casos), por detrás del cáncer de mama que registrará aproximadamente 34.750 casos. En cuanto a la mortalidad, los últimos datos facilitados por la SEOM indican que el CCR fue el segundo tipo tumoral más letal para ambos sexos en 2020 (935.173 defunciones). En varones, fue el segundo más mortal por detrás del cáncer de pulmón. En mujeres, el CCR fue el tercero con más mortalidad, por detrás del cáncer de mama y pulmón³³.

2.2. Clasificación del CCR

En la actualidad, para el diagnóstico del CCR se emplean fundamentalmente dos clasificaciones: la clasificación TNM (*tumor node metastasis*) y la clasificación histológica. Cabe destacar que la clasificación del CCR es de gran importancia tanto para el diagnóstico como para el tratamiento, ya que este tipo de tumor muestra cambios en la supervivencia y en la respuesta a la quimioterapia en función del subtipo de CCR que se presente³⁴.

2.2.1. Clasificación TNM

La clasificación TNM es el método de estadiaje de neoplasias establecido por el *International American Joint Committee on Cancer* (AJCC)³⁵ y la Unión Internacional contra el Cáncer. Se trata de un sistema dual que determina el tamaño del tumor y hasta dónde se ha extendido la enfermedad en el organismo. Hoy en día, el sistema de clasificación TNM se emplea en la práctica clínica para evaluar el estadio en el que se encuentra el tumor en el momento del diagnóstico³⁶. Se basa en la observación de tres parámetros fundamentales (Tabla 1):

- El tamaño (T) y el grado de invasión local del tumor primario.
- El grado de afectación ganglionar (N). Según la AJCC, el número mínimo de ganglios que deben analizarse es de 10-13 ganglios linfáticos perirrectales y/o pericólicos³⁶.
- Presencia o ausencia de metástasis a distancia (M).

Tabla 1. Sistema de clasificación internacional TNM para el cáncer colorrectal.

T: Tumor primario	
Tx	No es posible evaluar el tumor primario
T0	No existe evidencia de tumor primario
Tis	Carcinoma in situ intraepitelial o con invasión de la lámina propia
T1	Tumor infiltrante a nivel de submucosa
T2	Tumor infiltrante a nivel de lámina muscular propia
T3	Tumor infiltrante a nivel de serosa
T4	Tumor perforante (peritoneo, vesical o invasión de órganos adyacentes)
N: Afectación ganglionar	
Nx	No se pueden evaluar los ganglios linfáticos regionales
N0	No existe diseminación hacia los ganglios linfáticos regionales
N1	Entre 1 y 13 ganglios linfáticos perirrectales o pericólicos afectados
N2	≥4 ganglios linfáticos perirrectales o pericólicos afectados
M: Metástasis a distancia	
Mx	No puede determinarse
M0	No existe evidencia de metástasis a distancia
M1	Presencia de metástasis a distancia

2.2.2. Clasificación histológica

A nivel histológico, el ADC colorrectal se define como un tumor maligno del tejido epitelial de la mucosa intestinal. Presenta alteraciones en las criptas de Lieberükn, invaginaciones del epitelio que ocupa desde la luz intestinal hasta la mucosa, mostrándose más alargadas que el tejido normal y con pérdida de la estructura normal de la mucosa. Estas células tumorales se caracterizan por su elevada tasa mitótica y un aumento del tamaño de los núcleos. Desde el punto de vista histopatológico, la Organización Mundial de la Salud (OMS) clasifica los ADC colorrectales en cinco subtipos³⁷ (Figura 5):

- **Adenocarcinoma mucinoso:** más del 50% de estos tumores están formados por células de Globet productoras de mucina. Es el subtipo más frecuente de CCR.

- **Adenocarcinoma de células en anillo de sello:** aunque también está formado por más del 50% de células de Globet, la mucina sintetizada se localiza en grandes vacuolas citoplasmáticas, generando un desplazamiento del núcleo y una morfología típica de anillo de sello.

- **Carcinoma medular:** su característica principal es la aparición de células tumorales en láminas con rasgos morfológicos bien definidos (núcleo prominente, núcleo vesicular y gran citoplasma) y con grandes focos de infiltración linfocitaria intraepitelial. Es un subtipo histológico poco frecuente.

- **Carcinoma adenoescamoso:** subtipo poco frecuente de CCR donde se observan entremezcladas células típicas de ADC y de carcinoma de células escamosas.

- **Carcinoma indiferenciado:** es un subtipo con histología variable que engloba a aquellos tumores donde sus células han perdido las características de diferenciación morfológica típicas del tejido normal.

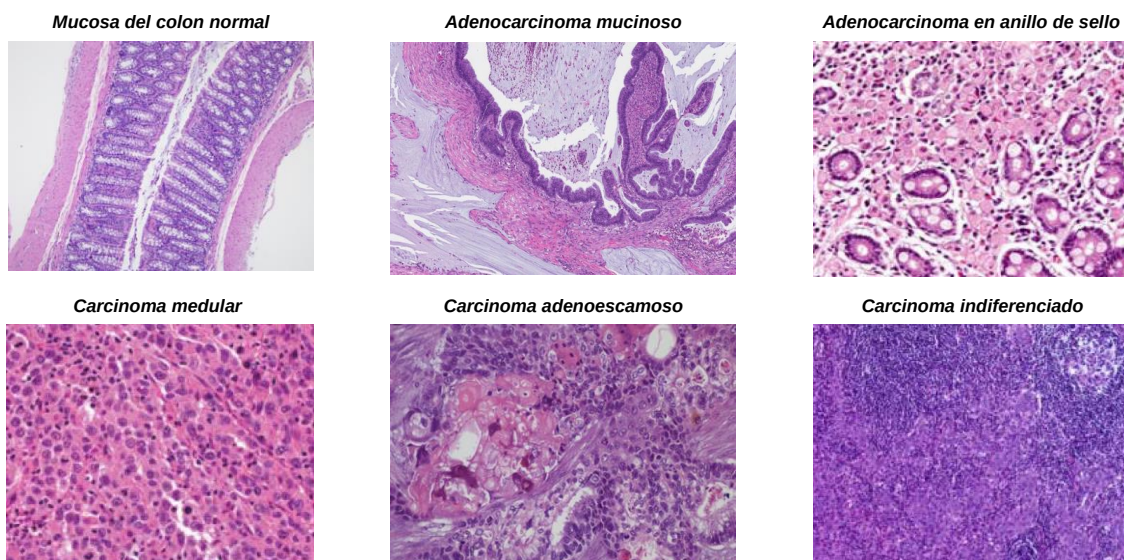


Figura 5. Histología del cáncer colorrectal. Se representan imágenes microscópicas tras tinción con hematoxilina/eosina de la mucosa del colon normal y sus variantes histopatológicas.

2.3. Tratamiento del CCR

Dados los avances en los tratamientos primarios y adyuvantes, el tiempo de supervivencia en el CCR ha ido mejorando. Por lo general, el tratamiento del CCR se basa en la extirpación completa del tumor y de las metástasis, lo que requiere en la mayoría de los casos una intervención quirúrgica³⁸. Sin embargo, a pesar de la mejora en los programas de cribado para reducir la incidencia del CCR, casi una cuarta parte se diagnostican en un estadio avanzado con metástasis. Además, el 20% de los casos restantes pueden desarrollar metástasis metacrónicas, lo que dificulta su seguimiento y el control quirúrgico^{39,40}. En el caso de los pacientes con lesiones no resecables o que no toleran la cirugía, el objetivo es reducir al máximo el tamaño del tumor y evitar su propagación y crecimiento. Para ello, la radioterapia y la quimioterapia son las principales estrategias terapéuticas. Cabe destacar que, en algunos casos, la quimioterapia o la radioterapia pueden aplicarse antes o después de la cirugía como tratamiento neoadyuvante o adyuvante para reducir y estabilizar el tumor⁴¹⁻⁴³.

2.3.1. Quimioterapia

La quimioterapia en el CCR incluye tanto la monoterapia como la terapia combinada, basadas principalmente en el tratamiento con 5-fluorouracilo (5-FU), oxaliplatino (OX), irinotecán (IRI) y capecitabina (CAP). Aunque hay estudios que sostienen que la monoterapia no es inferior al tratamiento combinado, este último sigue siendo la principal opción en primera línea con las combinaciones FOLFOX (5-FU+OX), FOXFIRI (5-FU+IRI), XELOX o CAPOX (CAP+OX) y CAPIRI (CAP+IRI)⁴⁴⁻⁴⁶. En las últimas décadas, se ha demostrado la eficacia de estos tratamientos en pacientes con CCR, especialmente los que presentan metástasis, donde se ha elevado la supervivencia global a casi 20 meses, por lo que la quimioterapia se ha convertido en la columna vertebral del tratamiento del CCR^{45,47,48}. Sin embargo, la quimioterapia genera cierta toxicidad sistémica, una tasa de respuesta insatisfactoria, una resistencia innata o adquirida, y una baja especificidad tumoral. Por este motivo, en los últimos años el tratamiento del CCR se ha perfeccionado con el uso de terapias dirigidas y de la inmunoterapia, donde el factor clave son las alteraciones moleculares y las mutaciones que caracterizan el tumor. En este sentido, se trabaja para conseguir un tratamiento de precisión de la enfermedad, con menor toxicidad y que aumente la respuesta y la supervivencia.

2.3.2. Terapia dirigida

Los primeros agentes dirigidos para el tratamiento del CCR fueron el cetuximab® y el bevacizumab®, aprobados en 2004. La terapia dirigida actúa frente a proteínas mediadoras de vías de señalización que median en el inicio, la progresión y diseminación del CCR, como la vía Wnt/ β -catenina, Notch, Hedgehog, TGF- β / SMAD, PI3K/AKT o la vía de RAS^{49,50}. En general, la terapia dirigida está basada en tratamientos con inhibidores del factor de crecimiento del endotelio vascular y su receptor (VEGF; VEGFR) o inhibidores del receptor del factor de crecimiento epidérmico (EGFR):

- **Terapia antiangiogénica (anti-VEGF/VEGFR):** tiene como objetivo detener la formación de nuevos vasos sanguíneos que sostengan al tumor y su diseminación. En este grupo se incluyen el bevacizumab®, regorafenib®, ziv-aflibercept® y ramucirumab®⁵¹.

- **Terapia anti-EGFR:** tiene como objetivo detener o retrasar el crecimiento del tumor. Estos fármacos son el cetuximab® y el panitumumab®. Sin embargo, estos tratamientos tienen baja eficacia frente a tumores con mutaciones en los genes *RAS* o *BRAF* (BRAFV600E). Frente a este último caso, se suele emplear una combinación de encorafenib® (inhibidor específico) y cetuximab®, normalmente en CCR metastásico⁵¹.

2.3.3. Inmunoterapia

La inmunoterapia está diseñada para estimular las defensas del paciente con el fin de neutralizar a las células tumorales, en base a los receptores que éstas presentan (PD-1 y CTLA-4) y que tratan de escapar del sistema inmunitario. Principalmente, la terapia se dirige a los puntos de control del sistema inmune (*immune checkpoint inhibitors*) para bloquear el escape del tumor de su detección por parte de las células T. La inmunoterapia frente al CCR comprende el uso de los siguientes anticuerpos monoclonales (mAbs):

- **Pembrolizumab®**: está dirigido a PD-1, un receptor de las células tumorales. Se utiliza para tratar el CCR no resecable o metastásico con inestabilidad de microsatélites (MSI) o con deficiencia en la reparación de los errores de emparejamiento (*mismatch repair deficiency*; dMMR), producidos durante la replicación del ácido desoxirribonucleico (DNA)⁵¹.

- **Nivolumab®**: está dirigido a PD-1 y se utiliza para el CCR metastásico con MSI o dMMR que no ha respondido a la quimioterapia convencional. Alternativamente, se puede utilizar la combinación nivolumab® (anti-PD-1) + ipilimumab® (anti-CTLA-4)⁵¹.

3. Etiología del CCR

Los factores implicados en el desarrollo del CCR se pueden dividir en los siguientes grupos:

3.1. Factores de riesgo asociados a la historia médica personal

Algunas enfermedades que afectan al tracto gastrointestinal inferior, como la enfermedad inflamatoria intestinal (enfermedad de Crohn y colitis ulcerosa), la presencia de pólipos en el colon o la diabetes, se asocian con un mayor riesgo de padecer CCR.

- **Enfermedad inflamatoria intestinal (EII)**: aunque la etiología de la EII es desconocida, se considera que su desarrollo facilita una inflamación intestinal crónica. Por ello, el riesgo de desarrollar CCR aumenta con la extensión anatómica, la gravedad y la duración de la EII^{52,53}.

- **Pólipos**: la presencia de pólipos también incrementa el riesgo de desarrollar CCR. Los pólipos adenomatosos de carácter vellosos y los adenomas túbulo-vellosos presentan el mayor poder de malignización. Además, aquellos pólipos mayores de 2 cm o la presencia de múltiples pólipos presentan un alto potencial de malignización⁵⁴.

- **Diabetes Mellitus**: los individuos con diabetes tipo II tienen un riesgo 2-3 veces mayor de desarrollar CCR en comparación con la población no diabética^{55,56}. En este caso, se cree que el desarrollo del CCR está vinculado a un aumento de la concentración de insulina y del componente inflamatorio que se asocia a la diabetes.

3.2. Factores de riesgo hereditario

Si bien es cierto que la mayoría de los casos de CCR son esporádicos, la existencia de antecedentes familiares aumenta significativamente el riesgo de desarrollarlo. En este sentido, se considera CCR familiar cuando se presentan dos o más familiares de primer grado afectados por este tumor. Este fenómeno comparte tanto la predisposición genética heredada (5%) como los factores relacionados con el estilo de vida¹⁹. Asimismo, el riesgo aumenta en aquellos individuos de familias con síndromes hereditarios, como la poliposis adenomatosa familiar o el síndrome de Lynch⁵⁷.

3.3. Factores de riesgo biológicos

3.3.1. Microbiota intestinal

En las personas sanas, el microbioma intestinal participa en el mantenimiento de la integridad de la barrera intestinal, protege contra los patógenos y desempeña un papel importante en la inmunomodulación. Recientemente, se ha descrito que la alteración en la composición y funcionalidad del microbioma intestinal establece un microambiente favorable para la iniciación y progresión del CCR⁵⁸⁻⁶².

3.3.2. Edad

Dado que aproximadamente el 90% de los nuevos casos de CCR se producen en individuos mayores de 50 años, la edad avanzada se considera uno de los factores de riesgo más significativos^{19,52}. La edad media en el momento del diagnóstico es de 66 años⁶³. El hecho de que el cáncer colorrectal sea una enfermedad relacionada con la edad explica su elevada incidencia en países con mayores esperanzas de vida⁶⁴. No obstante, cabe destacar que en los últimos 40 años la incidencia del cáncer colorrectal de aparición temprana (individuos menores de 50 años) ha aumentado considerablemente en Estados Unidos y Europa^{65,66}.

3.3.3. Género y raza

Según la Sociedad Americana del Cáncer (ACS), los hombres tienen un 30% más de riesgo de desarrollar CCR que las mujeres, así como un peor pronóstico y una mortalidad aproximadamente un 40% mayor. Sin embargo, las mujeres son más propensas a desarrollar cáncer de colon derecho, que suele diagnosticarse en un estadio más avanzado y parece ser más agresivo que los tumores del lado izquierdo⁶⁷. La incidencia del CCR

también varía sustancialmente según la raza. Los individuos negros no hispanos representan las incidencias más altas de todos los grupos raciales. Este grupo tiene una incidencia un 50% más alta que los asiáticos y un 20% más alta que los blancos no hispanos⁶⁸.

3.4. Factores de riesgo asociados al estilo de vida

3.4.1. Dieta

La IARC ha clasificado la dieta rica en carne roja y procesada como probablemente cancerígena para el ser humano. Existen estudios que demuestran que el consumo regular de este tipo de alimentos incrementa el riesgo de padecer CCR en un 18%, ya que contribuye a la obesidad y a la resistencia a la insulina^{69,70}. Por otro lado, se ha demostrado que las dietas ricas en fibra, fruta y verduras disminuyen el riesgo de desarrollar CCR en un 50% gracias a las propiedades antioxidantes y antiinflamatorias de estos alimentos⁵². También, el consumo elevado de productos lácteos como la leche está asociado a un menor riesgo de padecer esta enfermedad. Este hecho se ha atribuido especialmente al contenido en calcio que impide la modificación de la mucosa intestinal por parte de los ácidos biliares y de los ácidos grasos⁷¹.

3.4.2. Sobrepeso y obesidad

La acumulación anormal o excesiva de grasa, sobrepeso y obesidad, es un factor de riesgo importante para el desarrollo del CCR, incrementando el riesgo de padecerlo en un 50% y 20%, respectivamente. Se calcula que el riesgo global de CCR aumenta un 3% por cada cinco kilos de aumento de peso⁷². Se ha descrito que la acumulación anormal o excesiva de grasa provoca alteraciones en las secreciones de hormonas y citoquinas del tejido adiposo. Estos factores tienen efectos mitogénicos sobre las células epiteliales, inhiben la apoptosis de las células, promueven el estrés oxidativo, suprimen la respuesta inmunitaria y reducen la actividad de IGF-1, procesos que se asocian con el desarrollo y la progresión del cáncer^{73,74}.

3.4.3. Inactividad física

Los datos epidemiológicos indican que el aumento de la incidencia del CCR en los países desarrollados y en vías de desarrollo puede ser el resultado de un estilo de vida sedentario. Se estima que las personas físicamente inactivas tienen hasta un 50% más de riesgo de desarrollar CCR en comparación con las más activas físicamente. Se ha

demostrado que los ejercicios físicos regulares mejoran la función del sistema inmunitario, reducen la inflamación, disminuyen el estrés, optimizan la tasa metabólica, ayudan a regular el nivel hormonal y previenen la obesidad³⁹.

3.4.4. Tabaco y alcohol

Se ha descrito que las personas que fuman cigarrillos tienen un riesgo 2-3 veces mayor de desarrollar CCR, en comparación con los no fumadores, aumentando el riesgo con la dosis y la duración de la exposición⁷⁵. Además, se considera que al consumo de cigarrillos se le atribuye hasta el 12% de las muertes por cáncer colorrectal. Por su parte, el consumo de alcohol es otro factor que contribuye al desarrollo del CCR. Existe una correlación directa entre la cantidad de alcohol consumido y el riesgo de padecer CCR, llegando a existir un riesgo del 52% en aquellos individuos que toman más de cuatro bebidas al día¹⁹.

4. Síndromes de predisposición hereditaria al CCR

Las variantes patogénicas germinales de genes que confieren un riesgo alto o moderado de desarrollar cáncer se detectan en el 6-10% de todos los casos de CCR^{76–78}. El 20% de estas variantes se diagnostican antes de los 50 años. Los síndromes hereditarios de predisposición al CCR se dividen fenotípicamente en síndromes polipósicos y no polipósicos, basados en gran medida en el número y la histología de los pólipos colorrectales. Los síndromes más comunes son el síndrome de Lynch y la poliposis adenomatosa familiar (Figura 6).

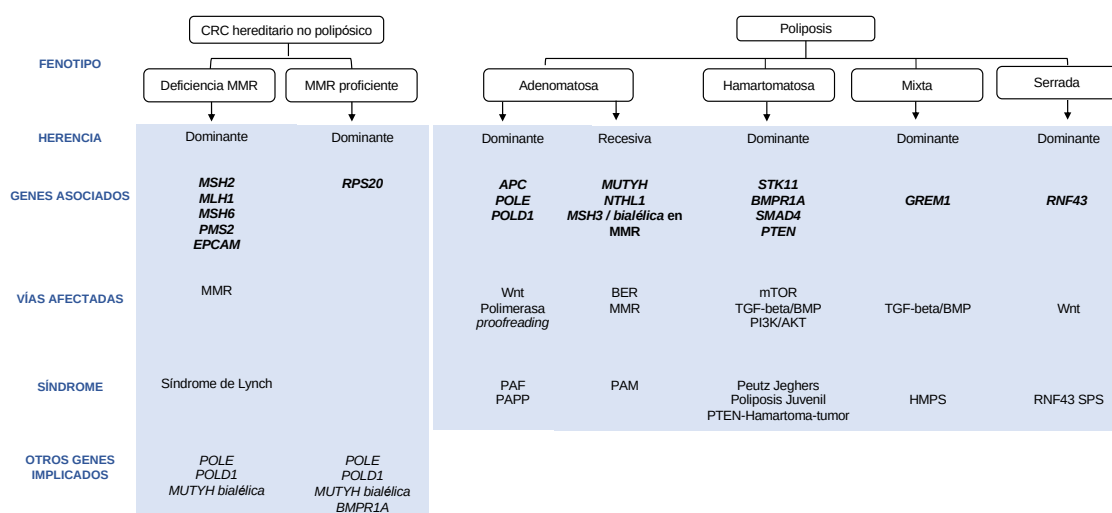


Figura 6. Clasificación fenotípica de los síndromes de predisposición hereditaria al CCR con y sin poliposis, modo de herencia, genes causales y vías moleculares afectadas.

4.1. CCR hereditario no polipósico (CCHNP) o síndrome de Lynch (SL)

El cáncer colorrectal hereditario no polipósico (CCHNP) o síndrome de Lynch (SL) constituye el síndrome hereditario más común que predispone al CCR, siendo el responsable del 3% de todos los nuevos diagnósticos⁷⁹. Incrementa el riesgo de padecer CCR en un 65-80% en varones y un 30-70% en mujeres. En estos pacientes, el CCR se produce a una edad más temprana (45-60 años) respecto de los pacientes que tienen CCR esporádico (69 años)⁸⁰, aunque la edad de aparición varía según el genotipo. El CCHNP se caracteriza por una aceleración de la carcinogénesis, donde la transformación de un adenoma en carcinoma tiene lugar en 2-3 años. Existen neoplasias extracolónicas asociadas al SL que incluyen el cáncer de endometrio (riesgo del 40-60%), ovario (9-12%), gástrico (2-13%), vías urinarias (4-6%) o tracto hepatobiliar (2%)⁸¹. Las variantes del SL incluyen el síndrome de Muir-Torre (SMT), que se caracteriza por adenomas sebáceos y otros tumores cutáneos (como queratoacantomas), y el síndrome de Turcot, que incluye el glioblastoma^{81,82}.

El fenotipo del SL se caracteriza por un predominio de los cánceres en el lado derecho del colon y una propensión al CCR sincrónico y metacrónico^{82,83}. Estos tumores suelen mostrar una diferenciación pobre con características mucinosas o un patrón de crecimiento medular, así como abundantes linfocitos infiltrados en el tumor^{84,85}. La hipermutación es el resultado de una deficiente reparación de los emparejamientos erróneos (MMR) de las bases del ADN, por mutaciones en los genes responsables de repararlos. Esto confiere una predisposición al CCR y a otros cánceres a lo largo de la vida. Los cánceres colorrectales con reparación deficiente de los emparejamientos erróneos se asocian a un estadio más temprano en el momento del diagnóstico y a un menor riesgo de desarrollar metástasis, en comparación con los tumores que no tienen defectos en genes de MMR.

El SL es una enfermedad hereditaria autosómica dominante con una penetrancia de las mutaciones en los genes afectados de hasta el 80%. Su diagnóstico se realiza mediante la identificación de mutaciones en línea germinal en alguno de los genes reparadores como *MLH1*, *MSH2*, *MSH6* o *PMS2*, o una delección en línea germinal de *EpCAM* (molécula de adhesión celular epitelial) que conduce a la inactivación epigenética de *MSH2*^{82,86}. Se ha descrito que las mutaciones en *MLH1* o *MSH2* son responsables del 60 al 80% de todos los cánceres asociados al SL. Las mutaciones en *MLH1* se asocian con un mayor riesgo de desarrollar CCR a una edad precoz, mientras que las mutaciones en *MSH2* incrementan el riesgo de desarrollar tumores extracolónicos, fundamentalmente cáncer de endometrio. Cuando uno de los genes de MMR muta en la línea germinal, existe una alta probabilidad de que se produzca una segunda mutación somática en el otro alelo de ese gen que

inhabilite la función reparadora y pueda provocar cáncer. Estos errores acumulados en el sistema MMR pueden afectar a regiones codificantes de genes asociados a la progresión tumoral, como *TGF β* , *IGFIIR*, *MSH3* o *BAX*. Además, esta reparación defectuosa conduce a la inestabilidad de los microsatélites (MSI) característica de muchos tipos tumorales⁸⁷.

4.2. Síndromes polipósicos

El CCR asociado a la poliposis hereditaria engloba a un conjunto de enfermedades con características variables. En los últimos años, se han descrito nuevos síndromes de predisposición asociados a la presencia de poliposis que se han detallado en la figura 6.

4.2.1. Poliposis adenomatosa familiar

La poliposis adenomatosa familiar (PAF) se caracteriza por la presencia de múltiples adenomas colorrectales que afecta por igual a varones y mujeres. Constituye la variante de poliposis más importante y suele aparecer a partir de los 20 años. La PAF está asociada a variantes patogénicas germinales en *APC*, un supresor tumoral que desempeña un papel importante en la regulación de la vía WNT. Aunque la PAF presenta una herencia autosómica dominante, aproximadamente el 30% de los individuos afectados no tienen antecedentes familiares y presentan mutaciones de *novo*⁵⁷. Los fenotipos son variables, algunos individuos presentan una poliposis clásica (de 100 a 1000 pólipos), mientras que otros pueden manifestar formas más sutiles (de 20 a 100 pólipos), a menudo denominadas poliposis adenomatosa familiar atenuada (PAFA). La PAFA se presenta en un 10% de los pacientes con PAF, y los pólipos o el CCR aparece unos 15 años después en comparación con la PAF⁸⁸. La mayoría de los individuos con PAF también desarrollan neoplasias en el tracto gastrointestinal superior, incluyendo pólipos en la glándula fúndica gástrica y adenomas duodenales. Estas manifestaciones incrementan el riesgo de cáncer papilar de tiroides o tumores desmoides. Los síndromes de poliposis adenomatosa se han actualizado recientemente con la inclusión de dos formas autosómicas recesivas raras causadas por mutaciones bialélicas en los genes *NTHL1*⁸⁹ y *MSH3*⁹⁰. Este último es un gen MMR no asociado al síndrome de Lynch.

4.2.2. Poliposis adenomatosa asociada a *MUTYH*

La poliposis adenomatosa asociada a *MUTYH* (PAM) es un síndrome autosómico recesivo asociado a variantes bialélicas germinales en el gen de reparación de escisión de bases *MUTYH*. Los individuos con MAP pueden mostrar una amplia gama de fenotipos, incluyendo la poliposis clásica y atenuada. Se ha descrito que las variantes monalélicas de

MUTYH están asociadas a un riesgo moderado (1,5-2 veces mayor) de padecer CCR, especialmente entre los individuos con una familiar de primer grado con CCR⁹¹.

4.2.3. Poliposis asociada a la actividad correctora de errores de la polimerasa

La poliposis asociada a la actividad correctora de errores de la polimerasa (PAPP) se asocia con variantes germinales patogénicas en el dominio exonucleasa de las polimerasas épsilon (*POLE*) y delta (*POLD1*)⁹². Los individuos pueden presentar poliposis clásica o atenuada autosómica dominante, CCR y otros tumores que presentan hipermutación somática, generalmente con fenotipos asociados a defectos en la reparación del DNA.

4.2.4. Síndrome de Peutz-Jeghers

El síndrome de Peutz-Jeghers se caracteriza por múltiples pólipos hamartomatosos en el tracto gastrointestinal y por un mayor riesgo de padecer varios tipos de cáncer, como el gastrointestinal (gástrico, colorrectal, pancreático), de mama, pulmón o el de cordón umbilical. Los individuos con este síndrome pueden presentar una prominente pigmentación mucocutánea y obstrucciones intestinales debido a invaginaciones de pólipos. Las variantes patogénicas germinales de *STK11* se identifican en el 50-70% de los individuos⁹³.

4.2.5. Síndrome de poliposis juvenil

El síndrome de poliposis juvenil (SPJ) se caracteriza por múltiples hamartomas gástricos y/o colónicos. Las variantes patogénicas germinales en *BMPR1A* y *SMAD4* se identifican en el 50-70% de los individuos afectados. El SPJ se asocia con un mayor riesgo de desarrollar cáncer gástrico y colorrectal. Los individuos con mutaciones en *SMAD4* tienen riesgo de padecer telangiectasia hemorrágica hereditaria (THH)⁹³.

4.2.6. Síndrome PTEN-Hamartoma-Tumor

El síndrome PTEN-Hamartoma-Tumor (SPHT) se asocia con una mayor predisposición al cáncer de mama, tiroides, endometrio y riñón como resultado de variantes patogénicas germinales en *PTEN*. El fenotipo gastrointestinal del SPHT puede incluir hamartomas gástricos y colorrectales, adenomas, pólipos dentados, pólipos hiperplásicos, lipomas y ganglioneuromas⁹⁴.

4.2.7. Síndrome de poliposis mixta hereditaria

La poliposis mixta hereditaria se caracteriza por la presencia de múltiples pólipos colorrectales de tipo histológico mixto, incluyendo lesiones dentadas, adenomas convencionales y hamartomas, y se asocia con un mayor riesgo de desarrollar CCR. Aunque la causa genética sigue siendo cuestionable en la mayoría de los casos, se han identificado duplicaciones en línea germinal del gen *GREM1* en algunos individuos afectados^{95,96}.

5. Carcinogénesis del CCR

El CCR esporádico constituye el subtipo genético más prevalente englobando más del 80% de los casos. La mayoría se desarrollan a partir de eventos mutagénicos que tienen lugar en la mucosa colónica normal. Este proceso comienza con una alteración inicial en las criptas intestinales que evoluciona a una lesión neoplásica precursora (un pólipo) y, finalmente, progresa a un adenocarcinoma en un periodo estimado de 10-15 años. La célula de origen en la mayoría de los CCRs es una célula madre o célula *stem* localizada en la base de las criptas intestinales^{97,98}. En estas células se produce una acumulación progresiva de alteraciones genéticas y epigenéticas que inactivan los genes supresores de tumores y activan los oncogenes permitiendo que se inicie la carcinogénesis colorrectal. Los primeros autores en describir un modelo de secuencia adenoma-carcinoma fueron Fearon y Vogelstein en 1980. Con este modelo, el epitelio del colon normal se transformaría en un adenoma que progresaría con los años hasta convertirse en un carcinoma con capacidad invasiva y metastásica⁹⁹ (Figura 7).

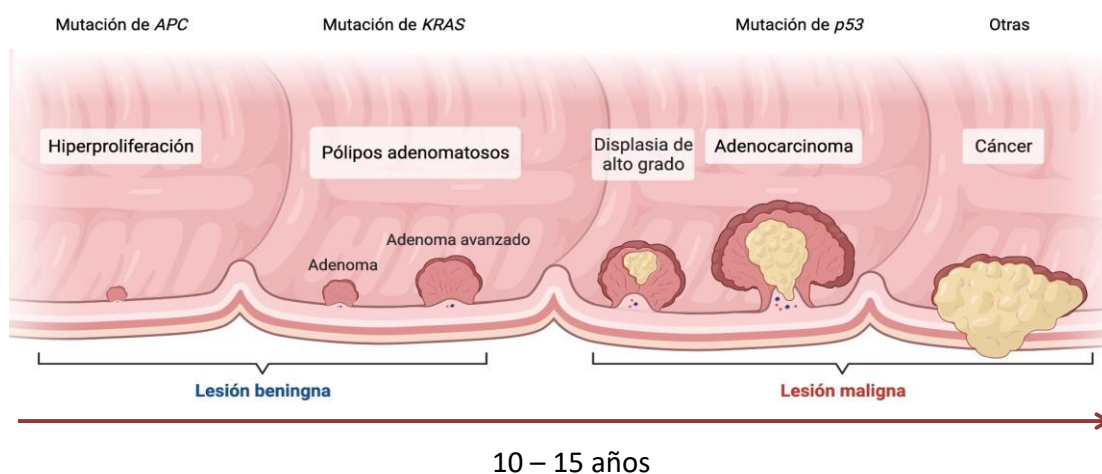


Figura 7. Modelo de secuencia adenoma-carcinoma propuesto por Fearon y Vogelstein.

Actualmente, existen tres vías principales que explican la carcinogénesis colorrectal: la inestabilidad cromosómica, la inestabilidad de microsatélites y el fenotipo metilador de islas CpG¹⁰⁰ (Figura 8).

5.1. Vía de inestabilidad cromosómica (CIN)

La inestabilidad cromosómica (CIN) es la vía más común, y es responsable del 70-80% de las transiciones que se dan desde el epitelio normal a adenocarcinoma en la génesis del CCR esporádico^{101,102}. La CIN resulta de defectos en la segregación cromosómica que generan aneuploidías, disfunciones teloméricas o defectos en los mecanismos de respuesta al daño en el DNA. La CIN se refiere a un aumento significativo de la ganancia o pérdida de la totalidad o de grandes porciones de cromosomas. Se caracteriza por la activación de oncogenes (*KRAS* y *BRAF*), la inactivación de genes supresores de tumores (*APC* y *TP53*), y una pérdida de heterocigosidad (LOH) en el brazo largo del cromosoma 18 (18qLOH), promoviendo así la tumorigénesis del CCR^{100,101}.

Según el modelo propuesto por Fearon y Vogelstein, el primer paso incluye el silenciamiento del gen *APC*, que impide la degradación de la β -catenina y conduce a la formación de un adenoma benigno en el epitelio intestinal. Seguidamente, se producen mutaciones del gen *KRAS* en el estadio adenomatoso. Esto genera un aumento de la señal proliferativa transformándose en un adenoma avanzado. Finalmente, la delección del cromosoma 18q (incluye genes como *SMAD2*, *SMAD4* o *DCC*) y la inactivación del gen *TP53* producen la transición hacia la malignidad^{99,103,104}. Además, recientemente se ha descrito que las alteraciones genéticas en *TGF- β* y *PI3KCA* están implicadas en el modelo adenoma-carcinoma^{105,106}. Además, las técnicas de hibridación genómica comparativa basada en arrays y en polimorfismos de nucleótido único (SNP) han permitido identificar pérdidas alélicas frecuentes en los brazos cromosómicos 1p, 5q, 8p, 17p, 18p, 18q, 20p y 22q. Por otro lado, se han identificados ganancias del cromosoma 7 y de los brazos cromosómicos 1q, 8q, 12q, 13q y 20q^{84,107-109}. Estas pérdidas y ganancias afectan a genes supresores de tumores y oncogenes, respectivamente, que facilitan la transformación de las células normales a cancerosas.

5.2. Vía de inestabilidad de microsatélites (MSI)

Otro tipo de inestabilidad genómica en el CCR es la inestabilidad de microsatélites (MSI) o vía mutadora, una característica distintiva de las células cancerosas. Los microsatélites son secuencias de nucleótidos cortas y repetitivas que se extienden a lo largo del genoma. Estas secuencias explican la alta tasa de errores que se da durante la

replicación del DNA. En estos casos, el sistema de reparación de errores del DNA (MMR) se encarga de reconocer y reparar estos desajustes generados durante la replicación. Los tumores se pueden clasificar en función de la presencia de marcadores STR (*short tandem repeat*). Así, se definen tumores con MSI-alta (MSI-H) si presentan 2 o más marcadores inestables; MSI-baja (MSI-L) si solo presentan un marcador inestable; o tumores estables (MSS) cuando no existe inestabilidad en ningún marcador.

La MSI es el sello distintivo de más del 95% de CCHNP o síndrome de Lynch, y sólo representa el 10-15% de todos los casos de CCR esporádico¹¹⁰. Los individuos con síndrome de Lynch desarrollan con frecuencia CCR-MSI debido a mutaciones en línea germinal en alguno de los genes MMR (*MLH1*, *MSH2*, *MSH6* o *PMS2*). Las mutaciones en el gen *MLH1* o *MSH2* conllevan un mayor riesgo (70-80%) de desarrollar cáncer, mientras que las mutaciones en el gen *MSH6* o *PMS2* tienen un riesgo comparativamente menor (25-60%)¹¹¹. La pérdida de función de estas proteínas ocasiona, además de la MSI, la acumulación de mutaciones somáticas en genes responsables del inicio y progresión de la enfermedad¹¹².

Por el contrario, los CCR-MSI esporádicos presentan con frecuencia una pérdida de la actividad MMR debido al silenciamiento del gen *MLH1* por metilación del ADN¹¹³. Los CCR-MSI esporádicos se correlacionan con la presencia de mutaciones en *BRAF* (*BRAFV600E*)^{114,115}. Clínicamente, el CCR con mutación de *BRAF* se correlaciona con un mal pronóstico y menor supervivencia global¹¹⁶. Además, se ha descrito que las mutaciones en *BRAF* asociadas a CCR presentan un peor pronóstico en comparación con las mutaciones en *RAS* (*KRAS* y *NRAS*)¹¹⁷.

5.3. Vía serrada o vía del fenotipo metilador de Islas CpG (CIMP)

La metilación del DNA consiste en la adición de un grupo metilo en el carbono 5 de la citosina mediante las enzimas DNA metiltransferasas, generándose una 5 metil-citosina. Esto tiene lugar en los dinucleótidos CpG, conocidas como islas CpG, que se localizan en la región promotora de los genes, lo que bloquea la transcripción de los mismos¹¹⁸. En las células tumorales, la hipermetilación dentro de la región promotora puede llevar a la inactivación de los genes supresores de tumores y de los implicados en la reparación del DNA. Por su parte, la hipometilación global se asocia a la inestabilidad genómica, aberraciones cromosómicas y estadios avanzados del cáncer¹¹⁹. Las alteraciones en el patrón de metilación pueden afectar a prácticamente todas vías de señalización, incluyendo la vía de TP53, TGF/SMAD, Wnt, NOTCH y receptores tirosina-quinasa (RTK), implicadas

en regulación del ciclo celular, transcripción, estabilidad del ADN, apoptosis, adhesión celular, angiogénesis, invasión y metástasis^{120,121}.

La vía del fenotipo metilador de islas CpG es la segunda vía más común en el CCR esporádico (15% de los casos). Los tumores CIMP suelen tener mutaciones en *BRAF* (*BRAFV600E*)¹²². El CIMP se subclasifica en base a la inestabilidad genética y epigenética integrada en CIMP-bajo (CIMP-L) y CIMP-alto (CIMP-H)^{123,124}. Así, los tumores CIMP-H se correlacionan significativamente con la edad, el sexo femenino, la localización proximal del colon y con mutaciones en *KRAS* y *BRAF*^{125,126}. Los marcadores CIMP más utilizados son *MLH1*, *p16*, *MINT1*, *MINT2* y *MINT31*¹²⁷.

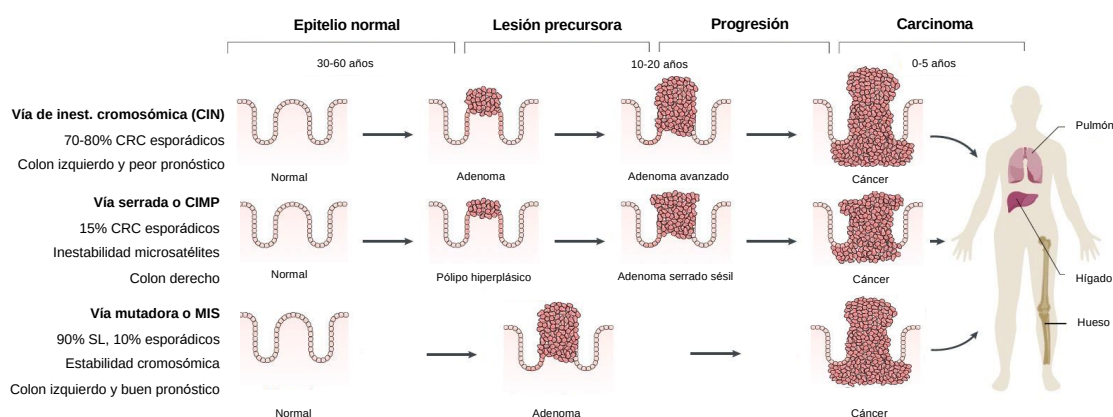


Figura 8. Vías de carcinogénesis del CCR. Adaptado de Keum *et al.* 2019.

5.4. Clasificación molecular del CCR

Aunque no existe una clasificación molecular estandarizada a nivel internacional, en 2015 se estableció el “*The Consensus Molecular Subtypes of Colorectal Cancer (CMS)*”¹²⁸ que clasifica la enfermedad en cuatro subtipos moleculares. Esta clasificación se basa en la integración de las tres vías de carcinogénesis atendiendo al estado de MSI y CIMP:

- **CMS1: CIMP-H / MSI-H:** presente en el 14% de los CCRs, con mayor frecuencia en el sexo femenino y localización en el colon proximal. Se suele diagnosticar en estadios avanzados y tienen peor pronóstico.
- **CMS2: CIMP-L / MSI-H:** abarca el 5% de los CCRs y se asocia al síndrome de Lynch. Tiene mayor localización proximal.
- **CMS3: CIMP-H / MSS:** presente en el 5-10% de los CCRs, se origina a partir de pólipos serrados y se asocia a mujeres ancianas con localización proximal.
- **CMS4: CIMP-L / MSS:** ocupa el 75% de los CCRs y se generan a partir de adenomas convencionales con presentación en el colon izquierdo. Estos pacientes tienen tasas de supervivencia más elevadas.

6. Cáncer colorrectal de aparición temprana (EOCRC)

La incidencia global del CCR ha disminuido en los últimos años gracias a la mejora de los programas de cribado y de las estrategias terapéuticas. Sin embargo, la incidencia en adultos menores de 50 años se ha incrementado de una forma alarmante en las últimas dos décadas. Además, se han evidenciado diferencias epidemiológicas, clínicas, genéticas, y biológicas que permiten definir el CCR de aparición temprana como una entidad diferente. Por ello, en función de la edad en el momento del diagnóstico, el CCR se puede clasificar en dos entidades: CCR de aparición temprana (*Early-Onset colorectal cancer*; EOCRC) y CCR de aparición tardía (*Late-Onset colorectal cancer*; LOCRC), con puntos de corte variables en la literatura, aunque normalmente situado en los 50 años¹²⁹.

6.1. Epidemiología del EOCRC

El CCR se considera una enfermedad del adulto mayor, con una edad media de aparición de 66 años, aunque puede afectar a toda la población⁶³. En los últimos 25 años, las tasas de EOCRC en Europa han aumentado drásticamente en pacientes de 20 a 49 años, con un predominio del cáncer de colon distal y de recto. Por ejemplo, las tasas de cáncer de recto aumentaron un 1,8% cada año desde 1990 a 2016, siendo un 3,5% en adultos de entre 20 y 29 años⁶⁶. Actualmente, esta tendencia se empieza a observar de forma global, incluyendo países como Estados Unidos, Australia o muchas regiones de Asia^{30,130}. Así, en este momento el EOCRC representa el 10% del total de diagnósticos de CCR, siendo para este grupo de edad el segundo tumor más común y la tercera causa de muerte por cáncer¹³¹. Para 2030, se estima que el 10% de los cánceres de colon y el 25% de los cánceres de recto se diagnosticarán en pacientes menores de 50 años³⁰. Las razones de este aumento desproporcionado se desconocen, en parte por la falta de datos epidemiológicos sólidos a nivel internacional y por el desconocimiento de los mecanismos moleculares implicados en la carcinogénesis del EOCRC.

6.2. Características clínico-patológicas del EOCRC

El panorama clínico-patológico y molecular del EOCRC muestra una considerable heterogeneidad, donde el estadio avanzado de la enfermedad, la presencia de metástasis tempranas y sus características histopatológicas desfavorables son definitorias en este grupo (Tabla 2).

Varios estudios poblacionales han mostrado una proporción preocupantemente alta de diagnósticos de EOCRC en estadios III y IV, que oscila entre el 54% y el 61,8%,

respectivamente^{132,133}. La baja sospecha de cáncer, el desconocimiento de la enfermedad, o la falta de reconocimiento de los síntomas preocupantes pueden llevar a un retraso en la evaluación de hasta 6 meses en comparación con el LOCRC¹³⁴. Estos datos ponen de manifiesto la importancia de considerar el CCR como un posible diagnóstico en adultos menores de 50 años (independientemente de los antecedentes familiares), así como trabajar en la concienciación de jóvenes y profesionales para garantizar un diagnóstico y una intervención oportuna. En cuanto a su localización, el EOCRC se localiza fundamentalmente en el colon distal y en el recto¹³⁵. Además, los pacientes jóvenes con CCR son más propensos a tener lesiones sincrónicas (en el momento del diagnóstico del tumor primario) y metacrónicas (durante el transcurso de la enfermedad)^{136,137}. A nivel histológico, el EOCRC presenta con mayor frecuencia características histopatológicas adversas: tumores con baja diferenciación y morfología mucinosa y/o de células en anillo de sello, lo que contribuye a una biología tumoral desfavorable, peor pronóstico, menor supervivencia y peor respuesta a la quimio-radioterapia neoadyuvante¹³⁸.

Tabla 2. Características clínico-patológicas y moleculares del CCR de aparición temprana y del CCR de aparición tardía.

Características clínicas y moleculares	EOCRC (<50 años)	LOCRC (>50 años)
Colon proximal		X
Colon distal y recto	X	
Etapas III y IV al diagnóstico	X	
Sincrónicos y metacrónicos	X	
Mucinoso/células en anillo de sello	X	
Diferenciación pobre	X	
Mayoritariamente MSS	X	
MSI por metilación de la región promotora del gen <i>MLH1</i>		X
CIN		X
CIMP (-)	X	
CIMP (+)		X
MSS y CIN (-): MACS	X	

6.3. Perfil molecular del EOCRC

Cada vez hay más pruebas que sugieren que el EOCRC puede representar una entidad diferente, definido por sus características biomoleculares específicas. Aunque la patogénesis del EOCRC está bien definida en individuos con SL, el 80% de los casos de CCR en este grupo son de origen esporádico, ya que no presentan mutaciones patogénicas en línea germinal en genes asociados a la predisposición de la enfermedad^{139,140}. Estos pacientes tienen un perfil molecular distinto que el CCR vinculado al SL¹⁴¹. En este sentido,

el EOCRC esporádico se podría asociar a un efecto acumulativo de alteraciones genéticas que muestran una penetrancia variable¹⁴². Sin embargo, el conocimiento de los mecanismos moleculares implicados en su génesis sigue siendo un reto, ya que es imprescindible para diseñar estrategias de prevención, diagnóstico y seguimiento más eficaces. Para ello, es necesario encontrar nuevos marcadores que puedan incluirse en los programas de cribado del CCR, con el objetivo de identificar a la población de riesgo de una manera precoz.

En este aspecto, la actividad de la telomerasa se ha postulado como un factor clave en el desarrollo del cáncer desde el punto de vista de la inmortalización celular, manteniendo el tamaño del telómero, o del envejecimiento y senescencia, acortando el mismo. Los telómeros cubren los extremos cromosómicos lineales para mantener su estabilidad, y se acortan en cada ciclo de la división celular. En las células sanas, la telomerasa pierde actividad y el acortamiento de los telómeros conduce finalmente a la senescencia celular y apoptosis. Sin embargo, en las células tumorales no se inhibe la actividad de la telomerasa y, en consecuencia, los telómeros no se acortan. Además, la división celular continuada provoca un acortamiento telomérico que puede conducir a la fusión de los extremos cromosómicos y a la CIN¹⁴³. Por tanto, este evento se considera como un proceso de envejecimiento asociado a la inestabilidad genética y a la oncogénesis. Por ejemplo, se ha demostrado que el acortamiento telomérico aumenta el riesgo de desarrollar algunos cánceres como el cáncer de cabeza y cuello, vejiga, renal, pulmón, esófago o CCR¹⁴⁴. Sin embargo, la relación entre la longitud telomérica (LT) y el riesgo de desarrollar EOCRC no ha sido bien documentada hasta el momento.

En el EOCRC, casi todos los tumores con MSI se asocian con el SL y rara vez presentan inactivación somática del gen *MLH1*¹⁴⁵. Los tumores EOCRC esporádicos con MSI están asociados mayormente a la inactivación del gen *MSH2*¹⁴⁶, mientras que la población mayor de 60 años con MSI presentan inactivación de *MLH1* (83%) y metilación de su región promotora (62%).

En general, los pacientes más jóvenes suelen tener tumores con estabilidad de microsatélites (MSS), hipometilación de *LINE-1* y mutaciones en los genes *TP53* y *CTNNB1*^{147,148}. Estos pacientes son menos propensos a albergar mutaciones en los genes *K-Ras*, *B-Raf* (BRAFV600E) y *APC*, o a mostrar metilaciones de islas CpG (CIMP-L)^{136,145,149}. La mayoría de los EOCRCs esporádicos (85%) que presentan estabilidad de microsatélites poseen un perfil distinto que el LOCRC-MSS, por lo que sigue siendo un subgrupo de tumores mal caracterizados. La ubicación del tumor en el colon distal, la baja frecuencia de otras neoplasias primarias y un componente familiar importante son, en este momento, las características más significativas del EOCRC-MSS¹⁵⁰. También, se ha

propuesto la existencia de un subgrupo específico de tumores de EOCRC, denominados MACs. Estos tumores presentan MSS, pero no tienen CIN ni pertenecen a la vía serrada de carcinogénesis. Abarcan el 30% de los EOCRC esporádicos y no se conocen los mecanismos implicados en su génesis¹⁵¹. Estos pacientes tienen tumores localizados en el colon distal y el recto, con características histológicas de pronóstico desfavorable, metástasis en el momento del diagnóstico, y muestran una recurrencia temprana de la enfermedad. Además, presentan una supervivencia inferior que los pacientes con MSI o CIN¹⁵².

6.4. Perfil cromosómico del EOCRC

La llegada de los estudios de hibridación genómica (*comparative genomic hybridization*) han permitido estudiar las alteraciones cromosómicas recurrentes que se dan en determinados tipos tumorales. Esto facilita la identificación de genes candidatos y de rutas de señalización afectadas, como por ejemplo las vías WNT/beta-catenina, RAS-MAPK, PI3K o TGF- β , frecuentemente alteradas en CRC¹⁴⁸. Estas vías, que parecen estar sobre-reguladas en la enfermedad de inicio temprano, desempeñan un papel fundamental en el inicio, progresión y metástasis del EOCRC.

Para el CCR, existen regiones cromosómicas con pérdida o ganancia de material genético que parecen estar afectadas de forma recurrente, por lo que se denominan “regiones recurrentemente alteradas” (RAR), las cuales podrían funcionar como marcadores pronósticos. En estas regiones se han descrito ganancias de material genético en 7p (que contiene el gen *EGFR*), 8q (*MYC*), 13q y 20q (*ZNF217*, *CYP24* y *AURKA*). Por otra parte, se han identificado pérdidas en 8p (contiene el gen *FGFR1*), 17p (*TP53*) y 18p/q (*SMAD4* y *DCC*)¹⁵³. Sin embargo, en el momento actual no existe una correlación clara entre el perfil clínico-patológico y molecular del EOCRC y las ganancias y pérdidas de material genómico, a diferencia de lo que ocurre en el LOCRC. Esto podría deberse a que los tumores de aparición temprana poseen diferentes mecanismos moleculares implicados en su génesis y a que conforman un subtipo distinto dentro de la vía CIN.

En un estudio realizado recientemente con la colaboración de nuestro grupo, se comparó el perfil cromosómico tumoral de una cohorte de pacientes diagnosticados de CCR primario que englobaba los dos subgrupos de edad: EOCRC y LOCRC. El análisis de alteraciones en el número de copias del DNA (CNAs) mostró que el 97,3% de todos los tumores presentaban inestabilidad genómica y que el subgrupo de pacientes jóvenes mostraba una mayor frecuencia de cambios recurrentes. Por un lado, se pudieron observar alteraciones recurrentes comunes a los dos grupos de edad, lo que indicaría que los genes

afectados confieren una ventaja adaptativa a la célula tumoral independiente de la edad de inicio de la enfermedad. Por otro lado, se identificaron alteraciones recurrentes presentes en los dos subgrupos, pero con frecuencias estadísticamente significativas entre ellos. En concreto, las ganancias en 1p35.2-p31.1 y 3p21.31-p21.1, y las pérdidas en 1p36, 1p12, 1q21, 9p13, 14q11, 16p13 y 16p12 son más frecuentes en los tumores colorrectales de inicio temprano, mientras que las ganancias en 7q11 y 7q22 predominan en los tumores de inicio tardío¹⁵⁴. Estas diferencias podrían explicar los mecanismos implicados en la carcinogénesis del EOCRC esporádico, su etapa de formación, y sus características clínico-patológicas y moleculares distintivas en función de la edad¹⁵⁵.

6.5. Región cromosómica 16p13.12-p13.11 en EOCRC

Interesantemente, los estudios de hibridación realizados por nuestro grupo revelaron una deleción recurrente en el brazo corto del cromosoma 16, en concreto en la región 16p13.12-p13.11, que se presentaba de forma más recurrente en pacientes jóvenes (33% EOCRC vs 16% LOCRC). La pérdida de esta pequeña región podría presentarse como una deleción única o junto a otras alteraciones cromosómicas¹⁵⁶. En concreto, se trata de un *hotspot* genómico rico en repeticiones de pequeñas secuencias de DNA (*low copy repeats*, LCRs), altamente homólogas, con alta probabilidad de presentar CNAs mediante recombinaciones homólogas no alélicas. Este tipo de recombinaciones podrían explicar la génesis de múltiples enfermedades como consecuencia de microdeleciones y microduplicaciones. Por ejemplo, las personas con microdeleciones en la región 16p13.11 pueden presentar retrasos en el desarrollo, microcefalia, estatura pequeña, estrabismo o discapacidad intelectual con o sin anomalías congénitas que afectan al riñón, huesos, dientes, dedos o corazón¹⁵⁷. Las personas con microduplicaciones en esta región son comunes en pacientes con esquizofrenia, epilepsia, autismo y trastorno por déficit de atención e hiperactividad¹⁵⁸.

Desde un punto de vista molecular, esta región cromosómica, recurrentemente alterada en EOCRC, abarca un conjunto de genes que se detallan en la tabla 3.

Tabla 3. Genes localizados en la región cromosómica 16p13.12-p13.11.

Gen	Trastorno asociado
<i>ABCC6</i>	Desarrollo de pseudoxantoma elástico (PXE)
<i>PDXDC1</i>	Esquizofrenia
<i>NTAN1</i>	Trastorno de la memoria y el aprendizaje
<i>BFAR</i>	Defectos en la apoptosis neuronal
<i>NPIPA2</i>	Desconocido
<i>PLA2G10</i>	Síndrome torácico agudo y enfermedad de células falciformes
<i>RRN3</i>	Desconocido
<i>PKD1P6</i>	Enfermedades hepáticas y renales poliquísticas
<i>NOMO1</i>	Desarrollo de pseudoxantoma elástico (PXE)

En un contexto de cáncer, el gen *NOMO1* (*Nodal Modulator 1*) cobra especial interés. Este gen codifica para una proteína que forma parte de un complejo proteico junto a Nicalina (NCLN) y TMEM147. Principalmente, este complejo actúa como antagonista de la vía de señalización Nodal durante el desarrollo embrionario¹⁵⁹, la cual también presenta una alta actividad en distintos tipos tumorales, entre ellos el cáncer colorrectal¹⁶⁰. Por este motivo, *NOMO1* podría postularse como un gen candidato *driver* en el desarrollo del EOCRC dada la recurrencia con la que se pierde la región cromosómica en la que se encuentra.

7. Papel de *NOMO1* en la carcinogénesis del EOCRC

Hasta la fecha, son relativamente pocos los estudios que han investigado el papel de *NOMO1* en carcinogénesis. Por ello, y dada la recurrencia con la que se pierde la región cromosómica 16p13.12-p13.11 en EOCRC, nuestro grupo se propuso estudiar el estado del gen *NOMO1* en una serie de pacientes con CCR de cualquier rango de edad.

En una primera aproximación, el 100% de los pacientes con EOCRC estudiados por RT-qPCR mostraban delección homocigota del gen *NOMO1* en el tumor, independientemente del estado de la región 16p13.12-p13.11. Sin embargo, en el subgrupo de pacientes con CCR de inicio tardío no existió ningún caso con delección de *NOMO1* cuando la región cromosómica estaba conservada. Además, en aquellos casos de LOCRC donde si existía alteración en 16p13.12-p13.11, el estado de *NOMO1* era variable, de forma que la delección homocigota del gen solo estuvo presente en el 11,7% de los casos¹⁵⁶. En una segunda aproximación, se amplió la serie de tumores colorrectales incluidos en el estudio hasta un total de 75 casos de EOCRC y 67 casos de LOCRC. Globalmente, un 81,3% de pacientes EOCRC presentaron delección del gen *NOMO1* en homocigosis, un 9,3% en heterocigosis y un 9,3% conservaban los dos alelos. En el grupo de LOCRC, únicamente el 4,5% de los pacientes mostraban la delección en homocigosis, un 13,4% en heterocigosis y un 82,1% no

tenían alteración de *NOMO1*. Además, los pacientes con EOCRC se dividieron según el fenotipo de inestabilidad de microsatélites. En este caso, la delección homocigota de *NOMO1* se observó en más del 90% de los tumores EOCRC-MSS, mientras que sólo estaba presente en el 43% de los tumores EOCRC-MSI. Sin embargo, la delección en heterocigosis era más recurrente en los pacientes jóvenes inestables de microsatélites (31,2% EOCRC-MSI vs 3,3% EOCRC-MSS)¹⁵⁶.

Por otro lado, el estado del gen *NOMO1* se analizó en DNA de sangre periférica de pacientes con CCR y en tejido polipoide. En el primer caso, *NOMO1* no mostró ninguna alteración en línea germinal, concluyendo que la pérdida de este gen es de carácter somático. En el segundo caso, se observó que el 42,5% de los pólipos presentaban delección heterocigota del gen. Así, el gen *NOMO1* podría perder una primera copia en un estadio inicial de la enfermedad (lesión precursora), de forma que una segunda mutación en el otro alelo facilitaría la progresión hacia carcinoma.

Por ello, estos datos preliminares permiten postular a *NOMO1* como un biomarcador de malignización asociado al EOCRC, particularmente del subtipo MSS.

7.1. Aspectos generales de *NOMO1*

NOMO (también conocida como pM5) es la proteína codificada por el gen *NOMO1*, el cual está localizado en la región afectada del brazo corto del cromosoma 16¹⁵⁹. Se trata de una proteína transmembrana de tipo I localizada en el retículo endoplasmático (RE), altamente conservada, y que tiene un tamaño de 130 KDa¹⁶¹. Está formada por un total de 1222 aminoácidos repartidos en tres dominios principales: un dominio extracelular, un dominio de transmembrana helicoidal y un dominio citoplasmático. En cuanto a su expresión, diferentes estudios han identificado niveles altos o moderados en tejidos como el cerebro, músculo esquelético, páncreas, pulmón, colon, hígado o estómago¹⁶². Aunque su función en tejido adulto se desconoce, esta proteína ha sido identificada como parte de un complejo proteico que actúa como antagonista de la vía Nodal, la cual es esencial para la especificación de la simetría corporal, y para la formación del mesodermo y endodermo durante el desarrollo embrionario^{163,164}.

7.2. Complejo NOMO/NCLN/TMEM147

Existen evidencias sobre la participación de NOMO en un complejo proteico localizado en la membrana del RE. Este complejo tiene una estructura similar al que forma la γ -secretasa, el cual es fundamental para el procesamiento proteolítico de varias proteínas

de transmembrana, como la proteína precursora β -amiloide asociada a la enfermedad del Alzheimer¹⁶⁵. Este complejo está formado por cuatro proteínas de transmembrana: Presenilina 1 (actividad enzimática), Nicastrina (proteína de transmembrana tipo I), PEN2 y A β -1¹⁶⁶.

En un primer acercamiento hacia la comprensión del complejo proteico al que pertenece NOMO, se identificó en el RE una proteína con una estructura muy similar a la Nicastrina, denominada Nicalina (NCLN, *Nicastrin like protein*). Esta proteína tiene 60 KDa y forma parte de un complejo proteico de alto peso molecular, independiente del complejo de la γ -secretasa, que presenta entre 200-220 KDa¹⁶¹. Diferentes estudios han afirmado que este complejo está formado por NOMO y NCLN, aunque no está claro la manera en la que se forma. También, se ha demostrado que NOMO y NCLN juegan un papel imprescindible en la formación del complejo, controlando de manera recíproca su expresión. Diversos estudios han corroborado la estrecha relación entre ambas proteínas, sugiriendo que son inestables cuando no están unidas entre sí, de forma que el complejo tiene un efecto estabilizador al igual que ocurre con los componentes del complejo γ -secretasa^{163,164}.

Dado el tamaño del complejo proteico identificado en el RE (200-220 KDa), y el tamaño que tienen individualmente las proteínas NOMO (130 KDa) y NCLN (60 KDa), parecía evidente que no eran los únicos componentes del complejo. En este sentido, se confirmó la presencia de una proteína de transmembrana de 25 KDa (TMEM147) que presenta una distribución en los tejidos humanos idéntica a la de las otras dos proteínas¹⁶³. Además, los niveles de expresión de TMEM147 parecen estar regulados también por NCLN, confirmando el papel como factor limitante en la formación del complejo de esta última. En cuanto a la formación de este complejo, primero se produciría la interacción entre NOMO y NCLN a través del dominio luminal y, posteriormente, se acoplaría TMEM147 al complejo a través del dominio transmembrana¹⁶⁴.

Aunque se desconoce a qué nivel de la ruta de señalización actúa el complejo NOMO/NCLN/TMEM147, se ha propuesto que actúa como antagonista de la vía de señalización Nodal durante el desarrollo embrionario. Sin embargo, se desconoce el papel que este complejo desempeña en el contexto de cáncer.

8. Vía de señalización Nodal

La vía de señalización Nodal desempeña funciones ancestrales bien conservadas y esenciales durante el desarrollo embrionario. Por ejemplo, interviene en la formación del mesodermo y endodermo, en el mantenimiento de la pluripotencia y renovación de las

células madre embrionarias, en la simetría del embrión a lo largo del tercer eje izquierda-derecha y en la promoción de la tumorigénesis^{160,167,168}.

Los ligandos nodales son miembros de la familia del factor de crecimiento transformador beta (TGF- β), el cual desempeña un papel fundamental en inflamación, inmunidad y desarrollo embrionario, regulando la expresión de varios factores de transcripción críticos, como Mix-like, GATA, Sox y Fox¹⁶⁹. La señalización Nodal/TGF- β se inicia por la unión del ligando Nodal a complejos hetero-multiméricos formados por receptores serina/treonina quinasa de tipo I y tipo II que activan a las proteínas SMADs. Se trata del receptor de la activina de tipo I ALK4/7 (*ACVR1B/ACVR1C*) y de tipo II ActRII-B (*ACVR2B*), y el correceptor Cripto-1 (*TDGF1*). La fosforilación y activación del complejo ALK4/7/ActRII-B permite la fosforilación de las proteínas citoplasmáticas SMAD2 y/o SMAD3. La activación de estas permite su interacción con SMAD4, dando lugar a la formación de un complejo transcripcional que se transloca al núcleo para activar a otros reguladores transcripcionales como FOXH1, TP53 y MIXER^{160,170} (Figura 9).

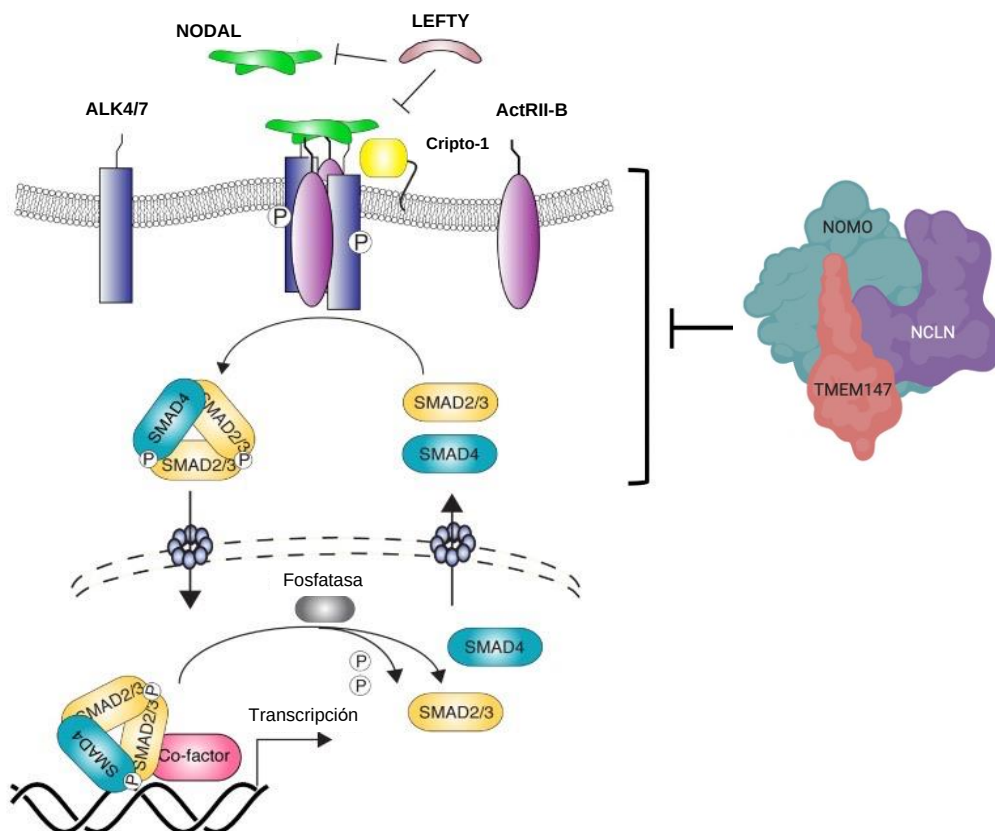


Figura 9. Esquema de la vía de señalización Nodal y complejo antagonista NOMO/NCLN/TMEM147.

Para la formación del complejo hetero-multimérico, es necesario una activación previa mediada por el correceptor Cripto-1. Así, Cripto-1 debe unirse previamente a ALK4 para su activación y posterior reclutamiento de ActRII-B. Además, Cripto-1 participa en la activación de otras vías de señalización implicadas en carcinogénesis como la vía RAS/RAF/MAPK y la vía PI3K/AKT¹⁷¹. Las proteínas que actúan como correceptores de la vía Nodal confieren especificidad a los ligandos nodales para unirse al complejo receptor. En ausencia de correceptores, los ligandos nodales no podrían unirse y por lo tanto la vía no podría activarse¹⁶⁴.

En cuanto a la regulación de la vía Nodal, existen inhibidores específicos como la proteína extracelular Lefty, la proteína de membrana Tomorregulina1 o la proteína DRAP. Lefty se encarga de antagonizar la vía Nodal a través de su interacción con los correceptores o con los ligandos nodales, de manera que impide la formación del complejo receptor¹⁶⁷. Por su parte, la Tomorregulina1 antagoniza la vía uniéndose directamente al correceptor¹⁷², mientras que DRAP lo hace interaccionando con el factor de transcripción FOXH1, previniendo así su unión al DNA¹⁷³. En último lugar, a pesar de que se desconoce a qué nivel de la ruta actúa, el complejo antagonista NOMO/NCLN/TMEM147 ha mostrado evidencias de que su pérdida de función genera fenotipos muy similares a los descritos para mutantes de la vía de señalización Nodal¹⁶⁴.

8.1. Papel de la vía Nodal en CCR

Hoy en día, está bien establecido en la comunidad investigadora que el cáncer es una patología heterogénea y que dentro de las células tumorales suele haber una subpoblación denominada células madre cancerosas (*Cancer Stem Cells*; CSC). Las CSC embrionarias poseen un nicho independiente que las protege, ofrecen quimio-resistencia y tienen capacidad de autorrenovación. Se cree que son las células iniciadoras del cáncer, así como el origen de las recaídas. Otra característica difundida en las CSC, pero no limitada a ellas, es la reactivación de algunas vías de señalización embriogénicas que son fundamentales para las células madre "normales" ¹⁷⁴. Esto suele deberse a la expresión aberrante de algunas proteínas implicadas en vías de señalización que suelen estar activas en las células embrionarias. La reactivación de estas vías permite que las células cancerosas adquieran un fenotipo más agresivo, acelerando la transición epitelio-mesénquima (EMT) y, por tanto, la invasión y la metástasis.

La vía Nodal presenta un papel importante en el mantenimiento de las CSC y en la carcinogénesis. Así, se ha descrito que un incremento en los niveles de expresión de las proteínas que actúan como ligandos de la vía favorece la progresión del melanoma maligno,

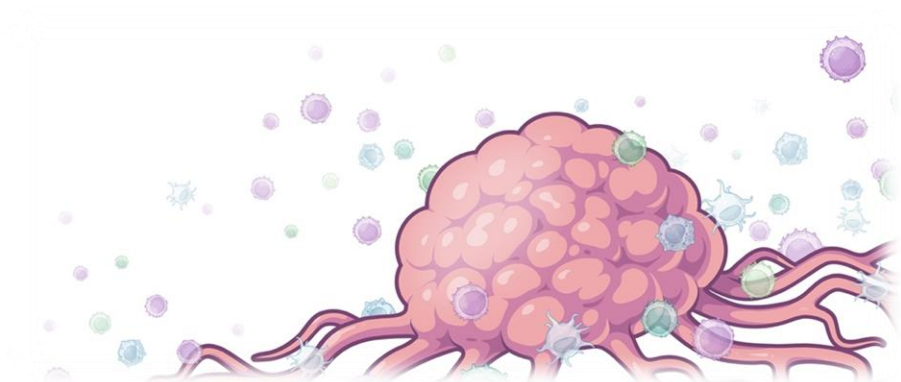
mientras que una inhibición de la vía disminuye la tumorigénesis¹⁷⁵. Aunque la expresión de la mayoría de los genes que intervienen en la vía Nodal es infrecuente, existen evidencias de una alta actividad de la vía en determinados tipos tumorales¹⁷⁶. Por ejemplo, la vía Nodal está implicada en la vascularización del cáncer de mama y en la aparición de metástasis en tumores como el melanoma, colon, ovario, próstata, endometrio y páncreas.

Numerosos estudios han encontrado que los ligandos nodales y el correceptor Cripto-1 están constitutivamente sobreexpresados en varios cánceres humanos, donde su alta expresión está fuertemente asociada con un peor pronóstico^{177,178}. Entre ellos, destaca el cáncer colorrectal, el cáncer de mama, ovario, gástrico o pulmón¹⁷¹. Concretamente, Cripto-1 realiza funciones de regulación relacionadas con el desarrollo embrionario, el mantenimiento de la pluripotencia y la capacidad de autorrenovación de las células madre embrionarias. Dada su baja expresión en tejido sano, Cripto-1 se ha postulado junto a otros intermediarios de la vía Nodal como posible diana terapéutica en distintos tipos tumorales¹⁶⁰.

En el caso del CCR, se ha demostrado en líneas celulares de cáncer de colon y en tejido tumoral humano que la sobreexpresión de los ligandos nodales y sus receptores es significativamente mayor, por lo que se puede asociar la ganancia de función de estas proteínas con la progresión del CCR¹⁷⁹. Especialmente, se ha observado un aumento de la expresión de Cripto-1 en lesiones premalignas como adenomas, metaplasias intestinales o metaplasias de la mucosa gástrica, que se va incrementando durante la progresión hacia carcinoma. Algunos estudios en modelos animales han tratado de dilucidar los mecanismos por los cuales la vía Nodal facilita la progresión del CCR y sugieren que la activación de Cripto-1 puede promover la transición epitelio-mesénquima¹⁸⁰. Además, se han estudiado otros mecanismos de señalización en los que interviene Cripto-1, como su interacción con la chaperona GRP78, un regulador de la homeostasis del RE que se expresa en las células tumorales. En CCR, existe una expresión diferencial de GRP78 entre el tejido normal y tumoral¹⁸¹.

Con todo esto, tanto el complejo antagonista NOMO/NCLN/TMEM147 como la desregulación de los intermediarios de la vía de señalización Nodal podrían tener un papel crucial en la génesis del cáncer colorrectal de aparición temprana.

Hipótesis



Hipótesis

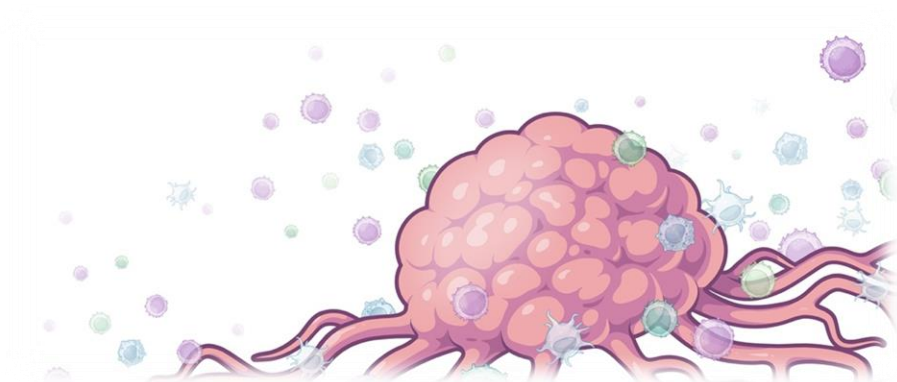
A nivel global, el cáncer colorrectal es el tercer tumor más frecuente y la segunda causa de muerte por cáncer, generando más de 900.000 fallecimientos cada año. El CCR se considera una enfermedad del adulto mayor donde la edad es el factor más influyente para su aparición, en consonancia con la mayoría de las enfermedades neoplásicas. Sin embargo, aunque la incidencia del CCR está en aumento debido al envejecimiento poblacional, la mortalidad por este tipo de tumor ha ido disminuyendo. Esto se debe a la implementación y mejora de los programas de cribado para el diagnóstico precoz a los que se someten las personas mayores de 50 años, así como a la aparición de estrategias terapéuticas más eficaces.

En cambio, la incidencia del cáncer colorrectal en pacientes menores de 50 años (EOCRC) ha aumentado considerablemente en las últimas décadas, representando en el momento actual el 10% del total de casos de CCR. Así, el EOCRC esporádico (75-80% de todos los casos de EOCRC) presenta importantes limitaciones, tanto para su diagnóstico precoz como para la identificación de biomarcadores específicos que permitan diagnosticar, pronosticar y diseñar planes de tratamiento eficientes. Este subgrupo de pacientes se caracteriza por un diagnóstico de la enfermedad en estadio avanzado, peor pronóstico y por presentar fenotipos más agresivos del tumor. Sin embargo, a diferencia del CCR de aparición tardía, las bases genéticas y los mecanismos de carcinogénesis implicados en el EOCRC son, hoy en día, una incógnita por resolver.

Un estudio preliminar de nuestro grupo, con el objetivo de identificar diferencias moleculares entre el CCR de aparición temprana y tardía, identificó una delección recurrente en la región cromosómica 16p13.12-p13.11 en pacientes con EOCRC esporádico. Dentro de esta región se localiza el gen *NOMO1*, delecionado en homocigosis de forma somática en más del 80% de los casos de EOCRC, independientemente del estado de la región 16p¹⁵⁶. Además, la pérdida en heterocigosis del gen *NOMO1* también fue identificada en lesiones benignas precursoras de la enfermedad (pólipos).

Con todo esto, la delección del gen *NOMO1* se posiciona como una alteración recurrente en EOCRC, aunque se desconocen los mecanismos moleculares alterados en tumores con este perfil genómico. Por ello, en este trabajo de tesis doctoral se hipotetiza que la pérdida del gen *NOMO1* podría ser un evento precoz en el desarrollo del EOCRC, por lo que podría utilizarse como un biomarcador de malignización.

Objetivos



Objetivos

Vistos estos antecedentes, en este trabajo se propusieron los siguientes objetivos con el fin de identificar los mecanismos moleculares que podrían estar alterados en aquellos tumores con inactivación de *NOMO1*:

1. Caracterizar funcionalmente la pérdida de *NOMO1* en cáncer colorrectal de aparición temprana (EOCRC).

- 1.1 Generar distintas líneas celulares *knockout* (KO) para *NOMO1* mediante el sistema de edición génica CRISPR/cas9.

- 1.2 Caracterizar el fenotipo de las líneas celulares *NOMO1-KO*.

- 1.3 Identificar posibles rutas de señalización alteradas tras la inactivación de *NOMO1* mediante el análisis del perfil transcriptómico y proteómico de las líneas celulares carentes de este gen.

2. Determinar si la pérdida de *Nomo1* es una mutación *driver* o *passenger* en EOCRC.

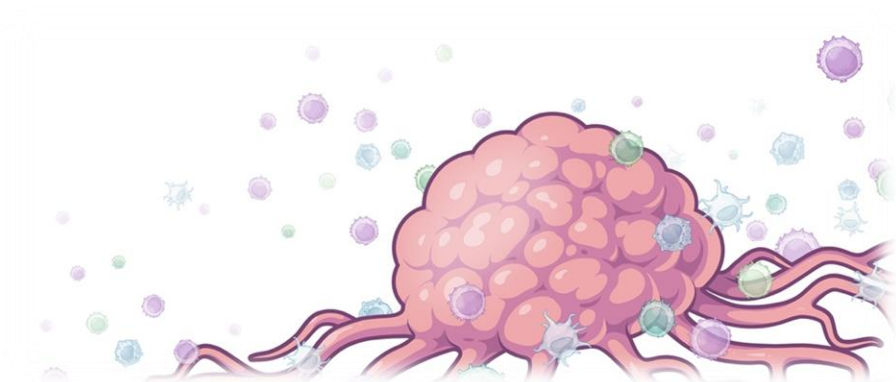
- 2.1 Generar un modelo murino condicional para *Nomo1* mediante la herramienta de edición génica CRISPR/cas9.

- 2.2 Realizar un análisis anatomopatológico de los modelos murinos con inactivación de *Nomo1*.

3. Analizar la sensibilidad al tratamiento con 5-fluorouracilo, irinotecán, oxaliplatino y cisplatino de las líneas celulares carentes de *NOMO1*.

4. Estudiar la longitud telomérica y analizar su implicación en EOCRC.

Material y métodos



Material y métodos

1. Selección de pacientes

Para el estudio de LT, se seleccionaron un total de 96 casos consecutivos de EOCRC, de los que se excluyeron los casos con mutaciones patogénicas en alguno de los genes de susceptibilidad al CCR, quedando la muestra definitiva del estudio en 87 pacientes con EOCRC. Los casos se seleccionaron del estudio multicéntrico prospectivo definido como *Spanish Early-Onset Colorectal Cancer Consortium* (SECOC)¹⁸². Todos los pacientes fueron diagnosticados antes de los 50 años, sin antecedentes de EII, y con diagnóstico histopatológico de adenocarcinoma. Las características clínico-patológicas se determinaron a partir de una revisión detallada de las historias clínicas. Las variables incluyeron el sexo, la edad en el momento del diagnóstico de CCR, el índice de masa corporal (IMC), el estadio y localización del tumor en el momento del diagnóstico, las características histológicas (grado de diferenciación, componente mucinoso, células en "anillo de sello"), las neoplasias primarias múltiples (CCR sincrónico y metacrónico) y los antecedentes de cáncer familiar. Las muestras de sangre periférica (SP) de todos los pacientes incluidos en el estudio se recogieron antes de cualquier tipo de tratamiento quirúrgico, neoadyuvante (quimioterapia o radioterapia) o adyuvante.

2. Técnicas de manipulación de ácidos nucleicos

La extracción de ácidos nucleicos se llevó a cabo a partir de SP, líneas celulares, fibroblastos embrionarios de ratón (MEFs) y tejido fresco procedente de ratón.

2.1. Extracción de DNA a partir de sangre periférica

La extracción de DNA procedente de SP de pacientes con cáncer colorrectal se realizó a través del método tradicional de fenol-cloroformo. Estas muestras de DNA se utilizaron para el estudio de longitud telomérica. Para ello, se aislaron las células nucleadas mediante centrifugaciones repetidas y lisis osmótica de los eritrocitos con H₂O estéril. El pellet celular leucocitario se resuspendió en tampón Fornace (0,25M sacarosa, 50mM Tris-HCL pH 7,5, 25mM KCl, 5mM MgCl₂) y se centrifugó a 1500 rpm, durante 10 minutos a 4°C. El botón celular obtenido se resuspendió en los siguientes componentes: EDTA (ácido etilendiaminotetraacético) 10 mM pH 8,0, para posibilitar la inactivación de nucleasas; 50 µg/mL de proteinasa K (ApliChem), para degradar las proteínas; SDS (dodecilsulfato sódico) al 1% de concentración, para romper la membrana celular. Esta mezcla se incubó durante 16-24 horas a 55°C. 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), seguido de una centrifugación a 1800 rpm durante 10 minutos para separar las fases. El DNA, que se encuentra en la fase acuosa, se limpió de otros residuos celulares mezclándolo con CIAA y se centrifugó nuevamente para separar las fases. Tras esto, se recogió la fase acuosa y se añadió etanol absoluto frío para precipitar el DNA. En último lugar, se realizó un lavado con etanol al 70% para eliminar sales y restos proteicos, y se resuspendió el DNA en 200 μ L de ddH_2O estéril.

2.2. Extracción de DNA a partir de líneas celulares y MEFs

Para la extracción de DNA a partir de líneas celulares y de MEFs, las células se recogieron de las placas de cultivo utilizando 1,5 mL del agente de disociación TrypLE 1X (Gibco™), se lavaron con 5 mL de PBS (tampón fosfato salino) y se centrifugaron a 1200 rpm durante 5 minutos. El precipitado celular se resuspendió en tampón Fornace, al que se le añadió EDTA 10 mM pH 8,0 para posibilitar la inactivación de las nucleasas, 50 μ g/mL de proteinasa K (ApliChem) para degradar las proteínas, y SDS a una concentración final del 1% para romper las membranas celulares. La mezcla se incubó a 55°C durante 16-24 horas y se continuó de la misma manera que en la extracción de DNA procedente de sangre periférica.

2.3. Extracción de DNA a partir de tejido fresco de ratón

En el caso del tejido fresco procedente de ratón, principalmente de tubo digestivo y cola, se tomaron muestras de 100-200 mg y se homogenizaron en tampón Fornace con el homogeneizador *Polytron® System PT 1200 E* (Thermo Fisher Scientific). Una vez disgregado el tejido, se incubó en una solución de lisis (Tris pH 8,0 10 mM, EDTA pH 8,0 5 mM, NaCl 200 mM SDS 0,2% y proteinasa K) durante 12-16 horas a 55°C. Finalmente, se realizó la extracción y purificación de DNA mediante el método tradicional de fenol-cloroformo empleando el mismo procedimiento que en la extracción de DNA a partir de líneas celulares.

2.4. Extracción de RNA a partir de líneas celulares y MEFs

Para la extracción del ácido ribonucleico (RNA), las células se lavaron con PBS, se recogieron de las placas de cultivo utilizando el agente de disociación TrypLE 1X (Gibco) y se centrifugaron a 1200 rpm durante 5 minutos. El pellet celular se resuspendió en 1 mL de TRIzol (Life Technologies) por placa de 10 cm de diámetro. La solución se pasó a tubos nuevos (Eppendorf®), se añadió 200 μ L de cloroformo, se agitó en el vórtex durante 15

segundos y la mezcla se dejó a temperatura ambiente durante 10-15 minutos. A continuación, se centrifugó a 12000 g, durante 15 minutos, a 4°C. Se observaron tres fases diferenciadas: una fase inferior que contiene proteínas, una fase intermedia que contiene DNA, y una fase acuosa superior que contiene RNA. La fase acuosa superior se pasó a otro tubo y se añadieron 500 µL de isopropanol frío para precipitar el RNA. Esta mezcla se conservó a -80°C durante unas 12-16 horas para aumentar el rendimiento de la precipitación. Posteriormente, se centrifugó la mezcla a 12000 g, a 4°C, durante 10 minutos. Después, se aspiró el sobrenadante observándose el pellet de RNA en el fondo del tubo que fue lavado con 1 mL de etanol frío al 70% mediante una centrifugación a 12000 g, durante 10 minutos, a 4°C. Se eliminó el sobrenadante y se dejó secar el pellet de RNA durante 1 hora. Finalmente, se resuspendió en 40 µL de agua libre de RNasas.

2.5. Extracción de RNA a partir de tejido fresco de ratón

Para la extracción de RNA a partir de tejido fresco de ratón, se tomaron muestras de 100-200 mg y se homogenizaron en tampón Fornace con el homogeneizador *Polytron® System PT 1200 E*. El tejido disgregado se resuspendió en 1 mL de TRIzol y se pasó a un nuevo tubo, se añadieron 200 µL de cloroformo y se agitó en el vórtex durante 15 segundos. A partir de este punto, se procedió a su purificación y precipitación de la misma forma que en la extracción de RNA a partir de líneas celulares.

2.6. Cuantificación de ácidos nucleicos

La concentración de DNA y RNA procedente de sangre periférica, líneas celulares, MEFs y tejido fresco de ratón se determinó en un espectrofotómetro *NanoDrop® 2000* (Thermo Fisher Scientific) midiendo la absorbancia de la muestra a 260 nm. Así, el cociente entre las absorbancias A260/A280 permitió determinar el grado de pureza de las muestras, estableciendo el rango 1,8-2,0 como adecuado para su posterior manejo. En los casos donde el valor de pureza superó el límite inferior o superior establecido, la muestra no se consideraría apta para su uso debido a la contaminación por proteínas o solventes orgánicos. En ambos casos, el DNA y RNA fueron nuevamente purificados y cuantificados.

Para evitar una degradación progresiva o posible contaminación, las muestras de DNA y RNA se almacenaron en tubos Eppendorf® y se conservaron a -20°C y -80°C, respectivamente.

2.7. Amplificación de fragmentos de DNA mediante PCR

El DNA extraído de las líneas celulares, de los MEFs y del tejido fresco de ratón fue amplificado mediante la reacción de amplificación PCR (*polymerase chain reaction*) en un volumen final de reacción de 25 μ L. Para ello se utilizaron 12,5 μ L del sistema PCR *Master Mix* 2X (Promega) (22 mM Tris-HCl a PH 8,4, 55 mM KCl, 1,65 mM de $MgCl_2$, 220 μ M de dNTP y 22 unidades de Taq DNA polimerasa), 1 μ L de cada uno de los oligonucleótidos (*forward* y *reverse*) a una concentración de 0,5 μ M cada uno y 9,5 μ L de agua libre de nucleasas. A cada reacción se añadió 1 μ L de DNA molde a una concentración de 200 ng/mL para su amplificación. En todos los casos, se utilizó como control negativo un tubo de reacción con todos los componentes excepto DNA. Los oligonucleótidos específicos diseñados para la amplificación de la región de interés de nuestros genes en estudio se muestran en la tabla 4.

Tabla 4. Oligonucleótidos empleados para la amplificación de la región de interés del gen *NOMO1* mediante PCR: *NOMO1-humano* (*h-NOMO1*) y *Nomo1-ratón* (*m-Nomo1*).

Gen	Oligonucleótido <i>forward</i> (5'-3')	Oligonucleótido <i>reverse</i> (5'-3')	Tamaño amplicón
<i>h-NOMO1</i> (exón 3)	CAGTGCTCAGTACCATGTAG	GGGAGGAATACAAACCCTC	337 pb
<i>m-Nomo1</i> (exón 3)	AGGGAGAAAGAGGATCAGCAG	CCACACAGATTAGGACTCTACT	786 pb

Las reacciones de amplificación se llevaron a cabo en un termociclador (Life Technologies-Invitrogen) utilizando un programa de PCR convencional, adaptado a la temperatura de anillamiento específica de cada pareja de oligonucleótidos (Figura 10).

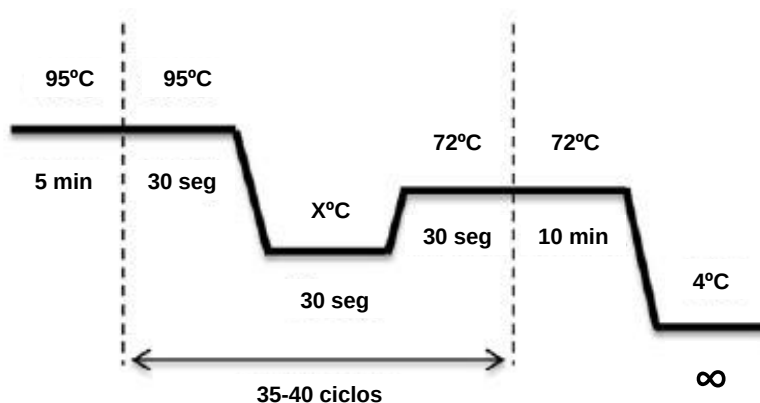


Figura 10. Programa de amplificación estándar empleado para la amplificación del fragmento de interés, donde X es la temperatura de anillamiento específica de cada pareja de oligonucleótidos.

2.8. Electroforesis en gel de agarosa

Los fragmentos de DNA amplificados se separaron por su tamaño molecular mediante una electroforesis horizontal en geles de agarosa al 2% en tampón TBE 0,5X (Tris 0,045 M, ácido bórico 0,045 M, EDTA 1,0 mM, pH 8,0). Las muestras migraron en paralelo con un marcador de peso molecular derivado del DNA del fago Φ X-174 cortado con la endonucleasa HaeIII, que genera amplicones con los siguientes tamaños: 1,353, 1,078, 0,872, 0,603, 0,310, 0,281, 0,271, 0,234, 0,194, 0,118 y 0,072 kb. El tampón de carga añadido a la muestra de DNA incorporaba dos colorantes que permitió monitorizar su migración en el gel: el xileno de cianol, que migra aproximadamente con los fragmentos de 5 kb en un gel de agarosa al 0,8%, y el azul de bromofenol, que migra junto a los fragmentos de aproximadamente 0,5 kb. La electroforesis se llevó a cabo durante 40 minutos con un potencial constante de 120 voltios. Posteriormente, los fragmentos de DNA amplificado se observaron gracias al componente SYBR® *Safe DNA Gel Stain* (1:10000) (Invitrogen, Life Technologies), que se intercala entre las bases nitrogenadas del DNA y emite fluorescencia al ser expuesto a la luz ultravioleta (UV) (254 nm). Los resultados obtenidos se capturaron gracias a un sistema de fotografía digital acoplado al software de tratamiento de imágenes KODAK *iD software v3.5.4* (Kodak).

2.9. Purificación de fragmentos de DNA amplificados mediante PCR

La purificación de los fragmentos de DNA obtenidos tras la amplificación se realizó utilizando el kit *PureLink® PCR Purification Kit* (Life Technologies-Invitrogen, California, U.S.A). Para ello, se mezcló el producto de PCR con 4 volúmenes de la solución “*Binding*”. Esta mezcla se pasó a una columna y se centrifugó a 10000 rpm, durante un minuto. Seguidamente, se limpió el DNA añadiendo a la misma columna 500 μ L de la solución “*wash buffer*” y se centrifugó a 10000 rpm, durante un minuto. Se decantó el filtrado y se volvió a centrifugar para secar la columna. Finalmente, el DNA adherido a la membrana se eluyó en 40 μ L de “*elution buffer*” con una última centrifugación de 10000 rpm, durante un minuto. La concentración y pureza del filtrado obtenido se cuantificó en un *NanoDrop®* y se conservó en un nuevo tubo Eppendorf® a -20°C.

2.10. Secuenciación mediante el método de Sanger

Se analizaron mediante secuenciación automática los productos de PCR purificados del gen *NOMO1* procedentes de las líneas celulares, de los MEFs y de las muestras primarias de ratón empleadas en este trabajo.

Para la secuenciación automática, se utilizó un total de 50-80 ng de DNA purificado con 3 pmol del oligonucleótido “*forward*” o “*reverse*” correspondiente en un volumen final de 8 µL. Para la reacción de secuenciación Sanger se utilizó el Kit *BigDye Terminator*® v.3.1. Este procedimiento se llevó a cabo en un secuenciador automático *ABI PRISM 377 DNA Sequencer* (Applied Biosystems, Inc.), perteneciente al Servicio de Secuenciación del ADN de la Universidad de Salamanca.

La lectura y el análisis de las secuencias automáticas se realizó con la ayuda del programa *SnapGene*® *Viewer v.5.2* (Figura 11). La homología con las secuencias de referencia depositadas en las bases de datos *GenBank* se realizó con el programa *BLAST* de los servidores <http://genome.ucsc.edu>.

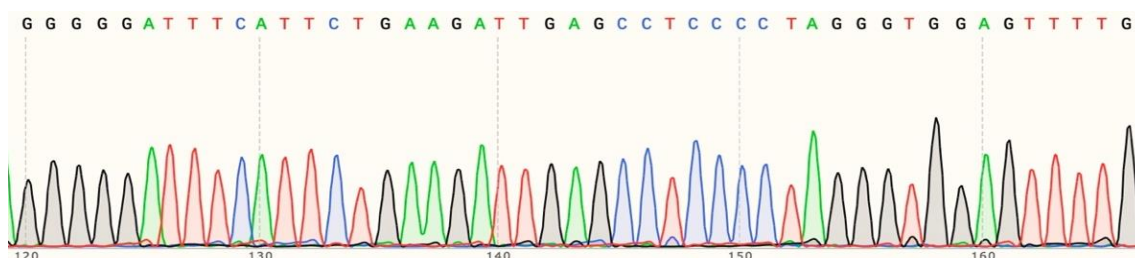


Figura 11. Secuencia correspondiente al exón 3 del gen *NOMO1* (46 pb).

2.11. Síntesis de DNA complementario (cDNA) mediante transcripción reversa (RT)

Para cuantificar la expresión génica fue necesario sintetizar el DNA complementario (cDNA) en un paso previo. En este caso, el material de partida fue RNA mensajero (mRNA), que necesita ser retro-transcrito a cDNA a través de una reacción que se conoce como transcripción reversa (RT). La RT es una técnica que permite sintetizar el cDNA a partir de una molécula de RNA de cadena simple utilizando la enzima transcriptasa reversa que genera una cadena simple y complementaria a la cadena molde de RNA (Figura 12).

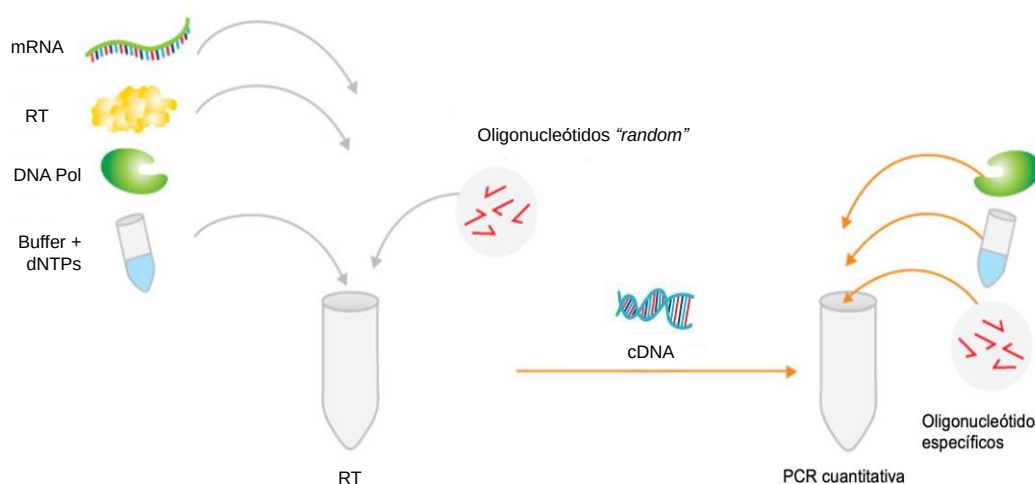


Figura 12. Esquema de síntesis de DNA complementario (cDNA) a partir de RNA mensajero (mRNA).

En este caso, se ha empleado el kit *High-Capacity cDNA Reverse Transcription Kit* (Applied Biosystems™). Para ello, se utilizó un máximo de 2 µg de RNA total en un volumen de 10 µL. A este volumen se añadió una mezcla de 10 µL compuesta por: 2 µL de 10X RT Buffer, 0,8 µL de 25X dNTP Mix (100 mM), 2 µL de 10X RT Random Primers, 1 µL de MultiScribe™ Reverse Transcriptase, 1 µL de RNase Inhibitor y 3,2 µL de H₂O. La preparación de esta reacción se hizo en hielo y se llevó a un termociclador para la reacción de síntesis con las condiciones que se detallan en la tabla 5. Finalmente, el cDNA obtenido se conservó a -20°C hasta su uso.

Tabla 5. Condiciones para la síntesis de cDNA mediante transcripción reversa (RT).

	Paso 1	Paso 2	Paso 3	Paso 4
Temperatura	25°C	37°C	85°C	4°C
Tiempo	10 min	120 min	5 min	∞

2.12. PCR cuantitativa-comparativa en tiempo real (RT-qPCR)

La PCR cuantitativa en tiempo real (RT-qPCR) es una variante de la PCR utilizada para amplificar y cuantificar de forma absoluta y simultánea un determinado fragmento de DNA¹⁸³. De esta manera, se puede detectar y evaluar la cinética de esta reacción de amplificación a lo largo de un determinado número de ciclos. Se fundamenta en la adaptación de un termociclador y un fluorímetro, que permite detectar la fluorescencia emitida por sondas fluorescentes o componentes intercalantes unidos al DNA. Así, la monitorización de la fluorescencia detectada permite determinar el número de copias del

fragmento de interés al inicio de la reacción y detectar variaciones tales como deleciones o amplificaciones.

Estos valores se pueden cuantificar con precisión a través del ciclo umbral (C_t), que indica el número de ciclos necesarios para que se produzca un aumento de fluorescencia diferencial (ΔR_n) respecto a la señal base de emisión y que es directamente proporcional a la cantidad inicial de DNA molde. A partir de este ciclo, comienza a producirse la amplificación de nuestro fragmento de interés, generándose una curva de amplificación (Figura 13) que se estabiliza conforme avanzan los ciclos, indicando la finalización de la reacción.

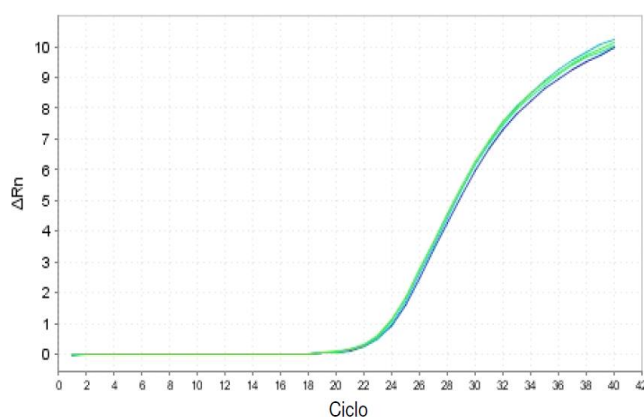


Figura 13. Curva de amplificación de DNA medida por el aumento de fluorescencia del agente intercalante.

La temperatura de fusión (T_m) es otro parámetro que permite caracterizar el producto amplificado y medir la especificidad de la reacción. Se representa en una curva de temperaturas, llamada curva de fusión (*Melt curve*) (Figura 14), que se genera al finalizar la etapa de extensión de la RT-qPCR, y donde T_m indica la temperatura a la cual se ha desnaturalizado el 50% de nuestro DNA genómico. El pico máximo de esta curva nos permite detectar la especificidad de la reacción en el sentido de no generar amplicones de otros fragmentos donde puedan anillar los oligonucleótidos. En este sentido, una curva con más de un “pico” en la curva de fusión indicaría una reacción inespecífica, donde estaríamos cuantificando DNA que no es de nuestro interés.

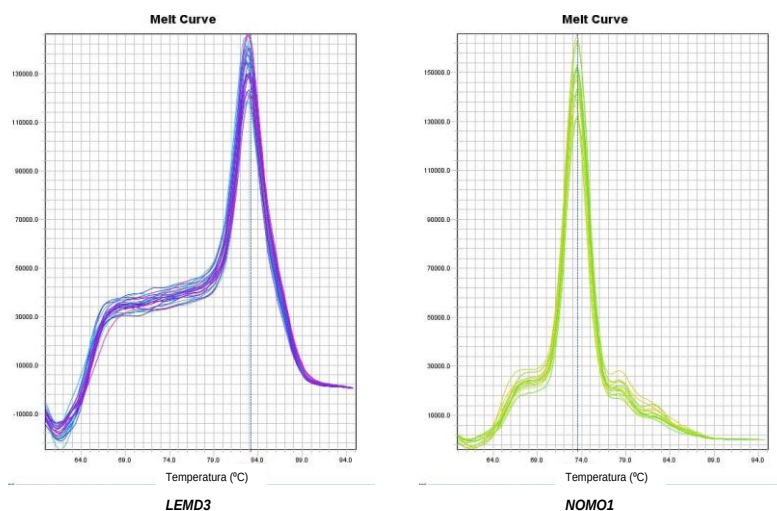


Figura 14. Representación gráfica de las curvas de fusión de los productos de amplificación para el control endógeno (*LEMD3*) y el fragmento de interés del gen *NOMO1*. En el eje de abscisas se representa el gradiente de temperatura y en el eje de ordenadas la fluorescencia detectada en unidades arbitrarias.

Para determinar, en las muestras en estudio, el número de copias exacto de nuestro fragmento de interés es necesario comparar su Ct con el Ct de una muestra control, de la cual conocemos el número de copias que posee. Estas reacciones de amplificación deben tener lugar de forma simultánea y en el mismo equipo. Para analizar los cambios relativos en la cuantificación de nuestro fragmento debidos a otros factores, se amplifica un fragmento de un gen de referencia o copia única (control endógeno) que no tiene variación en el número de copias en la patología a estudiar. Así, en este trabajo se ha utilizado el método de cuantificación relativa $\Delta\Delta C_t$, que determina los cambios relativos en la carga genética como pérdidas de heterocigosidad, deleciones o amplificaciones de nuestro gen de interés, normalizando con una muestra control y un control endógeno. En este caso, se ha utilizado como control endógeno un fragmento del exón 1 del gen *LEMD3* (Tabla 6).

Tabla 6. Oligonucleótidos utilizados en los ensayos de pérdida de heterocigosidad (LOH) mediante PCR cuantitativa-comparativa en tiempo real (RT-qPCR): *NOMO1* G1+G2 y *LEMD3*.

Gen	Oligonucleótido <i>forward</i> (5'-3')	Oligonucleótido <i>reverse</i> (5'-3')	Tamaño amplicón
<i>NOMO1</i> G1+G2 (exón 3)	CTTTTAGGGGGATTTCATTCTG	ACCAAACTCCACCCTAGGGG	52 pb
<i>LEMD3</i> (exón 1)	GCGGCTGCCGGGAGTCTAGACAG	GGTAGGTATGATTGGAGCCCG	343 pb

Las amplificaciones tuvieron lugar partiendo de una misma cantidad de DNA genómico al inicio de la reacción. Para ello, el DNA se cuantificó utilizando un *NanoDrop®* y se diluyó en ddH_2O hasta obtener una concentración final de 20 ng/ μL .

Cada pocillo de reacción tuvo un volumen final de 10 μL ajustado de la siguiente manera: se mezclaron 5 μL del agente intercalante SYBR® Green PCR Máster Mix (Roche, Basilea, Suiza), 0,5 μL de cada oligonucleótido, 4 μL de H_2O libre de nucleasas y 2 μL de DNA (40 ng) de la muestra correspondiente. El programa de amplificación necesario para cada fragmento de interés se ajusta en una puesta a punto, previa al estudio de las muestras (Figura 15).

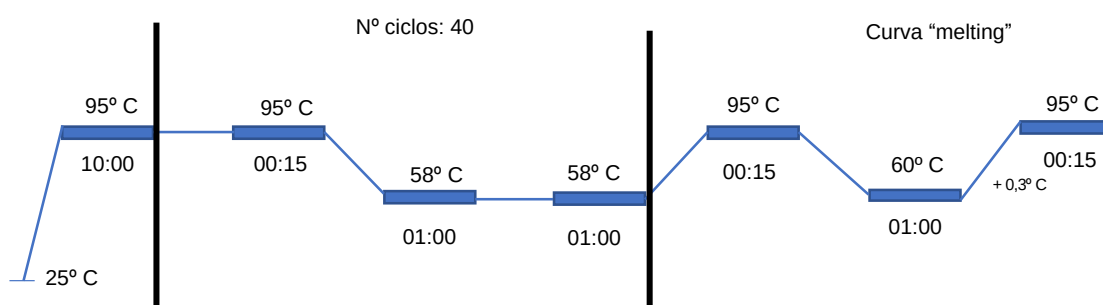


Figura 15. Representación gráfica del programa establecido en el termociclador para la cuantificación relativa del número de copias del fragmento de interés del gen *NOMO1* mediante RT-qPCR.

Cuando la muestra de partida es cDNA, obtenemos una cuantificación que indica los niveles de expresión relativa de nuestro fragmento de interés. En este caso, añadimos 200 ng totales de cDNA en un volumen de 2 μL a cada pocillo de reacción. Para los estudios de expresión de este trabajo, se ha utilizado como control endógeno el gen *GAPDH* (Tabla 7), un gen con expresión constante en el tipo de muestras estudiadas. Para el resto del procedimiento se procede de la misma manera que para la RT-qPCR en DNA genómico.

Tabla 7. Oligonucleótidos utilizados en los ensayos de expresión génica mediante PCR cuantitativa-comparativa en tiempo real (RT-qPCR).

Gen	Oligonucleótido <i>forward</i> (5'-3')	Oligonucleótido <i>reverse</i> (5'-3')	Tamaño amplicón
<i>h-NCLN</i> (exón 4-5)	AGTGACTGGCTGATTGCCA	CGTGCGCTTGTAGGTGT	204 pb
<i>h-TMEM147</i> (exón 1-3)	CAACGCCTTCTGGAAATGCG	TATCAGGTCTGCCACATCCA	160 pb
<i>m-Nomo1</i> (exón 3)	AGACGGACTGTGCTCCTA	CTTGCCATTACAGAGAAGC	200 pb

Todas las reacciones de RT-qPCR se llevaron a cabo 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.). Para cada muestra se cuantificaron, simultáneamente, el número de copias o la expresión relativa de los genes de interés y de los controles endógenos, según el caso. Para cada uno de ellos se establecieron 3 réplicas con el fin de minimizar la variabilidad en la concentración de partida de las muestras. Las placas se cubrieron con las láminas transparente ópticas *MicroAmp® Optical Adhesive Film* (Life Technologies-Invitrogen, California, U.S.A.). De esta manera, se evitó la evaporación de la muestra durante la reacción y se permitió la penetración del láser para detectar la fluorescencia emitida.

Se utilizó el equipo *StepOnePlus™ Real-Time PCR System* (Life Technologies-Invitrogen, California, U.S.A.) para llevar a cabo la qPCR. Se compone de un termociclador adaptado a un fluorímetro con varios canales de lectura de fluorescencia. El análisis de resultados se llevó a cabo con programa *StepOne software v.2.1* (Life Technologies-Invitrogen, California, U.S.A.).

2.13. Estudio de longitud telomérica mediante RT-qPCR

Para el estudio de LT en pacientes con EOCRC, se llevó a cabo una cuantificación absoluta mediante RT-qPCR empleando el kit *Absolute Human Telomere Length Quantification qPCR Assay Kit* (ScienCell™). El kit aporta los oligonucleótidos necesarios para amplificar la región de interés (TEL), el gen de referencia (SCR), y un DNA control cuya LT es conocida. Así, el oligonucleótido *TEL* reconoce y amplifica las secuencias teloméricas, mientras que el oligonucleótido *SCR* funciona como control endógeno. Este último reconoce y amplifica una secuencia de 100 pares de bases, localizada en el cromosoma 17 y que actúa como referencia para normalizar los datos. La normalización se llevó a cabo mediante el método de cuantificación relativa $\Delta\Delta C_t$. La longitud relativa de cada muestra en estudio se volvió a normalizar con la obtenida en la muestra control, obteniendo así la LT absoluta de cada paciente.

En este caso, cada pocillo de reacción tuvo un volumen final de 20 μL , ajustado de la siguiente manera: se mezclaron 10 μL del agente intercalante SYBR® Green PCR Máster Mix, 1 μL de cada oligonucleótido, 7 μL de H_2O libre de nucleasas, y 2 μL de DNA (10 ng) de la muestra correspondiente. El programa de amplificación necesario para cada fragmento de interés se detalla en la figura 16. Cada paciente fue estudiado por triplicado para *TEL* y por triplicado para *SCR*.

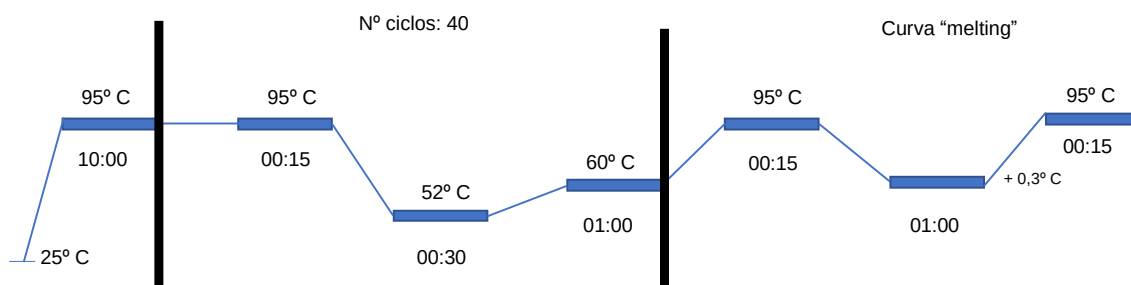


Figura 16. Representación gráfica del programa establecido en el termociclador para la cuantificación absoluta de la longitud telomérica de pacientes con EOCRC.

2.14. Discriminación alélica mediante PCR con sondas TaqMan®

Durante la discriminación alélica con sondas TaqMan®, la amplificación y detección del genotipo tiene lugar de forma simultánea. Las sondas TaqMan son oligonucleótidos que hibridan de forma específica con cada alelo, en la región donde se sitúa el polimorfismo. Estas sondas están marcadas con un fluorocromo donador en el extremo 5', que emite fluorescencia al ser excitado, y con un fluorocromo aceptor o “*quencher*” en el extremo 3', el cual absorbe la fluorescencia emitida por el donador mientras se encuentren próximos. Durante la amplificación de la secuencia diana, la sonda hibrida con su cadena complementaria y la DNA polimerasa, a través de su actividad 5'-3' exonucleasa, hidroliza el extremo 5' libre de la sonda y produce la liberación del fluorocromo donador. Así, la fluorescencia captada por el lector es proporcional a la cantidad de DNA amplificado, permitiendo en todo momento reconocer y registrar la cinética de la reacción.

En nuestro estudio, hemos utilizado una sonda específica para detectar el polimorfismo rs2853669 (A>G) del gen *hTERT*, el cual ha sido asociado con el acortamiento telomérico. Esta sonda estaba marcada con los fluorocromos VIC y FAM para cada alelo. Para llevar a cabo la reacción de amplificación mediante PCR, se utilizaron 40 ng de DNA, a los que se añadieron 5 µL del TaqMan® Universal PCR Master Mix. Para la detección de fluorescencia se empleó el mismo equipo que en la RT-qPCR.

2.15. Estudio del transcriptoma

Los cambios en los perfiles de expresión génica, generados por la inactivación de *NOMO1* en las líneas celulares empleadas en este trabajo, se estudiaron a través de dos aproximaciones: los microarrays de expresión (WT PLUS Affymetrix) y la secuenciación de RNA (RNA-seq). Como paso previo, se extrajo RNA total de 3 clones independientes *NOMO1-WT* y 3 clones independientes *NOMO1-KO* de las líneas celulares HCT-116 y HS-

5. Para ello, se utilizó el kit *RNeasy® Mini Kit* (Qiagen, Hilden, Alemania) siguiendo las instrucciones del fabricante. Durante este procedimiento también se realizó un tratamiento de digestión de DNA genómico empleando el set *RNase-Free DNase Set* (Qiagen, Hilden, Alemania). El RNA total extraído se cuantificó y se midió su calidad mediante un NanoDrop 2000.

2.15.1. Microarrays de expresión (WT PLUS Affymetrix)

El análisis del transcriptoma mediante array de expresión se llevó a cabo en el Servicio de Genómica del Centro de Investigación del Cáncer (CIC) de Salamanca.

En primer lugar, se evaluó la integridad del RNA total utilizando el bioanalizador Agilent 2100 (Agilent Technologies, CA, U.S.A). Posteriormente, las muestras fueron hibridadas y etiquetadas siguiendo el protocolo de la plataforma *Affymetrix PrimeView Human Gene Expression Array #901838* (ThermoFisher Scientific) que permite la lectura de más de 36000 transcritos por cada muestra. Tras el etiquetado, se amplificaron y marcaron 100 ng de RNA total de cada clon, utilizando el kit *GeneChip® WT PLUS Reagent Kit* de Affymetrix, y se hibridaron en *Clariom S Human Array*. Para el lavado y escaneo de las muestras se utilizó el sistema GeneChip de Affymetrix siguiendo las instrucciones del fabricante (GeneChip Hybridization Oven 645, GeneChip Fluidics Station 450 y GeneChip Scanner 7G). Los datos brutos de los microarrays se normalizaron por cuantiles, se corrigieron por el fondo y se transformaron en log2 mediante el procedimiento *Robust Multi-array Average* (RMA) en el paquete “oligo R” (v.1.54.1) utilizando una referencia *cdf* personalizada de BrainArray (v.25.0.0). Los análisis de expresión génica diferencial se llevaron a cabo con el paquete limma (v.3.46.0). En este caso, se establecieron dos puntos de corte para considerar la expresión diferencial de genes en función de su valor absoluto de expresión o *fold change* (FC):

FC superior o igual a 1,5: sobre-expresado.

FC inferior o igual a -1,5: infra-expresado.

2.15.2. Secuenciación de RNA (RNA-seq)

La secuenciación del RNA de los distintos clones *NOMO1-WT* y *NOMO1-KO* de las líneas celulares utilizadas en este trabajo se realizó a través de la plataforma Macrogen, Inc. (Seúl, Corea del Sur).

Se enviaron al menos 0,5 µg de RNA total de cada muestra con un RIN>7 y un rRNA>1. En un primer paso, se realizó una depleción del RNA ribosómico (rRNA) con Ribo-

Zero y se preparó la librería de cDNA con el kit *Truseq stranded total RNA with Ribo-Zero* (Illumina®). Posteriormente, se llevó a cabo la secuenciación de la librería en el equipo *NovaSeq6000 150PE (150 pb x 2)*, siguiendo las instrucciones del fabricante. Se obtuvieron un rango de 61,8-78,7 millones de lectura por cada muestra. El esquema de trabajo que sigue Illumina para la secuenciación de RNA mediante *Next Generation Sequencing* (NGS) tiene 4 etapas básicas (Figura 17):



Figura 17. Esquema de trabajo para la secuenciación de RNA.

1. Preparación de la muestra: para la construcción de la librería, se extrajo RNA total y se realizó el control de calidad (QC).
2. Construcción de la librería: para preparar la librería se realizó una fragmentación aleatoria del DNA o cDNA presente en la muestra, seguida de la ligación de los adaptadores en los extremos 5' y 3'. A continuación, los fragmentos ligados y adaptados se amplifican mediante PCR y se purifican en gel.
3. Secuenciación: para la generación de clúster, la librería se cargó en una celda de flujo donde los fragmentos fueron capturados por oligonucleótidos complementarios a los adaptadores de la librería. Estos fragmentos se amplificaron en grupos clonales distintos de manera que cada clúster quedó preparado para su secuenciación.
4. Procesamiento de datos: el secuenciador de Illumina generó imágenes de datos en bruto utilizando el software RTA (Real Time Analysis). Los datos se convirtieron en formato *FASTQ* utilizando el paquete de Illumina "*bcl2fastq*" para realizar la lectura de las secuencias.

- Análisis de datos: Las secuencias adaptadoras y las bases de baja calidad se recortaron con Trimmomatic (v.0.39). La calidad de las lecturas crudas y recortadas se analizó mediante la herramienta FASTQC (v.0.11.9) y las lecturas emparejadas se alinearon con el genoma humano de referencia hg19 mediante el alineador STAR (v.2.7.9a). La secuencia del genoma de referencia (hg19, *Genome Reference Consortium* GRCg37) y los datos de anotación (v.87) se descargaron del sitio web Ensembl (www.ensembl.org). Los recuentos a nivel de genes se calcularon utilizando el paquete HTSeq (v.0.12.4) y se normalizaron a través del método de la mediana de ratios. Finalmente, el análisis de expresión génica diferencial se realizó con el paquete DESeq2 R (v.1.30.1).

3. Técnicas de manipulación de proteínas

La extracción de proteínas se llevó a cabo de las líneas celulares y de las muestras de tejido fresco de ratón.

3.1. Extracción de proteínas a partir de líneas celulares

Para la obtención de extractos proteicos totales, procedentes de las líneas celulares, las células se tripsinizaron, previo lavado con PBS, en las placas de cultivo con 2 mL del agente de disociación TrypLE 1X (Gibco). Posteriormente, las células se centrifugaron a 1200 rpm, durante 5 minutos, y el pellet celular se incubó durante 30 minutos en hielo con un buffer de lisis suave (BLS) (NaCl 140 mM, EDTA 10 mM, 10% glicerina, 1% Nonidet P-40, Tris 20 mM, pH 8,0). El BLS fue suplementado con los siguientes inhibidores de proteasas: PMSF 1 mM, pepstatina 1 μ M, aprotinina 1 μ g/mL, leupeptina 1 μ g/mL y ortovanadato sódico 1 mM. Una vez terminada la incubación con BLS, se centrifugaron los extractos a 13200 rpm, durante 15 minutos, a 4°C, y se transfirieron los sobrenadantes resultantes a un nuevo tubo Eppendorf®. Estos extractos finales se almacenaron a -80°C hasta su uso.

3.2. Cuantificación de proteínas por el método Bradford

Los extractos proteicos totales obtenidos se cuantificaron mediante el método *Bradford*, basado en el uso del colorante “azul brillante G-250 Coomassie”. Cuando este reactivo se une a las proteínas presentes en la muestra, vira de un color pardo a un color rojo-azulado, indicativo de que se ha alterado su máximo de absorción (de 465 nm a 595 nm). De esta manera, la absorbancia medida a 595 nm tiene una relación directa con la cantidad de proteína presente en la muestra¹⁸⁴. Se utilizó el reactivo comercial *Bio-Rad Protein Assay* (Bio-Rad) y se realizó una curva patrón con concentraciones crecientes y conocidas de la albúmina sérica bovina (BSA). Tras la cuantificación, se prepararon extractos con la misma cantidad de proteína total, a los que se añadió tampón de carga (Tris 100 mM pH 6,8, 20% glicerol, 4% SDS, 0,05% azul de bromofenol y 2% β -mercaptoetanol). Finalmente, los extractos se desnaturalizaron a 100°C durante 5 minutos en un baño seco.

3.3. Electroforesis en SDS-PAGE

Las proteínas se separaron en función de su tamaño mediante electroforesis vertical en geles SDS-PAGE (*sodium dodecylsulfate – polyacrilamide gel electrophoresis*). El SDS es un detergente aniónico encargado de desnaturalizar y unirse a las proteínas,

confiriéndoles una carga negativa uniforme gracias a sus grupos sulfato. Así, la carga de la proteína queda enmascarada por las moléculas de SDS de manera que se alcanza una relación carga: masa uniforme. Esto permite que las proteínas migren de igual manera hacia el polo positivo en función de su masa molecular.

Para separar las proteínas en función de su tamaño se preparó un gel separador (*resolving*) a un porcentaje de acrilamida variable en función del tamaño de las proteínas a separar (8-12% acrilamida/bis-acrilamida 30%) (Bio-Rad) en tampón Tris-HCl 1,5 M, pH 8,8 y SDS al 10%. Encima de este gel separador, se preparó un gel concentrador (*stacking*) compuesto por acrilamida al 4,8% y bis-acrilamida al 0,128% en tampón Tris-HCl, pH 6,8, a 0,125 M y 3,5 mM de SDS. Para la polimerización de ambos tipos de geles, se añadió un agente iniciador de la polimerización, el persulfato de amonio (APS) y un catalizador de la formación de radicales libres, el TEMED (N, N, N', N'-tetrametiletilenodiamina).

Se cargaron un total de 50 ng/pocillo de cada extracto proteico analizado en este trabajo. La electroforesis se realizó a un voltaje constante (120 voltios) en condiciones desnaturizantes en tampón de carrera (Tris 20 mM, Glicina 192 mM y SDS al 1%). Como marcador de peso molecular se utilizó el marcador *PageRuler™ Prestained Protein Ladder Plus* (Thermo Fisher Scientific) que muestra los siguientes tamaños en KDa: 170, 130, 100, 70, 55, 35, 25, 15 y 10.

3.4. Transferencia semi-seca y western blot (WB)

Después de la electroforesis, las proteínas se transfirieron a una membrana *PVDF Immobilon-P* (Millipore Corporation) mediante el sistema de transferencia semi-seca TRANS BLOT® (Bio-Rad) durante 45 minutos a 25 voltios. El tampón de transferencia estaba compuesto por 25 mM Tris-HCl, pH 8,5, 192 mM Glicina y 20% metanol.

Tras la transferencia, y para llevar a cabo el *western blot* (WB), la membrana se bloqueó durante 1 hora, en rotación y a temperatura ambiente, en una solución de leche desnatada en polvo al 5% en tampón TBST-T (Tween-20 al 0,1% en TBS) con el fin de saturar todos los sitios inespecíficos de unión a proteínas. A continuación, la membrana se incubó a 4°C, durante toda la noche, con el anticuerpo primario (Tabla 8) necesario para detectar nuestra proteína de interés a la dilución correspondiente en TBS-T (3-5% BSA), según las recomendaciones del fabricante. Después de la incubación, se realizaron 3 lavados durante 10 minutos cada uno con TBS-T en agitación y se incubó la membrana con el anticuerpo secundario (Tabla 8) a una dilución 1:10000 conjugado a distintos fluorocromos en función de su reactividad. Esta incubación se realizó durante una hora y en agitación.

Finalmente, se realizaron 3 lavados con TBS-T y se reveló la membrana en el equipo *LI-COR Odyssey Infrared Imaging System*, que detecta la fluorescencia emitida por los anticuerpos secundarios reconociendo la fracción constante de los anticuerpos específicos de las proteínas a estudiar en cada caso. Para la cuantificación de la intensidad de las bandas obtenidas se empleó el software *ImageJ*.

3.5. Anticuerpos

En este trabajo se han estudiado los niveles de expresión proteica de las siguientes proteínas (Tabla 8): NOMO, NCLN, TMEM147, CRIPTO-1, ALK4, ACTRII-B, SMAD2/3, pSMAD2/3, SMAD4, HMGA1, LMNB1, CTND1, E-Cadherina, Vimentina y β -Catenina. Como control de carga se ha utilizado la ACTINA.

Tabla 8. Anticuerpos primarios y secundarios utilizados en este trabajo para la determinación de expresión proteica mediante western blot (WB).

Anticuerpo	Tipo	Origen	Conc. WB	Tamaño proteína (kDa)	Casa comercial
ACTINA	Primario	Ratón	1:10000	43	SigmaAldrich (A5441)
NOMO	Primario	Cabra	1:250	130	R&D Systems (AF3755)
NCLN	Primario	Conejo	1:500	62	SigmaAldrich (SAB1501379)
TMEM147	Primario	Conejo	1:1000	25	Abcam (ab97624)
CRIPTO-1	Primario	Conejo	1:1000	18	Novus Biologicals (100-1598)
ALK4	Primario	Conejo	1:1000	57	Abcam (ab109300)
ACTRII-B	Primario	Ratón	1:500	58	Santa Cruz (sc-390977)
SMAD2/3	Primario	Ratón	1:1000	55-60	Santa Cruz (sc-398844)
pSMAD2/3	Primario	Conejo	1:1000	55-60	Abcam (ab272332)
SMAD4	Primario	Ratón	1:1000	61	Santa Cruz (sc-7966)
HMGA1	Primario	Conejo	1:1000	12	GeneTex (GTX122110)
LMNB1	Primario	Conejo	1:1000	66	Sigma-Aldrich (ZRB1143)
CTND1	Primario	Conejo	1:1000	108	GeneTex (GTX130448)
E-Cadherina	Primario	Ratón	1:1000	120/80	Santa Cruz (sc-8426)
Vimentina	Primario	Ratón	1:1000	57	Santa Cruz (Sc-6260)
β -Catenina	Primario	Ratón	1:1000	92	BD Biosciences (610154)
Anti-Mouse	Secundario	Cabra	1:10000		ThermoFisher (35518)
Anti-Rabbit	Secundario	Cabra	1:10000		ThermoFisher (SA5-35571)
Anti-Goat	Secundario	Burro	1:10000		Santa Cruz (sc-2020)
Anti-Rat	Secundario	Cabra	1:10000		ThermoFisher (A-11006)

3.6. Estudio del proteoma mediante cromatografía líquida acoplada a espectrometría de masas (LC-IMS/MS)

Los cambios en los perfiles de expresión proteica tras la inactivación de NOMO se analizaron mediante cromatografía líquida por espectrometría de masas (LC-IMS/MS). El procesamiento de las muestras se llevó a cabo en la Unidad de Proteómica del Parque Científico y Tecnológico de Vizcaya. El diseño y análisis de los datos se realizó en colaboración con la Unidad de Proteómica del CIC de Salamanca. Para ello, se compararon un total de 4 réplicas de un clon NOMO-WT y 4 réplicas de un clon NOMO-KO de las líneas celulares HCT-116 y HS-5.

Todas las células purificadas (2000-3000 células) fueron procesadas con solución de lisis e inhibidores de fosfatasas y proteasas¹⁸⁵. Posteriormente, se redujeron 0,5 µg de proteína total con ditioneitol (DTT) 10 mM + yodacetamida 55 mM a temperatura ambiente durante 45 minutos. Las proteínas se digirieron con tripsina (1:50) a 37 °C durante 18 horas. Luego, la mezcla de péptidos se acidificó con TFA al 0,1 % y se desalinizó con C18 StageTips. Las muestras se almacenaron a -20 °C hasta que se realizaron las pruebas LC-IMS/MS.

Para el análisis LC-IMS/MS se utilizó un sistema nanoUPLC (NanoElute, Bruker) con una columna C18 de 15 cm x 75 µm, con partículas C18 de 1.6 µm (Ion Optics Inc.) y 120 minutos de gradiente (3-50% de ACN a 300 nL/min.), que se acopla a un TimsTOF Pro (Bruker). El TimsTOF Pro se configuró en modo PASEF (Fragmentación En Serie De Acumulación Paralela) utilizando Compass Hystar 5.036.0. Los ajustes para 11 muestras por día fueron los siguientes: rango de masa 100 a 1700 m/z, 1/K0 Inicio 0,6-Fin 1,6 V·s/cm², tiempo de rampa 110,1 ms, ciclo de trabajo de bloqueo 100 %, volumen capilar 1600 V, gas seco 3 l/min, temperatura seca 180 °C, configuración de PASEF: 10 escaneos MS/MS (ciclo total 1,27 s), rango de carga 0-5, exclusión activa 0,4 min, intensidad objetivo programada 10000, umbral de intensidad 2500, Energía de colisión CID 42eV¹⁸⁶.

Todos los archivos sin procesar fueron analizados con el software MaxQuant v1.6.6.0 utilizando el motor de búsqueda integrado Andromeda. La búsqueda de datos se realizó contra el proteoma de referencia Uniprot humano y una base de datos de señuelo inverso separada que utiliza una especificidad de tripsina estricta que permite hasta 2 escisiones perdidas. La longitud mínima del péptido se estableció en 7 aminoácidos considerando la carbamidometilación de cisteínas como modificación fija y la N-acetilación y oxidación de metioninas como modificaciones variables. La masa máxima de péptidos (Da) se estableció en 8000. Todas las demás configuraciones en los parámetros globales o de

grupo fueron las predeterminadas para MaxQuant. El cribado de PTMS se realizó siguiendo PTMScan Direct PTMScan® m (Cell Signaling Technology, EEUU) con las ligeras modificaciones descritas^{187,188}.

4. Análisis de enriquecimiento funcionales

Una vez se generaron las lecturas de los transcritos (mediante los microarrays de expresión y secuenciación de RNA) y proteínas (mediante el proteoma), se realizó un análisis de enriquecimiento de rutas utilizando la herramienta *WebGestalt* (WEB-based GENE SeT AnaLysis Toolkit; <http://www.webgestalt.org>). Esta herramienta de análisis permite realizar estudios de enriquecimientos funcionales derivados de la desregulación de un conjunto de genes. Esta herramienta incluye hasta 12 organismos diferentes, reconociendo más de 342 identificadores diferentes de genes y dispone de 155175 categorías funcionales distintas donde poder clasificar estos genes expresados de manera diferencial¹⁸⁹.

El funcionamiento de la herramienta web se puede resumir en los siguientes pasos:

- Se introdujo el organismo de estudio: *Homo sapiens*.
- Se seleccionó el tipo de análisis: *Over Represent Analysis (ORA)* o *Gene Set Enrichment Analysis (GSEA)*. La plataforma permite analizar según Wikipathways, Pathway Commons, dianas de factores de transcripción, dianas de microRNAs, redes de interacción de proteínas, localización cromosómica, asociación a enfermedades, asociación a fármacos o fenotipos.
- En el sistema de identificación de rutas “*pathway*” se seleccionó la base de datos para identificar las rutas enriquecidas: “*reactome*” o “*KEGG*”. En el sistema de rutas “*community-contributed*” se eligió la base de datos “*Hallmarks*”.
- Se introdujo el listado de genes o proteínas diferencialmente expresadas utilizando los siguientes identificadores: *Gene symbol* (array de expresión), *Ensembl* (RNA-seq) o *UniProt ID* (proteoma).
- Se seleccionó un conjunto de datos de referencia. En nuestro caso, “genome Protein-coding”.
- Se estableció el método de ajuste para testeo múltiple y nivel de significancia: el número de mínimo de genes por categorías fue “2” para el análisis ORA y “3” para el análisis GSEA, con un máximo de “2000” y un nivel de significancia “top 20”.

5. Cultivo de líneas celulares

Las líneas celulares y el medio de cultivo utilizado para su correcto crecimiento se detallan en la tabla 9. Todas las líneas se cultivaron en monocapa en placas de cultivo de 10 cm de diámetro (BD Falcon™), en placas de 6, 12, 24 o 96 pocillos, según los requerimientos del ensayo. Los cultivos se mantuvieron en un incubador (Cultek S.L.U., Madrid, España) a 37°C, 5% de CO₂ y 98% de humedad relativa. Las líneas celulares utilizadas en este trabajo se cultivaron con el medio de cultivo *Dubelcco's Modified Eagle Medium* (DMEM) (Sigma-Aldrich) con glucosa (4,5 g/L), L-glutamina y L-piruvato. El medio fue suplementado con un 10% de suero fetal bovino (FBS) y un 1% de los antibióticos penicilina-estreptomicina (GIBCO-Life Technologies, California, Estados Unidos). La presencia de micoplasma en el cultivo celular fue testada de manera rutinaria utilizando el kit *MycoAlert kit* (Lonza, Basilea, Suiza) siguiendo las instrucciones del fabricante.

Tabla 9. Principales características de las líneas celulares utilizadas.

Línea celular	ATCC® nº	Organismo / Tipo	Tumoral	Genes mutados	Mutaciones
HCT-116	CCL-247™	Humano / Adenocarcinoma colorrectal (48 años)	Si	<i>CDKN2A</i> <i>CTNNB1</i> <i>KRAS</i> <i>MLH1</i> <i>PI3KCA</i>	c.68_69insG; p.R24fs*20 c.133_135delTCT;p.S45d el c.38G>A;p.G13D c.755C>A;p.S252* c.140A>G; p.H1047R
HS-5	CRL- 11882™	Humano / Médula ósea (30 años)	No		

5.1. Contaje de células

El contaje de células se llevó a cabo en una cámara de Neubauer. Se trata de una placa gruesa de cristal con forma de portaobjetos que presenta 2 cámaras a cada uno de sus extremos. Cada cámara consta de 4 cuadrantes sobre los que se depositan las células. Para ello, se cubrió una de las cámaras con un cubreobjetos y se añadieron 40 µL del cultivo celular en el borde de este para que la muestra penetre en la cámara por capilaridad. Posteriormente, se contaron los 4 cuadrantes en un microscopio óptico y se calculó el número total de células por cada cuadrante. El número total de células por cada mL de nuestra muestra se estableció acorde al factor de dilución de la cámara (10000):

$$\frac{\text{Nº células totales}}{4 \text{ cuadrantes}} \times 10000 = \text{Nº células por mL de muestra}$$

5.2. Fármacos

Para estudiar la viabilidad celular en respuesta a agentes quimioterápicos se han utilizado 4 fármacos diferentes proporcionados por el Servicio de Oncología Médica del Hospital Universitario de Salamanca:

5-fluorouracilo (5-FU) (Accord Healthcare S.L.U): Se preparó una dilución 1:1000, en un volumen final de 1 mL, a partir de un stock de 0,384 M. Se utilizó como vehículo el medio de cultivo propio de las líneas celulares tratadas (DMEM).

Irinotecán (Fresenius Kabi): Se preparó una dilución 1:10, en un volumen final de 200 μ L, a partir de un stock de 34,1 mM. Se utilizó DMEM como vehículo.

Oxaliplatino (Sun Pharma). Se preparó una dilución 1:10, en un volumen final de 500 μ L, a partir de un stock de 12,58 mM. Se utilizó DMEM como vehículo.

Cisplatino (Pfizer S.A): Se utilizó directamente desde un stock de 3,3 mM.

Las concentraciones utilizadas de cada fármaco para cada uno de los ensayos de viabilidad celular (MTT y Anexina V/Yoduro de propidio (PI)) se detallan en la tabla 10.

Tabla 10: Representación de las dosis utilizadas en los ensayos de viabilidad celular en respuesta al tratamiento con 5-FU, irinotecán, oxaliplatino y cisplatino en líneas celulares.

HCT-116	MTT						ANEXINA V/PI		
Fármaco / Dosis (μ M)	D1	D2	D3	D4	D5	D6	D1	D2	D3
5-FU	2	3	6	8	10	15	3	6	-
Irinotecán	1	2,5	5	10	25	50	5	15	25
Oxaliplatino	1	2,5	5	10	25	50	5	15	25
Cisplatino	1	2,5	5	10	25	50	5	15	25
HS-5	MTT						ANEXINA V/PI		
Fármaco / Dosis (μ M)	D1	D2	D3	D4	D5	D6	D1	D2	D3
5-FU	1	5	10	15	25	50	5	25	50
Irinotecán	1	5	10	15	25	50	5	15	25
Oxaliplatino	1	5	10	15	25	50	5	15	25
Cisplatino	0,25	0,50	0,75	1	2,50	5	5	15	25

5.3. Ensayo de viabilidad celular

El crecimiento y supervivencia celular en condiciones basales o en respuesta a los tratamientos mencionados en el apartado anterior se analizó mediante el ensayo de viabilidad MTT¹⁹⁰. Esta técnica permite el marcaje de las células metabólicamente activas a través del bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolio (MTT) (Sigma-Aldrich). Este compuesto de color amarillento es reducido por la enzima succinato deshidrogenasa (SDH), presente en las mitocondrias de las células, dando lugar a un compuesto hidrofóbico de color azul oscuro, el formazán. El formazán se solubiliza con DMSO (dimetilsulfóxido), lo que permite medir absorbancia de la muestra en un espectrofotómetro y así determinar la tasa de proliferación. La absorbancia detectada a 570 nm se correlaciona de forma directa con la cantidad de células en cultivo metabólicamente activas y, por tanto, con la tasa de proliferación (Figura 18).

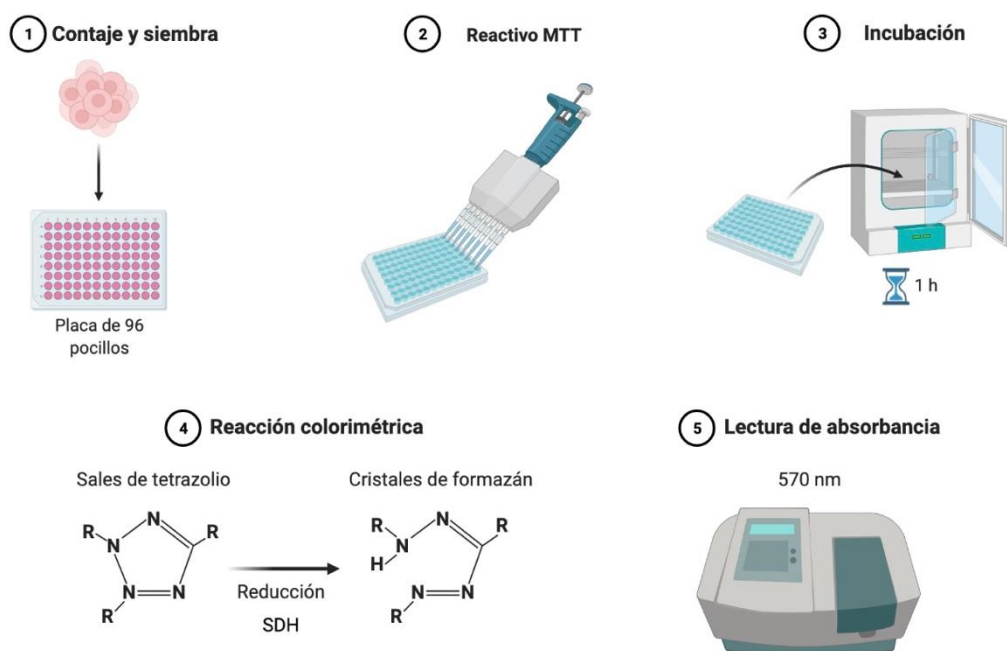


Figura 18. Representación gráfica del procedimiento para medir la viabilidad celular mediante el ensayo de MTT.

Para llevar a cabo este ensayo se sembraron un total de 4000 células por pocillo, en placas de 96 pocillos, de cada una de las dos líneas celulares empleadas en este trabajo (HCT-116 y HS-5). Las células permanecieron en cultivo durante 24 horas, momento en el que se sustituyó el medio de cultivo por uno nuevo con las concentraciones crecientes de fármacos mencionadas en el apartado anterior. Las dos líneas celulares fueron tratadas con 5-fluorouracilo, irinotecán, oxaliplatino y cisplatino. En todos los casos, se dejaron células sin tratamiento como control para medir la tasa de proliferación basal de cada línea celular

a 0, 24, 48 y 72 horas. La tasa de proliferación de las placas con tratamiento se midió tras 24, 48 y 72 horas de incubación. Para realizar la medición a cada uno de los tiempos, se añadió a cada pocillo 1 mg/mL del reactivo MTT (10 μ L) y se incubó durante 1 hora a 37°C. Tras esto, se aspiró el medio y se añadieron 100 μ L de DMSO para disolver los cristales de formazán depositados en cada pocillo durante 5 minutos en agitación. La absorbancia de la muestra se determinó en el lector de microplacas *UltraEvolution* (Tecan®) a 570 nm. Cada experimento, con o sin tratamiento, fue replicado al menos 2 veces.

5.4. Análisis de muerte celular mediante citometría de flujo

Para estudiar la muerte celular mediante citometría de flujo se utilizó el kit *Anexina V FITC Apoptosis detection kit* (Immunostep S.L). En este caso, la anexina se une a la fosfatidilserina (PS) presente en la membrana celular, indicativo de células que inician apoptosis. El yoduro de propidio, por su parte, penetra en las células que han perdido la integridad de sus membranas, indicativo de las células necróticas. Así, podemos distinguir 4 poblaciones diferentes: células vivas (Anexina V-/PI -; cuadrante inferior izquierdo), células apoptóticas (Anexina V+/PI-; cuadrante inferior derecho), células necróticas (Anexina V-/PI+; cuadrante superior izquierdo) y células necro-apoptóticas (Anexina V+/PI+; cuadrante superior derecho) (Figura 19).

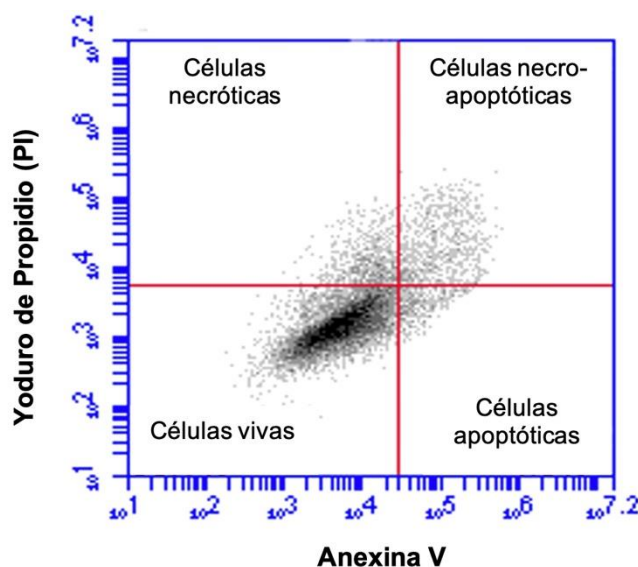


Figura 19. Representación gráfica de las poblaciones detectadas mediante citometría de flujo en el análisis de muerte celular (Anexina V/Yoduro de propidio).

Para el análisis de muerte celular mediante citometría de flujo se sembraron 30000 células por pocillo de cada línea celular (HCT-116 y HS-5) en placas de 12 pocillos. Las células se incubaron durante 24 horas, momento en el que se añadieron las concentraciones

crecientes de cada uno de los fármacos mencionados en el apartado 5.2 de esta metodología. Las dos líneas celulares fueron tratadas con 5-fluororuracilo, irinotecán, oxaliplatino y cisplatino. La cantidad de fármaco a añadir se calculó en base a la siguiente fórmula:

$$C_i \cdot V_i = C_f \cdot V_f$$

Donde C_i ; concentración inicial del fármaco, V_i ; volumen de fármaco a añadir, C_f ; concentración final del fármaco, V_f ; volumen del medio de cultivo (150 μ L).

La muerte celular se evaluó tras 72 horas de incubación con cada fármaco. Para ello, se recogió el sobrenadante de cada pocillo y se añadió a un tubo de citómetro previamente rotulado. Las células adheridas a los pocillos se lavaron con 300 μ L de PBS y se tripsinizaron con 300 μ L del agente de disociación TrypLE 1X (Gibco). Una vez despegadas de la placa, se añadieron al tubo de citómetro correspondiente donde ya se había añadido su sobrenadante y se centrifugaron durante 5 minutos a 1000 rpm. Tras retirar el sobrenadante, el pellet celular resultante se resuspendió en 100 μ L de la reacción de marcaje (98 μ L de Binding Buffer, 1 μ L de Anexina V, 1 μ L de PI). La mezcla se incubó durante 15 minutos en oscuridad y a temperatura ambiente. Finalmente, se añadieron 150 μ L de Binding Buffer a cada tubo y se adquirieron las muestras en un citómetro de flujo Accuri™ C6 (BD Biosciences). Los datos fueron analizados en el programa asociado al citómetro y posteriormente se trasladaron los porcentajes de cada población a un archivo Excel para su análisis final.

5.5. Análisis del estado *stem cell* mediante citometría de flujo

Para estudiar la implicación de NOMO en la diferenciación celular, se realizó un análisis de expresión de las proteínas de superficie celular CD44 y CD24. Ambas proteínas se consideran marcadores de células madre de cáncer de colon (CCSC)¹⁹¹ y están asociadas principalmente a procesos de adhesión celular, migración y metástasis¹⁹². En este caso, se compararon 2 clones NOMO-KO y 2 clones NOMO-WT de las líneas celulares HCT-116 y HS-5.

Para ello, se realizó un marcaje directo empleando un anticuerpo específico conjugado con un fluorocromo que emite señal en un determinado canal que es detectada mediante citometría de flujo. En primer lugar, se contaron y sembraron 1 millón de células de cada condición en una placa de cultivo de 10 cm de diámetro. A las 48 horas, las células se lavaron con 5 mL de PBS, se levantaron con 1,5 mL del agente de disociación TrypLE 1X (Gibco™) y se recogieron en medio de cultivo. Posteriormente, se realizó una

centrifugación durante 5 minutos a 1000 rpm y se resuspendió el pellet celular en 200 μ L de PBS al que se añadió el volumen correspondiente a 1 test de los anticuerpos anti-CD44 conjugado en FITC (fluoresceína-5-isotiocianato) o anti-CD24 conjugado en APC (aloficocianina) (Tabla 11), según corresponda. Tras 15 minutos de incubación en oscuridad con cada anticuerpo, se añadieron 150 μ L adicionales de PBS y se analizaron los porcentajes de células positivas para cada marcador mediante citometría de flujo. Para ello, se empleó el citómetro de flujo FACS Aria™ III (BD Biosciences).

Tabla 11. Anticuerpos utilizados para la determinación del estado de diferenciación mediante citometría de flujo.

Anticuerpo	Fluorocromo	Origen	Clase	Reactividad	Dilución	Casa comercial
CD44	FITC	Ratón	Monoclonal	Humano	20 μ L / test	BD Biosciences (#347943)
CD24	APC	Ratón	Monoclonal	Humano	5 μ L / test	ThermoFisher Scientific (#47-0247-42)

6. Estudio de migración celular

El análisis de la capacidad migratoria de las líneas celulares generadas en este trabajo se realizó mediante el ensayo de *scratch* o *cierre de herida* y el ensayo de migración celular mediante *transwell*.

6.1. Ensayo de *scratch* o *cierre de herida*

El ensayo de *cierre de herida* consiste en observar el comportamiento de una monocapa confluyente de células, que migran hasta establecer nuevos contactos célula-célula con el fin de cerrar una “herida” previamente realizada en la placa de cultivo. Con este experimento, podremos calcular la velocidad de migración de las células que se desplazan por la “herida”, ya que obtendremos medidas sobre el tiempo necesario para ocupar un área determinada por las mismas¹⁹³.

Para analizar los cambios en la capacidad migratoria tras la inactivación de NOMO, se sembraron 500.000 células por pocillo de las líneas celulares HCT-116 y HS-5. El ensayo se realizó en placas de cultivo de 6 pocillos (BD, Falcon™) y se estudiaron 3 réplicas de cada clon empleado en el estudio (NOMO-KO y NOMO-WT). Las células permanecieron en cultivo durante 48 horas hasta alcanzar un 95% de confluencia. Posteriormente, se sustituyó el medio de cultivo por un medio con menor concentración de FBS (2% FBS, 1% penicilina),

con el objetivo de parar su capacidad proliferativa sin comprometer la viabilidad celular. El cultivo celular se mantuvo en estas condiciones de privación de suero durante 48 horas. En este momento, se retiró el medio de cultivo, se lavó con PBS y se hicieron 3 “*scratches* o *heridas*” verticales en cada pocillo. Se volvió a añadir un nuevo medio de cultivo al 2% de FBS y se procedió a la grabación de las imágenes.

La toma de imágenes se realizó 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. El experimento se realizó en una atmósfera a 37°C y 5% de CO₂.

6.2. Ensayo de migración celular mediante *transwell*

La habilidad de las células carentes de NOMO para migrar a través de una membrana porosa fue evaluada mediante el ensayo de *transwell*. Este ensayo se realizó en placas de cultivo de 24 pocillos, donde se colocó un inserto que porta una membrana de policarbonato con poros de 8 µm (SARSTEDT Ag & Co. KG, Alemania). El inserto consta de dos cámaras. En la cámara inferior se agregaron 600 µL de medio de cultivo DMEM suplementado con un 20% de FBS. En la cámara superior se agregaron 50.000 células de cada condición en 300 µL de DMEM al 0% de FBS y se incubaron durante 24 horas. Posteriormente, se cuantificaron las células que atravesaron la membrana mediante fijación con formaldehído 3% durante 15 minutos a temperatura ambiente. A continuación, se realizó una tinción con cristal violeta 1% durante 30 minutos. Las células que no atravesaron la membrana se descartaron con un hisopo estéril y se ubicó el inserto en un portaobjeto. Las células que migraron hacia el lado inferior de la membrana se observaron bajo el microscopio de contraste de fases.

Para calcular el porcentaje de migración se utilizó el programa ImageJ®, que permitió el tratamiento de las imágenes captadas en el microscopio. En el ensayo de cierre de herida, se empleó la herramienta *MRI Wound Healing Tool* que mide el área libre de la herida a un tiempo concreto. Así, la disminución del área libre a lo largo del tiempo para cada una de las 3 réplicas de cada clon permitió calcular el porcentaje de área ocupada de cada línea celular y, por extensión, su velocidad de migración. En el ensayo de migración celular mediante *transwell*, las imágenes digitales fueron tomadas con un aumento de 10X y se cuantificaron las células que migraron a través de la membrana mediante el “Análisis de Partículas” en modo binario.

7. Edición génica mediante la tecnología CRISPR/cas9

Actualmente, el desarrollo de la edición genómica dirigida ocupa una de las grandes áreas de la biología molecular y de la terapia génica. Existen distintas técnicas basadas en la acción de proteínas con dominios de unión a secuencias de DNA específicas que utilizan su capacidad endonucleasa para generar la escisión de una secuencia concreta¹⁹⁴ (Figura 20).

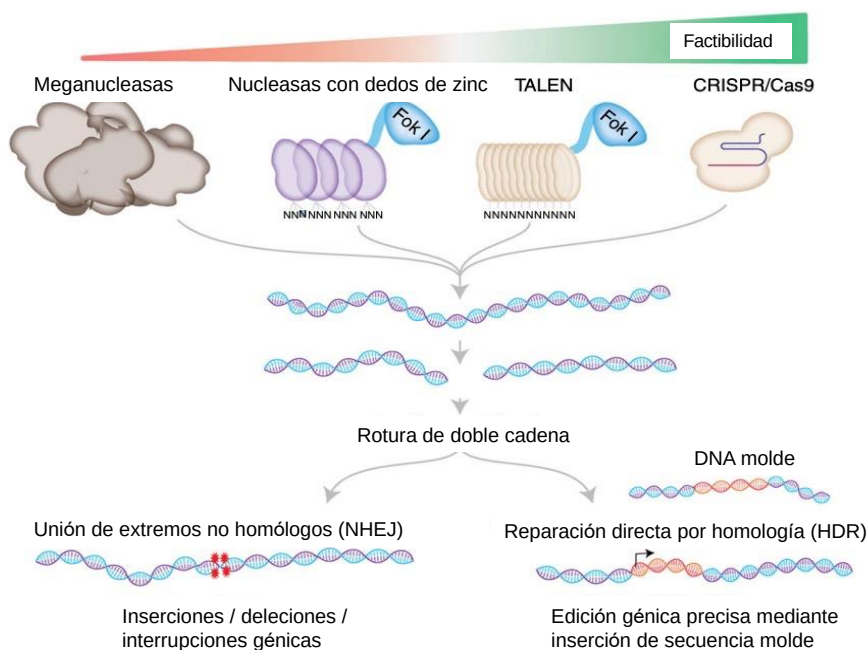


Figura 20. Funcionamiento básico de los principales métodos de edición del genoma. Adaptado de Mazhar Adli, 2018.

Entre estas técnicas de edición génica se encuentra la herramienta CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat). CRISPR se identificó como un mecanismo de defensa adaptativo en el sistema inmune de bacterias y arqueas ante la infección por patógenos como los bacteriófagos. Se trata de secuencias repetitivas palindrómicas cortas, agrupadas y regularmente espaciadas, que junto con la endonucleasa Cas forman el complejo CRISPR/cas. Se fundamenta en la incorporación del DNA del patógeno entre estas secuencias palindrómicas repetidas para generar un RNA específico (*CRISPR-RNA*; crRNA) al transcribirse. Así, ante una segunda infección por el patógeno, el crRNA reconoce por complementariedad su secuencia genómica y dirige la Cas a la misma para su degradación^{195,196}.

Se han identificado tres tipos de sistemas CRISPR (tipo I-II-III) en una amplia gama de huéspedes de bacterias y arqueas que difieren en la enzima utilizada para ejercer el corte: Cas3 en CRISPR-I, Cas9 en CRISPR-II y Cas6 en CRISPR-III¹⁹⁷.

El sistema de edición CRISPR-II es el más utilizado y el primero en ser adaptado para la edición de genomas en células eucariotas^{198,199}. En este caso, para generar el corte de doble hebra en el DNA se requieren dos componentes: la endonucleasa Cas9 y una molécula de RNA que actúa como guía (*single-guide RNA*; sgRNA). La enzima Cas9 se caracteriza por una arquitectura bilobular: el lóbulo REC reconoce la región diana y el lóbulo NUC contiene los dominios nucleasa HNH y RuvC. En la región C-terminal del dominio NUC se encuentra la región de reconocimiento de la secuencia PAM (*protospacer adjacent motif*), que varía en función del organismo de procedencia de la Cas9. Para dirigir eficientemente la Cas9 al punto de corte en el DNA se necesita generar un sgRNA funcional, formado por el crRNA y un RNA-CRISPR trans-activador (tracrRNA). El sgRNA tiene aproximadamente 20 nucleótidos que hibridan por complementariedad de bases con la región diana a editar. Adyacente a este punto se localiza la secuencia PAM, la cual se compone de una serie de nucleótidos localizados en la región 3' de la hebra de DNA diana que se une al sgRNA (Figura 21).

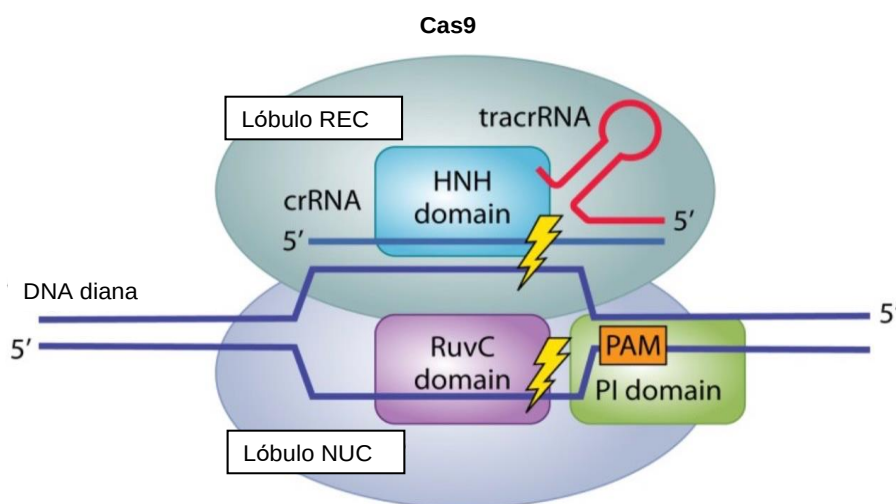


Figura 21. Arquitectura bilobular de la enzima Cas9.

Una vez la enzima Cas9 reconoce la secuencia PAM en la región diana, ejerce un corte de doble hebra en el DNA bicatenario (DSB). Este corte se localiza entre 3-4 nucleótidos antes de la secuencia PAM (en el sentido 5'-3'). Una vez se produce el corte, la maquinaria celular de reparación del daño en el DNA repara el corte a través de dos vías: mediante unión de extremos no homólogos (NHEJ) o mediante reparación directa por homología (HDR). En la reparación por NHEJ, los dos fragmentos bicatenarios de DNA son religados, produciéndose modificaciones en forma de inserciones o deleciones (InDel) que pueden cambiar el marco de lectura o generar codones de parada prematuros que originen una proteína truncada²⁰⁰. En la reparación por HDR, se necesita de una secuencia donadora

que ejerza de molde, por lo que el mecanismo de reparación incorpora la secuencia de interés en el genoma de la célula (Figura 22).

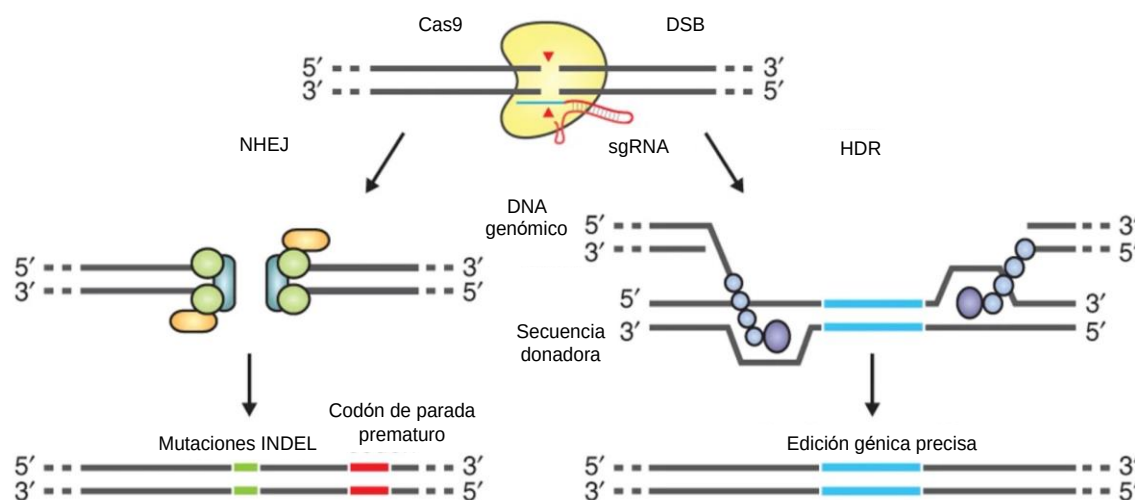


Figura 22. Mecanismos de reparación del daño en el DNA inducido por la acción de la enzima Cas9. Adaptado de Ran F. A. *et al*, 2013.

7.1. Diseño de sgRNAs

Para diseñar los sgRNAs, se ha realizado previamente una búsqueda bioinformática de la secuencia de referencia del gen a editar con el fin de seleccionar la región diana óptima para inactivar la proteína. Para ello, se ha utilizado la base de datos genómicos Ensembl (<http://www.ensembl.org>), donde se ha consultado la secuencia genómica del gen *NOMO1*. En este trabajo, se ha seleccionado como secuencia diana el exón 3 de *NOMO1* (31 exones en total). Además, también se han incluido las regiones intrónicas adyacentes de este exón. La selección de este fragmento del gen se debe a dos razones principales:

1. La región diana se encuentra al principio de toda la secuencia del gen, por lo que existe una larga secuencia posterior para generar un codón de parada prematuro tras la generación de InDels en el punto de corte de la cas9. Esto resultaría en la síntesis de una proteína truncada.
2. En el caso de *NOMO1*, la última base de su secuencia codificante forma un codón con las dos primeras bases del exón siguiente. Por lo tanto, si se elimina este fragmento del exón 3, se romperá el marco de lectura de la secuencia codificante dando lugar a una proteína truncada (Figura 23).

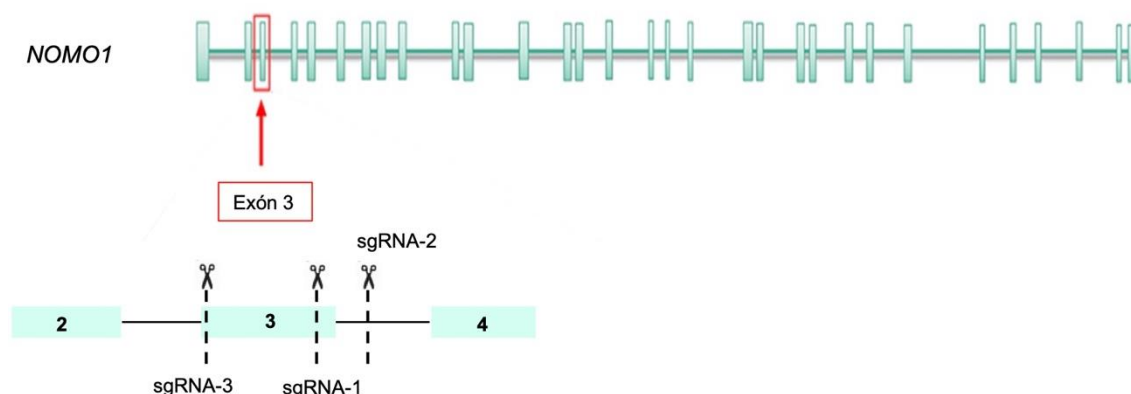


Figura 23. Representación gráfica de la secuencia genómica del gen *NOMO1*. Se marcan los 3 sgRNAs seleccionados con sus puntos de corte por la cas9 en la región diana. Las secuencias de nucleótidos de cada sgRNA presentan en su extremo 3' la secuencia PAM con el formato 5'-NGG-3'.

El diseño de los sgRNAs se ha realizado utilizando la herramienta informática Breaking-Cas (<https://bioinfogp.cnb.csic.es/tools/breakingcas/>), diseñada por el Centro Nacional de Biotecnología (CNB-CSIC). En primer lugar, se insertó la secuencia del fragmento que se quiere editar en formato *FASTA*, que debe tener entre 20 y 250 pb. En segundo lugar, se indicó la enzima Cas9 que porta el plásmido donde se clonaron los sgRNAs y que fue transfectado a las células. En este trabajo, se ha utilizado la enzima Cas9 de *Streptococcus pyogenes*, que reconoce secuencias PAM con el formato 5'-NGG-3' en la región diana²⁰⁰. Finalmente, la herramienta proporciona distintos sgRNAs candidatos y las secuencias de los oligonucleótidos necesarios para construirlos. Además, proporciona un *score* acerca de su eficacia y especificidad en la modificación del objetivo (modificación *on-target*) y evalúa el potencial de corte en regiones fuera de la secuencia diana (modificaciones *off-target*).

Para eliminar el exón diana, se seleccionaron 3 sgRNAs que presentaban un *score* óptimo (cercano al 100%). Las secuencias de los sgRNAs seleccionados del gen *NOMO1*, así como los oligonucleótidos correspondientes para su construcción, se especifican en la tabla 12.

Tabla 12. Secuencias dianas y oligonucleótidos necesarios para la construcción de los sgRNAs de *NOMO1*. En minúscula se indican las secuencias que facilitan la creación de los extremos cohesivos y su inserción en el plásmido.

	Secuencias diana <i>NOMO1</i>	<i>NOMO1</i> oligo-UP	<i>NOMO1</i> oligo-Low
sgRNA-1	5'-CCAAAACCTCCACCCTAGGGG-3'	5'caccgCCAAAACCTCC ACCTAGGGG3'	5'aaacCCCCTAGGGT GGAGTTTTGGc3'
sgRNA-2	5'-GATATGCCGAATATTAATTT-3'	5'caccgGATATGCCGA ATATTAATTT-3'	5'aaacAAATTAATATT CGGCATATCc3'
sgRNA-3	5'-CAGAATGAAATCCCCCTAAA-3'	5'caccgCAGAATGAAA TCCCCCTAAA-3'	5'aaacTTTAGGGGGA TTTCATTCTGc3'

7.2. Clonaje de sgRNAs

Para transfectar los componentes del sistema CRISPR/cas9 a las líneas celulares, se clonaron las secuencias de los sgRNAs en el vector plasmídico pSpCas9(BB)-2A-GFP (PX458), cuya estructura se encuentra en el repositorio “addgene #48138” (Figura 24).

En primer lugar, cada sgRNA fue clonado individualmente en el plásmido mediante ligación con una secuencia gRNA *scaffold* (1), la cual se encuentra bajo el control del promotor U6 (2). *Up-stream* del promotor U6, se localiza una diana de restricción para la enzima BbsI (3), que promueve el corte para la inserción y ligación de los sgRNAs. Para ello, en el diseño de estas secuencias se incorporaron los extremos 5'-*aaac*” y *c-3'* en el oligo LOW, y el extremo 5'-*caccg* en el oligo UP (Tabla 12).

Bajo el control del promotor CMV (4) se encuentran las secuencias de los genes hSpCas9 (5) y EGFP (6). Es decir, el vector PX458 posee la secuencia codificante para la cas9 y la proteína verde fluorescente (GFP). Esto nos permite introducir en la célula el segundo componente del sistema CRISPR/cas9 (la nucleasa) y monitorizar la eficacia de la transfección a través de la emisión de GFP.

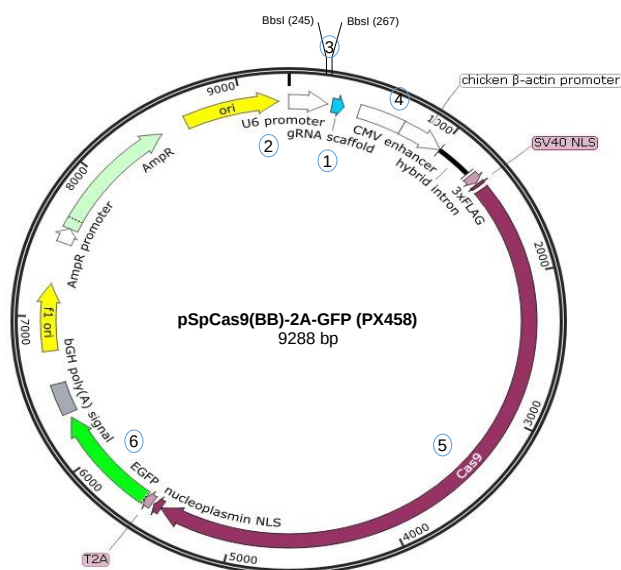


Figura 24. Esquema del vector PX458. Se detallan los puntos de interés para el funcionamiento del sistema de transfección: (1) gRNA scaffold; (2) promotor U6; (3) punto de corte por la enzima de restricción BbsI; (4) promotor CMV; (5) Cas9; (6) EGFP.

Cada sgRNA se clonó individualmente en el plásmido mediante hibridación por complementariedad de bases de los dos cebadores (oligos UP y LOW), necesarios para insertar la guía en el vector. Es decir, se construyó un oligonucleótido de doble cadena, cuyos extremos cohesivos permitieron la inserción del sgRNA en el vector por la enzima BbsI. Para ello, es necesario realizar el anillamiento de las secuencias y la fosforilación de los extremos cohesivos de cada pareja de oligonucleótidos utilizando la enzima polinucleótido quinasa T4 (PNKT4), la cual cataliza la transferencia de grupos fosfatos a los extremos cohesivos. Posteriormente, se realizó la digestión de la región del vector con la enzima BbsI y se ligaron los insertos en el vector digerido. Finalmente, se fortaleció la ligación con la exonucleasa PlasmidSafe para ayudar a eliminar posibles artefactos del procedimiento y prevenir posibles recombinaciones de productos.

Las etapas y reactivos del clonaje de sgRNAs se detallan en la tabla 13.

Tabla 13. Etapas y reactivos necesarios para el clonaje de los sgRNAs en el vector PX458.

Etapas	Reactivo	Volumen (μL)
Anillamiento y fosforilación de extremos cohesivos	Oligo-UP (100 μM)	1
	Oligo-LOW (100 μM)	1
	Buffer de ligación T4 (10X)	1
	PNK T4	0,5
	ddH ₂ O	6,5
	Volumen total	10
Digestión y ligación	Vector PX458 (100 ng/μL)	1
	Oligos UP/LOW 1:250	2
	Fast digest buffer	2
	DTT 10 mM	1
	ATP 10 mM	1
	Fast digest Bpl1	1
	DNA Ligasa T7	0,5
	ddH ₂ O	11,5
	Volumen total	20
Fortalecimiento de la ligación	Producto de ligación	11
	Buffer PlasmidSafe (10X)	1,5
	ATP 10 mM	1,5
	Exonucleasa PlasmidSafe	1
	Volumen total	15

El programa establecido en el termociclador para cada etapa en el clonaje de sgRNAs se especifica en la tabla 14.

Tabla 14. Condiciones de temperatura y tiempo para cada etapa en el clonaje de sgRNAs en el vector PX458.

Etapas	Temperatura (°C)	Tiempo (min)	Ciclos
Anillamiento y fosforilación de extremos cohesivos	37	30	1
	95	5	
	95 → 25	5°C/min	
Digestión y ligación	37	5	6
	23	5	
	4	∞	
Fortalecimiento de la ligación	37	30	1

7.3. Transformación de bacterias competentes *Escherichia coli* DH5-α

Tras la modificación del vector, se procedió a la transformación de bacterias competentes de la cepa DH5-α de *Escherichia Coli* trabajando, en todo momento, en un ambiente estéril.

En primer lugar, se descongelaron 150 µL de las bacterias en hielo durante 10 minutos exactos. Una vez atemperadas, se añadieron 2 µL del vector con cada una de las guías ligadas sobre 50 µL de bacterias. En paralelo, se añadieron también 2 µL del vector vacío (sin inserción de guías) sobre 50 µL de bacterias para utilizarlo como control. Las mezclas se incubaron en hielo durante 30 minutos. En segundo lugar, se sometieron las células a un choque térmico: 42°C durante 30 segundos seguido de 5 minutos en hielo. Tras el choque térmico, a cada uno de los tubos eppendorf (mezcla sgRNA y mezcla control) se añadieron 250 µL del medio 2XTY sin ampicilina, previamente calentado, y se incubaron en una estufa a 37°C, en agitación, durante 60 minutos. Durante esa hora, se atemperaron las placas TYE-agar suplementadas con ampicilina. Finalmente, se centrifugaron las bacterias a máxima velocidad durante 10 segundos, se aspiró el sobrenadante, se resuspendió el pellet celular en 50-100 µL de medio 2XTY y se sembraron con el asa de Digrafsky en las placas. Las placas se dejaron hacia arriba durante 15 minutos y se incubaron durante toda la noche a 37°C.

Tras 24 horas, se observó la formación de colonias en las placas y se sembraron 3 colonias/placa para su crecimiento y posterior selección. Para ello, trabajando en zona

estéril, se picaron las colonias con un palillo y se introdujeron en un tubo con 1 mL de medio 2XTY suplementado con 100 µg/mL de ampicilina. Los tubos permanecieron a 37°C, en agitación, durante 24 horas para permitir el crecimiento de las bacterias. Al día siguiente, se centrifugó el cultivo bacteriano durante 2 minutos a 10000 rpm, se descartó el sobrenadante, y se procedió a la extracción del DNA plasmídico.

7.4. Extracción y purificación de DNA plasmídico

Para extraer el DNA plasmídico de cada una de las colonias se utilizó el kit *Danagene plasmid miniprep kit* (DANAGEN-BIOTED S.L, Barcelona, España), siguiendo las especificaciones del fabricante. En resumen, se procedió de la siguiente manera:

- Se resuspendió el pellet bacteriano en 200 µL de la solución de resuspensión mediante vórtex.
- Se añadieron 200 µL de la solución de lisis y se invirtió el tubo entre 8-10 veces. Se incubó la mezcla durante 2 minutos. De esta manera, se genera la ruptura de la membrana celular.
- Se añadieron 100 µL de la solución de precipitación P3 y 150 µL de la solución de neutralización D3. Se mezcló por inversión durante 8-10 veces y se centrifugó el tubo a máxima velocidad durante 5 minutos.
- Tras la centrifugación, se trasladó el sobrenadante a un tubo nuevo con 450 µL de isopropanol frío. Se dio un vórtex y se volvió a centrifugar a máxima velocidad durante 3 minutos.
- Se descartó el sobrenadante y se resuspendió el pellet en 100 µL de la solución purificante D4. Se centrifugó a máxima velocidad durante 3 minutos y se volvió a descartar el sobrenadante.
- Se lavó el pellet de DNA con 750 µL de etanol al 70% mediante una centrifugación a máxima velocidad. Tras decantar el sobrenadante, se dejó secar el DNA durante toda la noche.
- Finalmente, se resuspendió el DNA en 40 µL de ddH_2O y se cuantificó mediante un *NanoDrop*.

7.5. Confirmación de la inserción del sgRNA mediante secuenciación Sanger

Para comprobar si los sgRNAs se habían insertado correctamente en el plásmido PX458, fue necesario secuenciar dicho vector a partir del promotor U6, en dirección *forward*. En este trabajo, se secuenció el DNA plasmídico de las bacterias transformadas con los plásmidos que integraban el sgRNA-1 y sgRNA-3, ya que habían mostrado ser eficaces en experimentos preliminares realizados en el laboratorio. Para ello, se secuenciaron por el

método de Sanger un total de 150 ng de cada uno de los productos de la miniprep, que se correspondían con cada colonia aislada. La reacción tuvo lugar en un volumen total de 8 μ L: 6 μ L del producto y 2 μ L del oligonucleótido *U6-F* 5'-GAGGGCCTATTTCCCATGATT-3'.

En este caso, todas las colonias secuenciadas mostraron una correcta inserción de los sgRNAs 1 y 3 (Figura 25).

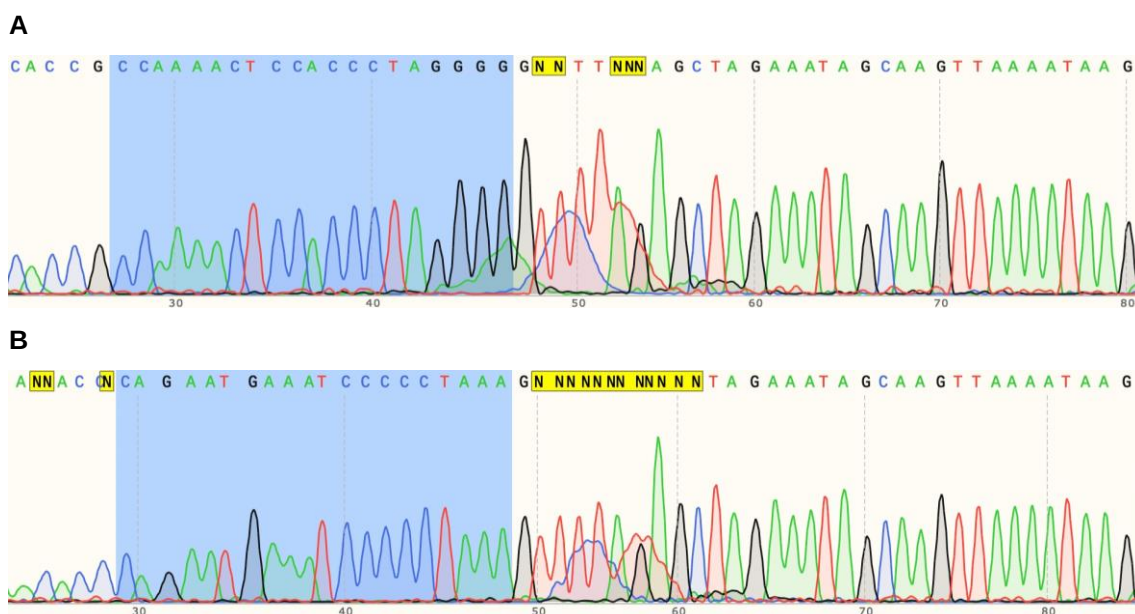


Figura 25. Secuencias correspondientes a los sgRNAs insertados en el vector PX458. Se representa con sombreado azul la secuencia específica de cada sgRNA. A) Secuencia de nucleótidos correspondiente al sgRNA-1. B) Secuencia de nucleótidos correspondiente al sgRNA-3.

Una vez identificadas las colonias que habían insertado correctamente las guías en el plásmido, se repitió la extracción de DNA empleando el kit *GeneJET Plasmid Maxiprep Kit* (Thermo Scientific™, U.S.A) para obtener mayores cantidades del material genómico. Para ello, se siguieron las instrucciones del fabricante.

7.6. Transfección y selección de las líneas celulares HCT-116 y HS-5

Una vez confirmado que los sgRNAs 1 y 3 se habían insertado individualmente en el vector, se llevó a cabo una transfección con la combinación de ambas guías en las líneas celulares HCT-116 y HS-5. Además, las dos líneas celulares también se transfectaron con el plásmido vacío a modo de control. Para mejorar la eficiencia de la transfección, la línea HCT-116 se transfectó mediante electroporación, mientras que en la línea HS-5 se llevó a cabo por lipofección.

- Transfección mediante electroporación de la línea HCT-116: la electroporación es un método de transfección capaz de modificar la permeabilidad de las membranas de las células aplicando un pulso eléctrico. Así, se consigue insertar el plásmido de interés en el interior celular por un aumento de la conductividad eléctrica y de la permeabilidad de la membrana plasmática. En este caso, se utilizaron 1×10^6 células, resuspendidas en 100 μL de *buffer* de transfección (solución V de Amaxa, Lonza), y 1 μg del plásmido correspondiente. La electroporación se llevó a cabo utilizando el programa D-032 y el equipo *Nucleofector Amaxa* (Lonza). Tras la transfección, se recogieron las células con una pipeta Pasteur y se sembraron en placas de Petri con 10 mL de medio DMEM.

Se transfectaron las siguientes combinaciones para la línea HCT-116:

A) Condición control: 1 μg de PX458 vacío.

B) Condición *NOMO1 knockout*: 0,5 μg de PX458 sgRNA-1 + 0,5 μg de PX458 sgRNA-3

- Transfección mediante lipofección de la línea HS-5: el reactivo utilizado para la lipofección fue el *PolyjetTM In Vitro DNA Transfection Reagent* (SignaGen® Laboratories). Este método consiste en la formación de un complejo de transfección entre el DNA plasmídico y el Polyjet. Este último es un polímero que recubre e inmoviliza las moléculas de DNA hasta introducirse por endocitosis en la célula. Una vez en el citoplasma, este polímero es degradado liberándose el DNA que acaba integrándose en el núcleo. Una vez en el núcleo, comienza el funcionamiento del sistema CRISPR/cas9. Para ello, se sembraron 1 millón de células de cada línea celular en placas de 10 cm de diámetro (placa de Petri) y se mantuvieron en cultivo durante 24 horas hasta alcanzar una confluencia del 90%. Antes de la transfección, se lavó el cultivo celular con PBS y se despegaron las células con 1,5 mL del agente de disociación TrypLE 1X (Gibco). Se centrifugaron a 900 rpm durante 5 minutos, se aspiró el sobrenadante y se añadió al pellet celular el complejo de transfección con un total de 8 μg de DNA plasmídico.

Se transfectaron las siguientes combinaciones para la línea HS-5:

C) Condición control: 8 μg de PX458 vacío.

D) Condición *NOMO1 knockout*: 4 μg de PX458 sgRNA-1 + 4 μg de PX458 sgRNA-3

El complejo de transfección para la lipofección se preparó de la siguiente manera:

- Tubo A: 4 μg de cada guía (8 μg total) + 400 μL de medio sin FBS. Vórtex y centrifugación a máxima velocidad (30 segundos).
- Tubo B: 32 μL de Polyjet + 400 μL de medio sin FBS. Vórtex y centrifugación a máxima velocidad (30 segundos).

- Se pasa el contenido del tubo B al tubo A. Vórtex e incubación a temperatura ambiente durante 15 minutos.

Una vez integrado el DNA en el Polyjet, se resuspendió el pellet celular en los 800 μ L del complejo de transfección y se sembraron en una nueva placa de Petri. Las células transfectadas se mantuvieron en cultivo durante 48 horas, momento en el que se comprobó mediante microscopía de fluorescencia la eficacia de la transfección a través de la detección de fluorescencia GFP. Así, aquellas células que habían incorporado la construcción emitirían luz verde gracias a la presencia de la proteína GFP que porta el plásmido. En este caso, ambas líneas se transfectaron eficientemente, si bien la eficiencia en la línea HCT-116 fue menor, probablemente debido al método de transfección empleado. La línea HS-5, por su parte, mostró una mayor presencia de células GFP+ (Figura 26).

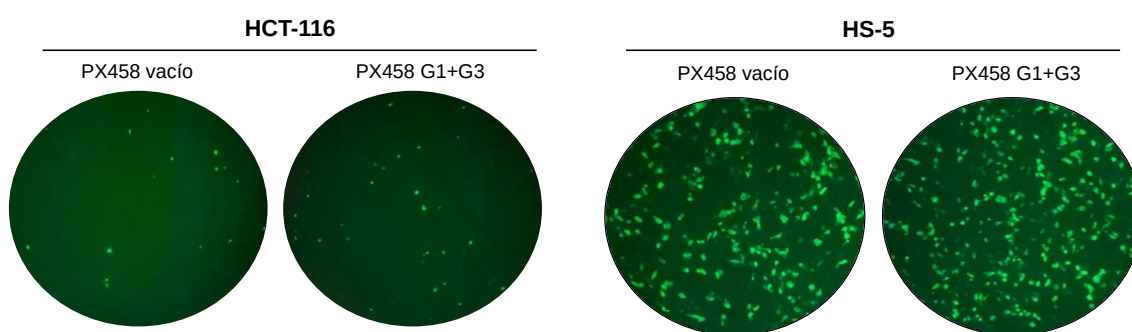


Figura 26. Imágenes de microscopía de fluorescencia de las líneas celulares HCT-116 y HS-5 transfectadas con el plásmido vacío (PX458) y con la combinación de sgRNAs (G1+G3). Se detectan en verde las células transfectadas eficientemente (GFP+).

7.7. Cell sorting mediante citometría de flujo

Tras 48 horas, se llevó a cabo la separación (*cell sorting*) de las células que habían incorporado el plásmido transfectado. Para detectar dicha fluorescencia y separar las células se utilizó el citómetro de flujo FACS Aria™ III (BD Biosciences) usando como marcador de selección la proteína GFP. En este caso, el número de células transfectadas con el plásmido vacío y con la combinación G1+G3 que emitían luz verde era lo suficientemente elevado como para separar la población celular en función de la emisión de GFP. Por ello, en cada condición de las líneas HCT-116 y HS-5 se separó mediante citometría de flujo “pooles” de 80.000 y 100.000 células GFP+, respectivamente. Además, se realizó un *single cell* para individualizar las células transfectadas y así poder obtener poblaciones clonales.

Para ello, se resuspendió el pellet celular de cada condición en 250 μ L de PBS, en un tubo de citometría, y se llevó al equipo en oscuridad y tapado con parafilm. Dado que el

objetivo principal es el de generar líneas celulares *KO* para *NOMO1*, se realizó una separación célula a célula ya que el porcentaje de células editadas en una población 100% GFP+ no suele superar el 5%. Esto es, no todas las células GFP+ serán *KO* para nuestro gen de interés. Por este motivo, se separaron para cada línea celular y cada condición (control y guías 1+3) un total de 3 placas de 96 pocillos con una única célula/pocillo en un volumen final de 150 μ L de medio de cultivo DMEM (20% FBS, 1% penicilina).

Todas las placas se mantuvieron en cultivo durante un mínimo de 3 semanas con el fin de generar poblaciones clonales. En la tabla 15, se indica el porcentaje de células que han incorporado el plásmido (GFP+) y el número de células GFP+ separadas en pool y en *single cell* para cada condición. La condición PX458 se corresponde con la transfección del plásmido vacío, y la condición G1+G3 con la transfección de la combinación de sgRNAs específicos para generar la delección de *NOMO1*. Adicionalmente, se realizó un *single cell* de las dos líneas celulares sin transfectar para comprobar su viabilidad tras la individualización.

Tabla 15. Análisis de la eficiencia de transfección en las líneas celulares HCT-116 y HS-5 mediante citometría de flujo.

Línea celular	HCT-116		HS-5	
Condición	PX458 vacío	PX458 G1+G3	PX458 vacío	PX458 G1+G3
Células GFP+	1%	2%	18%	22%
Pool GFP+ (nº células)	80.000	80.000	100.000	100.000
<i>Single cell</i> (nº placas 96 pocillos)	3	3	3	3
Clones generados	>20	15	>20	>20
Clones con genotipo validado	3	3	3	3

En todos los casos se obtuvieron al menos 15 clones viables. Estos clones se expandieron en placas de 24, 12 y 6 pocillos, hasta finalmente ser cultivados en placas de Petri. Tras ello, se extrajo DNA, RNA y proteínas para validar la edición génica mediante secuenciación Sanger, RT-qPCR y western blot. Es decir, cada clon fue genotipado para confirmar si el sistema CRISPR/cas9 había generado la modificación de interés en *NOMO1* en las células que habían sido transfectadas con la combinación G1+G3, o si conservaban el gen completo en las células transfectadas con el plásmido vacío.

8. Modelo murino para el estudio del cáncer colorrectal de aparición temprana

8.1. Bienestar animal

Los animales utilizados en este trabajo fueron estabulados siguiendo las directrices españolas y de la Unión Europea para la experimentación animal (RD1201/05, RD 53/2013 y 86/609/CEE). Los animales modificados genéticamente se generaron en la instalación de Transgénicos, perteneciente al Servicio de Experimentación Animal de la Universidad de Salamanca. Todos los ratones se estabularon en una instalación libre de patógenos específicos (*specific-pathogen free*, SPF), con temperatura controlada en jaulas de ventilación individual. Todos ellos siguieron una dieta estándar y un ciclo de luz-oscuridad de 12 horas. Los protocolos para la generación y manipulación de los animales utilizados en este trabajo han sido aprobados por el Comité de Bioética de la Universidad de Salamanca y la Consejería de Agricultura y Ganadería de la Junta de Castilla y León (ref 338).

8.2. Generación de un ratón condicional para *Nomo1* mediante la tecnología CRISPR/cas9

La modificación genética en ratones se ha basado durante muchos años en la introducción de mutaciones en células madre embrionarias mediante la electroporación de vectores de recombinación homóloga (HR). Estas mutaciones se microinyectan en un organismo silvestre para generar quimeras. Sin embargo, los avances en los sistemas de edición génica de los últimos años han permitido posicionar la tecnología CRISPR/cas9 como una técnica rápida y eficaz para generar organismos modificados genéticamente²⁰¹. En este sentido, es posible generar modelos murinos *knockout* y *knockin* (KI) para el estudio de numerosas enfermedades. Los animales KO se originan mediante la reparación de la rotura de doble cadena (generada por la cas9) mediante el sistema NHEJ. Los animales KI surgen de la reparación por el sistema HDR, el cual inserta en el genoma mediante HR una secuencia molde de DNA en el punto de corte de la cas9.

Para estudiar el efecto de la delección somática de *Nomo1* en EOCRC se generó un ratón condicional para este gen²⁰², ya que su pérdida durante el desarrollo embrionario es letal. Esta condicionalidad permite inducir su delección en un momento y contexto celular específico. Para ello, se empleó el sistema Cre-*loxP*, que consiste en flanquear una región del genoma con dos secuencias palindrómicas (de 34 pb cada una) que no están presentes en el genoma del ratón (secuencias *loxP*). De esta manera, si la recombinasa Cre es

expresada en la célula, reconoce las secuencias *loxP* que flanquean la región diana y media su escisión mediante recombinación de estas (Figura 27).

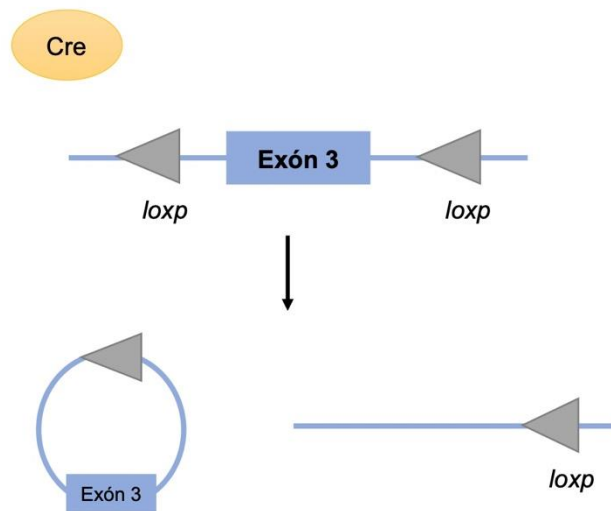


Figura 27. Funcionamiento del sistema Cre-loxP.

En este trabajo, se decidió flanquear el exón 3 de *Nomo1* por 3 motivos fundamentales:

1. El exón 3 es lo suficientemente pequeño (46 pb) para que la recombinasa Cre curse eficientemente.
2. La última base de la secuencia codificante del exón 3 forma parte de un codón cuyas dos bases restantes están en el exón 4 por lo que se rompería el marco de lectura, generando un alelo nulo de *Nomo1*.
3. *Nomo1* tiene 31 exones, por lo que tras la delección del exón 3 quedaría una larga secuencia codificante para generar un codón de parada prematuro.

El sistema Cre-loxP fue introducido en ratones de la cepa C57BL/6J de *Mus musculus*. Para ello, se microinyectaron en los cigotos los componentes necesarios para el funcionamiento del sistema CRISPR/cas9 (Figura 28):

- sgRNA-1: hibrida en la región codificante del exón 3.
- sgRNA-2: hibrida en una región del intrón 3-4.
- mRNA de la cas9.
- Un oligonucleótido de cadena simple (*single strand DNA*; *ssDNA*) que sirve como molde para la recombinación homóloga. Este ssDNA tiene 789 pb de longitud (ssDNA Megamer, Integrated DNA Technologies) y contiene el exón 3 completo de *Nomo1* flanqueado por los dos sitios *loxP*, separados entre sí por 338 pb. Además, contiene dos sustituciones de bases: G>A en el exón 3 y GG>AA en el intrón 3-4.

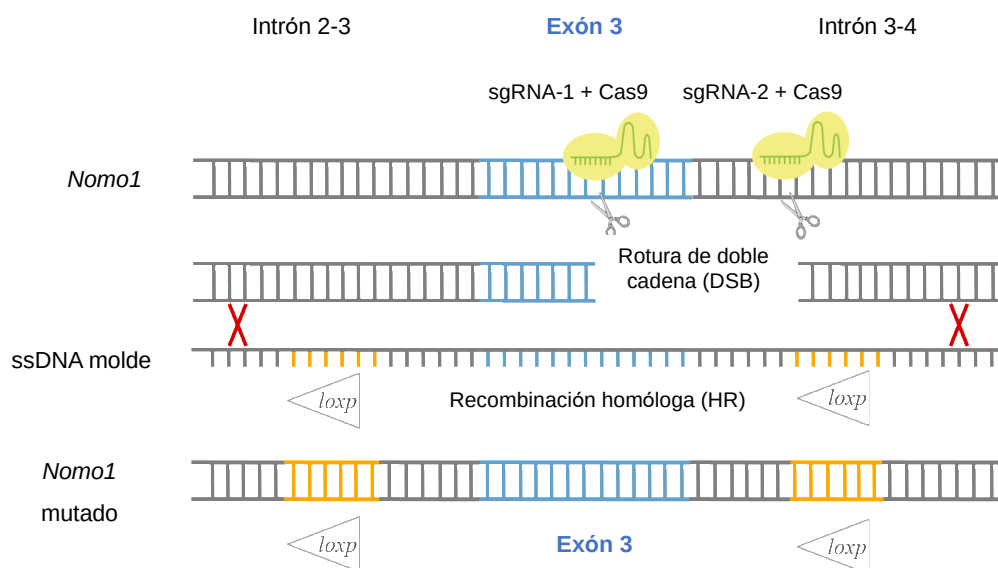


Figura 28. Representación esquemática de la generación de un alelo mutado de *Nomo1* mediante CRISPR/cas9 para el funcionamiento del sistema Cre-loxp.

Los oligonucleótidos específicos para construir los sgRNAs de ratón se diseñaron de la misma manera que en células humanas, cambiando únicamente la especie a la hora de buscar el genoma de referencia. Estos oligonucleótidos se especifican en la tabla 16.

Tabla 16. Secuencias dianas y oligonucleótidos necesarios para la construcción de los sgRNAs de *Nomo1* en ratón.

	Secuencias diana <i>Nomo1</i>	<i>Nomo1</i> oligo-UP	<i>Nomo1</i> oligo-LOW
sgRNA-1	5'-CCTGAAGATCGAACCTCCTC-3'	5'caccgCCTGAAGATC GAACCTCCTC3'	5'aaacGAGGAGGTTC GATCTTCAGGc3'
sgRNA-2	5'-GAGTAACGGGGCGCTTCGCA-3'	5'caccgGAGTAACGG GGCGCTTCGCA-3'	5'aaacTGCGAAGCGC CCCGTTACTCc3'

La clonación de ambas guías se realizó en el vector de expresión pIKO5.sgRNA.EFS.GFP, siguiendo los mismos protocolos que se han utilizado en las líneas celulares. A diferencia de las líneas celulares, los sgRNAs se insertaron en los cigotos en forma de mRNA por lo que fue necesaria una transcripción previa con el kit *MEGAscriptTM kit* (Thermo Fisher), siguiendo las instrucciones del fabricante.

Una vez obtenidos todos los componentes necesarios para el funcionamiento del sistema CRISPR/cas9, se procedió a la generación de los ratones condicionales para *Nomo1*. Para ello, se siguieron las siguientes etapas (Figura 29):

1. Obtención de cigotos de la cepa C57BL/6J.
2. Microinyección de los componentes del sistema CRISPR/cas9: sgRNA-1, sgRNA-2, cas9, DNA molde.
3. Transferencia a hembras pseudogestantes.
4. Genotipado: se extrajo DNA por biopsias de colas de ratón para identificar mediante PCR los distintos alelos generados de *Nomo1*.
5. Transmisión de los alelos generados: una vez identificados los ratones que tenían la mutación esperada en el gen *Nomo1*, se cruzaron con ratones de su misma línea que tenían los dos alelos de *Nomo1* *wild type*.

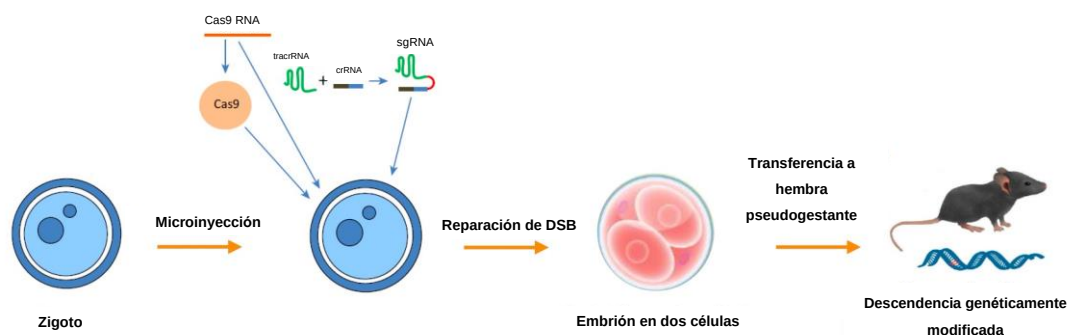


Figura 29. Proceso de edición génica en modelos murinos mediante la tecnología CRISPR/cas9. Adaptado de Williams & Warman, 2017.

Tras la generación del modelo condicional en heterocigosis, se cruzaron estos animales entre sí con el fin de generar ratones *Nomo1^{flox/flox}*, *Nomo1^{flox/wt}* y *Nomo1^{wt/wt}*. Todos estos ratones se mantuvieron en estabulación y fueron utilizados para reproducir la pérdida de *Nomo1* por inducción de la recombinasa Cre y así poder determinar su papel en la génesis del EOCRC.

8.3. Extracción de fibroblastos embrionarios de ratón (MEFs)

Tras el cruce de ratones condicionales en heterocigosis (*Nomo1^{flox/wt}*), se extrajeron los MEFs a los 13,5 días de gestación. Para ello, fue necesario sacrificar las hembras gestantes mediante dislocación cervical siguiendo un protocolo estándar²⁰³. Posteriormente, se extrajeron los embriones individualmente de los cuernos uterinos y se lavaron con PBS en una placa de Petri. Después, los embriones se cortaron en porciones de 1-2 mm y se homogenizaron por pipeteo en 1 mL de tripsina. Esta mezcla se pasó a un eppendorf y se

dejó toda la noche a 4°C. Al día siguiente, se filtraron los embriones y se sembraron por separado en una placa de Petri con medio de cultivo DMEM, suplementado con un 10% de FBS y un 1% de penicilina-estreptomicina. A partir de este punto, se trató el cultivo de MEFs bajo las mismas condiciones que el resto de las líneas celulares empleadas en este trabajo. Se excluyeron de la extracción de MEFs órganos como el hígado, corazón y cerebro.

El tejido cerebral descartado para la extracción de MEFs se utilizó para genotipar cada embrión mediante PCR y así distinguir los tres genotipos posibles: *Nomo1*^{flox/flox}, *Nomo1*^{flox/wt} y *Nomo1*^{wt/wt}. Los oligonucleótidos utilizados para amplificar la región de interés de *Nomo1* en el genotipado de los ratones se especifican en la tabla 4.

8.4. Deleción de *Nomo1* en modelos *in vitro* e *in vivo*

Para reproducir la pérdida de *Nomo1* en MEFs y en el intestino de los ratones condicionales fue necesario inducir la expresión de la recombinasa Cre mediante infección con el adenovirus Ad5CMVCre-eGFP (Viral Vector Core, University of Iowa). Este adenovirus porta la proteína GFP como marcador de selección.

Para validar el funcionamiento del sistema Cre-loxp *in vitro*, se sembraron 100.000 MEFs de cada embrión (*Nomo1*^{flox/flox}, *Nomo1*^{flox/wt} y *Nomo1*^{wt/wt}) en una placa de cultivo de 6 pocillos. Las células se incubaron durante 48 horas y se infectaron con 300 unidades infectivas (*Multiplicity of infection*, MOI) por condición. Para ello, se mezcló 1,5 µL de adenovirus y 2,5 mL de medio de cultivo DMEN sin FBS y se añadió a cada pocillo. Cada embrión se cultivó por duplicado para estudiar una condición infectada y otra sin infectar (control). Tras 48 horas de incubación, se observó la eficiencia de la infección mediante la detección de células GFP+ en un microscopio de fluorescencia. Además, se extrajo DNA y RNA para comprobar el funcionamiento del sistema Cre-loxp (deleción del exón 3 de *Nomo1*) mediante PCR, secuenciación Sanger y RT-qPCR.

Para validar el sistema Cre-loxp *in vivo*, se inyectaron 300 MOI del adenovirus Cre (Ad5CMVCre-eGFP) de forma local en el intestino de 9 ratones de 8 semanas de edad: 6 *Nomo1*^{flox/flox}, 2 *Nomo1*^{flox/wt} y 1 *Nomo1*^{wt/wt}. En paralelo, se estudió 1 ratón de cada genotipo sin infectar como grupo control. Todos los grupos de ratones incluidos en el estudio se observaron durante 20 meses tras la activación de la Cre, con el fin de identificar la formación de tumores o alguna otra patología. En ese momento, se sacrificaron por dislocación cervical y se inspeccionaron a nivel macroscópico. Además, en el momento de sacrificar cada animal se extrajo DNA y RNA del tubo digestivo para validar el funcionamiento del sistema Cre-loxp mediante PCR, secuenciación Sanger y RT-qPCR.

Los oligonucleótidos empleados para estudiar la delección de *Nomo1* a nivel de DNA genómico y de expresión se especifican en las tablas 4 y 7, respectivamente.

8.5. Generación de un ratón condicional *Nomo1* Flox / Msi1-Cre

La delección de *Nomo1* *in vivo* también se puede inducir por la expresión de la recombinasa Cre de manera selectiva en algunos tejidos. En este sentido, se han empleado ratones Musashi-1-Cre (Msi1-Cre), los cuales expresan la Cre de manera constitutiva en las criptas intestinales, pulmón, cerebro y piel. La expresión de esta recombinasa está controlada por el promotor Msi1, un marcador de células madre neuronales que tiene un papel importante en el mantenimiento del estado *stem*, la diferenciación y la tumorigénesis²⁰⁴.

Para ello, se cruzaron 3 ratones condicionales para *Nomo1* en heterocigosis (*Nomo1*^{flox/wt}) con 3 ratones Msi1-Cre +. De esta manera, se generaron ratones con la delección de *Nomo1* en heterocigosis desde estadios tempranos en las criptas intestinales. La delección homocigota de *Nomo1* no pudo ser reproducida en este modelo dado que la pérdida en homocigosis de este gen es letal durante el desarrollo embrionario. En total, se estudiaron 3 ratones con la delección heterocigota (*Nomo1*^{flox/wt} Cre+) y un ratón condicional en heterocigosis sin la delección que fue utilizado como control (*Nomo1*^{flox/wt} Cre-). Todos los animales permanecieron en observación durante 18 meses con el fin de identificar la formación de tumores o alguna otra patología, momento en el que fueron sacrificados por dislocación cervical para su estudio molecular y anatomopatológico. De cada animal se extrajo DNA y RNA del tubo digestivo para validar el funcionamiento del sistema Cre-loxp mediante PCR. Además, se extrajo material genómico de otros órganos donde se observasen indicios de patología.

8.6. Inducción de carcinogénesis colorrectal en ratones *Nomo1* Flox / Msi1-Cre

La 1,2-Dimetilhidrazina (DMH) es un potente agente carcinógeno que se ha utilizado durante muchos años para reproducir modelos de carcinogénesis colorrectal en roedores de laboratorio^{205–208}. Además, la formación de tumores en el colon de ratones inducida por DMH guarda una estrecha similitud con el cáncer colorrectal en humanos en cuanto a su morfología, patrones de crecimiento y manifestaciones clínicas²⁰⁹. El DMH es un agente alquilante que incorpora grupos metilo en el DNA, generando mutaciones puntuales que afectan a los procesos de proliferación, diferenciación y apoptosis en las células epiteliales del colon^{210–212}.

Por este motivo, en este trabajo hemos decidido estudiar el efecto de la delección de *Nomo1* en un contexto de carcinogénesis colorrectal generado por sucesivas inyecciones semanales de una concentración constante de DMH (Sigma-Aldrich). Para ello, se han seguido protocolos descritos anteriormente como modelo^{213,214}. En primer lugar, se han dividido un total de 15 ratones de 8 semanas de edad en 3 grupos acorde a su genotipo:

- a) 4 ratones *Nomo1*^{flox/wt}; Msi1-Cre + (delección heterocigota).
- b) 6 ratones *Nomo1*^{flox/wt}; Msi1-Cre- (grupo control 1).
- c) 5 ratones *Nomo1*^{wt/wt} (grupo control 2).

A todos los ratones de cada grupo se les administró de forma intraperitoneal una dosis semanal de 20 mg/kg durante un periodo de 20 semanas. El DMH fue diluido cada semana en EDTA 1 mM ajustado a pH 6,5 con NaOH 8 N. El volumen final de inyección fue de 250 μ L. A la semana 22, todos los ratones fueron sacrificados por dislocación cervical excepto en los casos donde se desarrollase alguna patología previamente que hiciera necesario sacrificar el animal. Finalmente, se observó si existían formaciones tumorales o alguna otra patología y se extrajo el tubo digestivo para su análisis anatomopatológico y molecular

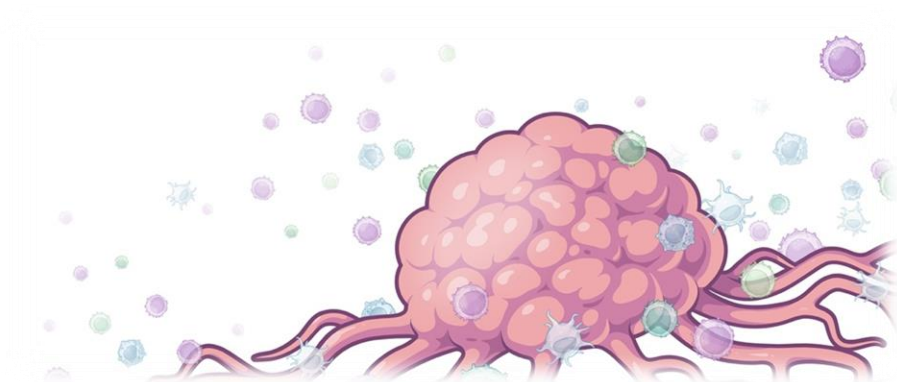
8.7. Estudio anatomopatológico

Las muestras obtenidas del tubo digestivo de todos los ratones fueron fijadas en PBS y se llevaron al Servicio de Patología Molecular Comparada del Centro de Investigación del Cáncer (IBMCC-CSIC) para su inclusión en bloques de parafina. Los estudios patológicos fueron realizados sobre secciones teñidas con hematoxilina-eosina con la colaboración del Servicio de Anatomía Patológica del Hospital Universitario de Salamanca.

9. Análisis estadístico

El análisis estadístico se realizó con el programa informático IBM/SPSS® Statistics v.26 (SPSS Inc., Chicago, IL). Se consideró la existencia de significación estadística cuando se halló un p-valor inferior a 0,05. Las variables cuantitativas se describieron en términos de tendencia central (media) y dispersión (desviación estándar). Esto aplica para los estudios de viabilidad celular, cuantificaciones de western blot, migración celular, estado *stem cell* y longitud telomérica. Las diferencias entre los grupos comparados se analizaron mediante la prueba t de Student cuando tenían una distribución normal validada mediante la prueba Kolmogorov-Smirnov (p-valor >0,05). En el caso de variables clínicas cualitativas, se analizó la distribución mediante la prueba χ^2 de Pearson.

Resultados



Resultados

CAPÍTULO I. CARACTERIZACIÓN FUNCIONAL DE LA PÉRDIDA DE *NOMO1* EN CÁNCER COLORRECTAL DE APARICIÓN TEMPRANA

1. Generación de líneas celulares *knockout* para *NOMO1* mediante el sistema de edición génica CRISPR/cas9

1.1. Validación de la edición génica de las líneas celulares HCT-116 y HS-5

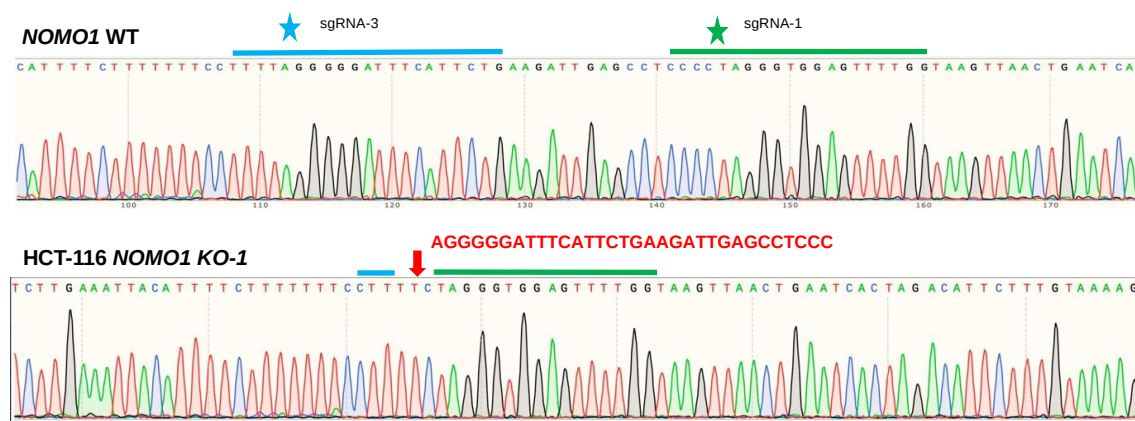
La validación de la edición génica generada por el sistema CRISPR/cas9 se realizó a nivel de DNA genómico y de expresión proteica. Para ello, primero fue necesario generar clones viables de cada condición de estudio, y extraer DNA y proteína de cada uno de ellos mediante los protocolos detallados en la metodología.

- **Confirmación mediante secuenciación Sanger:**

La extracción de DNA de los clones generados se llevó a cabo de las líneas celulares HCT-116 y HS-5 transfectadas con el plásmido vacío, con la combinación de guías G1+G3 y de las células sin transfectar.

En primer lugar, se amplificó mediante PCR la región diana del gen *NOMO1*, que comprende todo su exón 3 y parte de las regiones intrónicas adyacentes. De esta manera, se pudo observar como la endonucleasa cas9 había sido eficiente generando un corte de doble cadena en los puntos de corte marcados por las guías 1 y 3, localizados al inicio y final del exón 3. Estos productos de PCR se observaron en un gel de agarosa, se purificaron, y se secuenciaron mediante el método de Sanger (Figura 30).

A



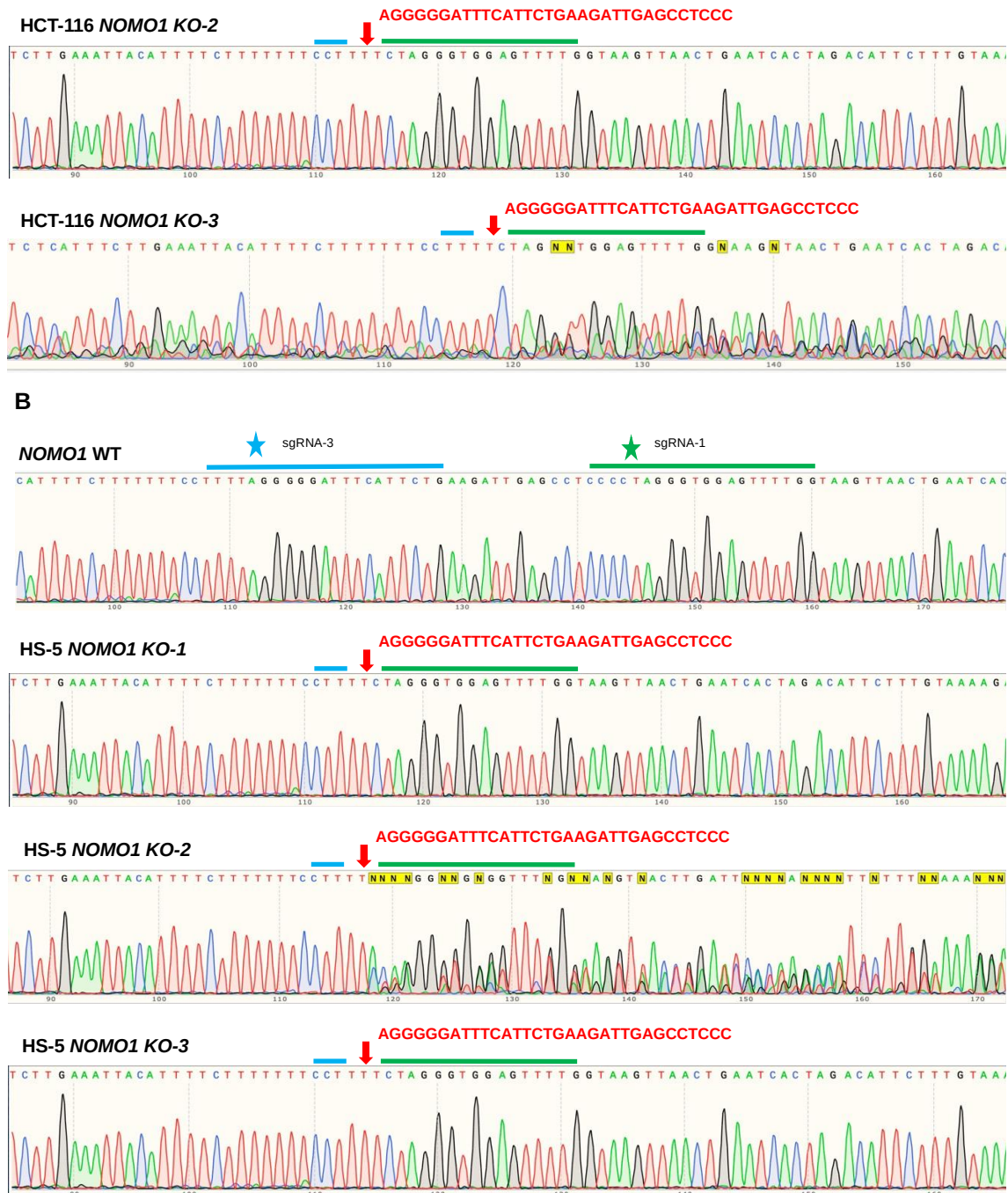


Figura 30. Secuencias de nucleótidos correspondientes al exón 3 de *NOMO1* y sus regiones intrónicas adyacentes. Se observan las secuencias modificadas tras el corte generado por la cas9 en las posiciones marcadas por las guías 1 y 3 (estrellas verde y azul). A) Secuencia WT y *NOMO1*-KO de los clones de la línea HCT-116. B) Secuencias WT y *NOMO1*-KO de los clones de la línea HS-5.

Así, se comprobó que el sistema CRISPR/cas9 había sido eficiente, generando la delección de la región comprendida entre las guías 1 y 3 (32 pb) en 3 clones de la línea HCT-116 y 3 clones de la línea HS-5. Además, se observó que el estado de *NOMO1* permanecía

germinal en todas las condiciones transfectadas con el plásmido vacío y en las células sin transfectar.

- **Confirmación mediante qPCR:**

Una vez observado que las guías 1 y 3 habían funcionado correctamente para escindir la región de interés del exón 3 de *NOMO1*, se validó la generación de los clones mediante la técnica de PCR cuantitativa en tiempo real (qPCR). En este sentido, se realizó la qPCR para 2 clones control y 3 clones *knockout* de cada línea celular. Así, la secuencia de DNA amplificada en la condición control se correspondía con la región delecionada, por lo que los clones *NOMO1-KO* no amplificarían ningún fragmento. Para ello, fue necesario utilizar los oligonucleótidos específicos de cada sgRNA (sgRNA-1 y sgRNA-3).

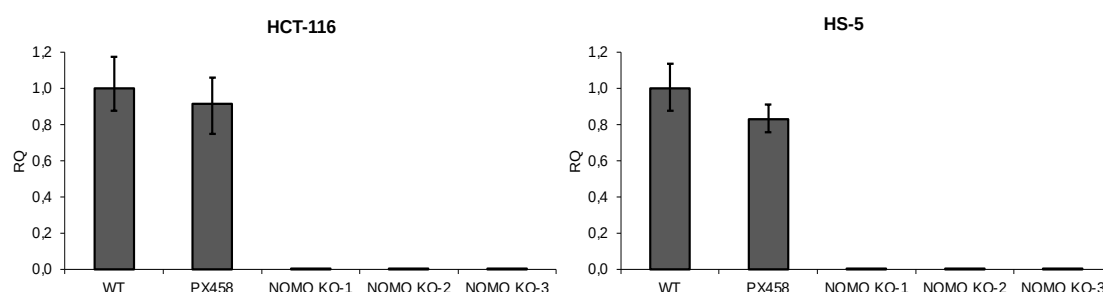


Figura 31. Normalización de la cantidad de DNA amplificado para el fragmento de interés del gen *NOMO1* mediante PCR cuantitativa en tiempo real. Se representan 2 clones control y 3 clones transfectados con la combinación G1+G3 para cada línea celular.

Como se muestra en la figura 31, los 3 clones *NOMO1-KO* de la línea HCT-116 y los 3 clones *NOMO1-KO* de la línea HS-5 no amplificaron ningún fragmento de DNA. Esto confirma que la edición a nivel genómico del exón 3 del gen *NOMO1*, producida por el sistema CRISPR/cas9, había sido la esperada.

- **Confirmación mediante western blot:**

Finalmente, se estudió la expresión proteica de NOMO mediante western blot, con el objetivo de confirmar si la deleción generada y validada a nivel genómico provocaba la ruptura del marco de lectura e inactivación de la proteína. Para esta validación, se utilizaron 2 clones control y 2 clones *NOMO1-KO* de cada línea celular, ya que serían los utilizados para la caracterización funcional en experimentos posteriores. Como muestra la figura 32, las condiciones control en la línea HCT-116 y HS-5 mostraron expresión de NOMO, mientras que en los clones editados genéticamente no había expresión de dicha proteína.

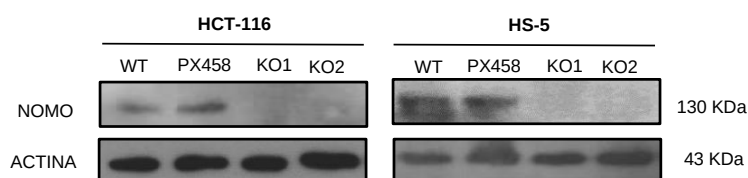


Figura 32. Análisis de la expresión proteica de NOMO en las líneas celulares HCT-116 y HS-5. Se representa la expresión de NOMO en los clones control y en los clones editados mediante CRISPR/cas9.

Por tanto, teniendo en cuenta las diferentes técnicas de validación empleadas, confirmamos la correcta generación de dos líneas celulares (HCT-116 y HS-5) *knockout* para *NOMO1* con el uso de la herramienta de edición génica CRISPR/cas9.

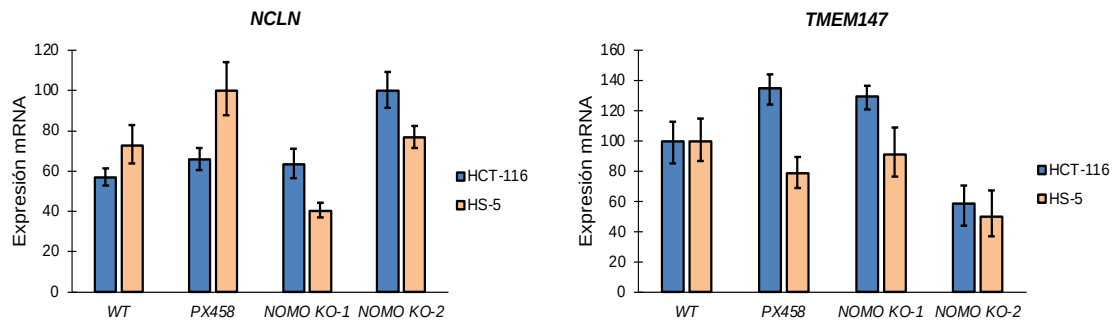
2. Caracterización fenotípica de las líneas celulares *NOMO1 knockout*

2.1. Análisis de expresión del complejo NOMO/NCLN/TMEM147

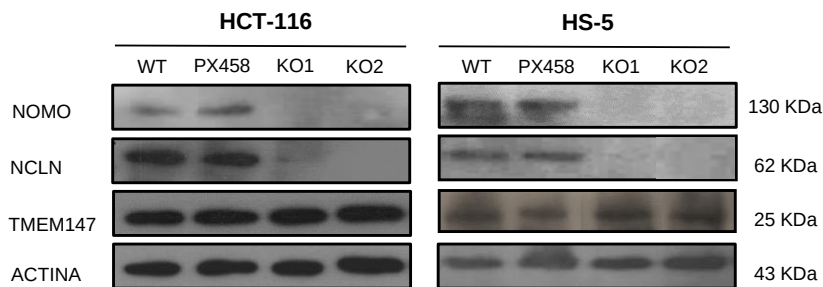
Diversos autores han descrito que las proteínas NOMO, NCLN y TMEM147 forman un complejo proteico localizado en el retículo endoplásmico que ejerce de antagonista de la vía de señalización Nodal durante el desarrollo embrionario. Además, estos estudios han demostrado que NCLN y NOMO se vuelven inestables y disminuyen su expresión en ausencia de alguno de los dos miembros. Por tanto, estas dos proteínas tendrían un efecto estabilizador del complejo. Por este motivo, tras la generación de las líneas celulares HCT-116 y HS-5 con inactivación de NOMO, quisimos estudiar los niveles de expresión de los miembros del complejo a nivel de RNA y proteína, mediante RT-qPCR y western blot, respectivamente. Para cada línea celular, se estudiaron 2 clones NOMO-WT y 2 clones NOMO-KO.

Como se observa en la figura 33, tal y como indican los estudios previos, la expresión de NCLN se redujo significativamente en los clones NOMO-KO de las dos líneas celulares. Sin embargo, esta reducción de NCLN a nivel proteico no fue acompañada por una disminución de los niveles de ARN mensajero, sugiriendo que la pérdida de expresión es causada por un mecanismo post-transcripcional. Del mismo modo, otros autores han demostrado que la expresión de TMEM147 se reduce significativamente en ausencia de NCLN y NOMO. Contrariamente, en este trabajo demostramos por primera vez que los niveles de expresión de TMEM147 no se ven afectados tras la inactivación de NOMO en las líneas celulares incluidas en este estudio. Por lo tanto, la pérdida de NOMO desestabiliza el complejo formado con NCLN, pero no afecta a la expresión de TMEM147.

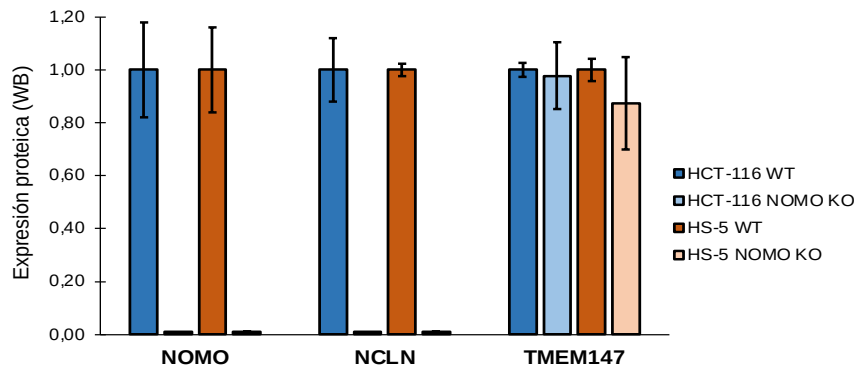
A



B



C



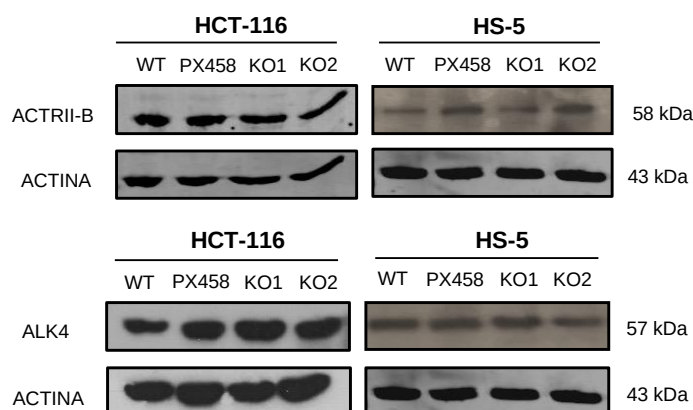
Cuantificación WB								
	HCT-116 WT	DS	HCT-116 NOMO KO	DS	HS-5 WT	DS	HS-5 NOMO KO	DS
NOMO	1,00	0,18	0,01	-	1,00	0,16	0,01	-
NCLN	1,00	0,12	0,01	-	1,00	0,02	0,01	-
TMEM147	1,00	0,03	0,98	0,13	1,00	0,04	0,87	0,17

Figura 33. Niveles de expresión de los miembros del complejo NOMO/NCLN/TMEM147 en las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO. A) Niveles de mRNA de *NCLN* y *TMEM147*. B) Niveles de expresión proteica de NOMO, NCLN y TMEM147 detectados mediante western blot. C) Cuantificación y normalización de los niveles de expresión proteica del complejo NOMO/NCLN/TMEM147. Se han cuantificado dos clones de cada condición y se representa su promedio. DS: desviación estándar.

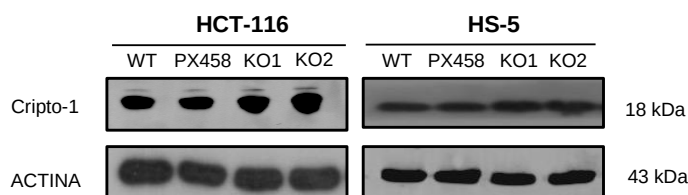
2.2. Análisis de expresión proteica de los receptores de la vía de señalización Nodal

Se ha demostrado que el complejo NOMO/NCLN/TMEM147 modula la señalización de la vía Nodal durante las primeras etapas del desarrollo embrionario. Aunque se sabe que ejerce un papel antagonista de la vía, actualmente se desconoce el mecanismo implicado en este proceso. Por ello, nos propusimos estudiar, en las dos líneas celulares carentes de NOMO generadas en este trabajo, si la inactivación de NOMO afectaba a la expresión de varios miembros de la vía de señalización Nodal. En primer lugar, se cuantificó la expresión proteica de los receptores ACTRII-B y ALK4, y del correceptor Cripto-1 mediante western blot. Para cada línea celular, se compararon 2 clones NOMO-WT y 2 clones NOMO-KO (Figura 34).

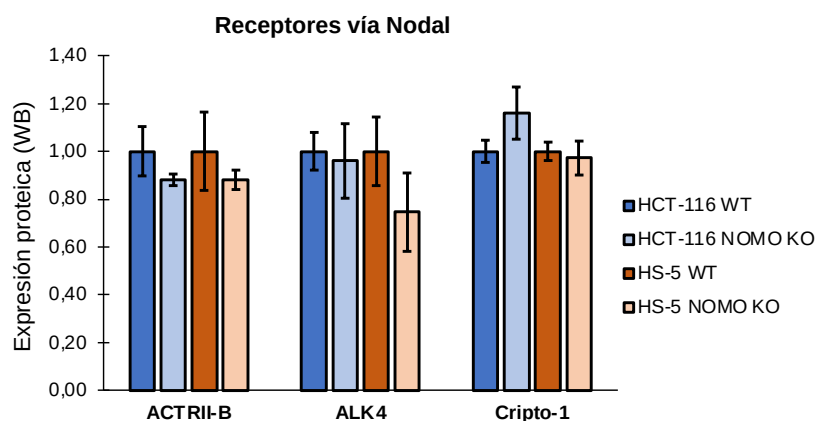
A



B



C



Cuantificación WB								
	HCT-116 WT	DS	HCT-116 NOMO KO	DS	HS-5 WT	DS	HS-5 NOMO KO	DS
ACTRII-B	1,00	0,10	0,88	0,03	1,00	0,16	0,88	0,04
ALK4	1,00	0,08	0,96	0,16	1,00	0,14	0,75	0,16
Cripto-1	1,00	0,05	1,16	0,11	1,00	0,04	0,97	0,07

Figura 34. Niveles de expresión proteica de los receptores (ACTRII-B y ALK4) y del correceptor (Cripto-1) de la vía de señalización Nodal en las líneas celulares HCT-116 y HS-5. A) Niveles de expresión de ACTRII-B y ALK4 mediante western blot. B) Niveles de expresión proteica del correceptor Cripto-1. C) Cuantificación y normalización de los niveles de expresión proteica de los receptores y del correceptor de la vía Nodal. Se han cuantificado dos clones de cada condición y se representa su promedio. DS: desviación estándar.

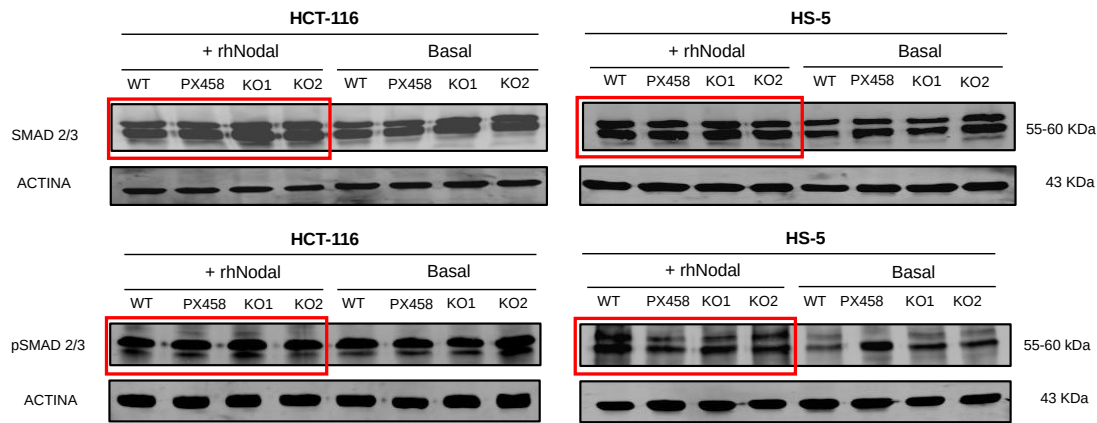
En este sentido, no se observaron alteraciones en los niveles de expresión proteica de ACTRII-B, ALK4 y Cripto-1 tras la inactivación de NOMO en ninguno de los clones analizados de las dos líneas celulares. Esto sugiere que NOMO no afecta a la expresión de los receptores y correceptores de la vía de señalización Nodal en células adultas, en consonancia con resultados previos obtenidos en nuestro laboratorio.

2.3. Estudio de la actividad de la vía Nodal

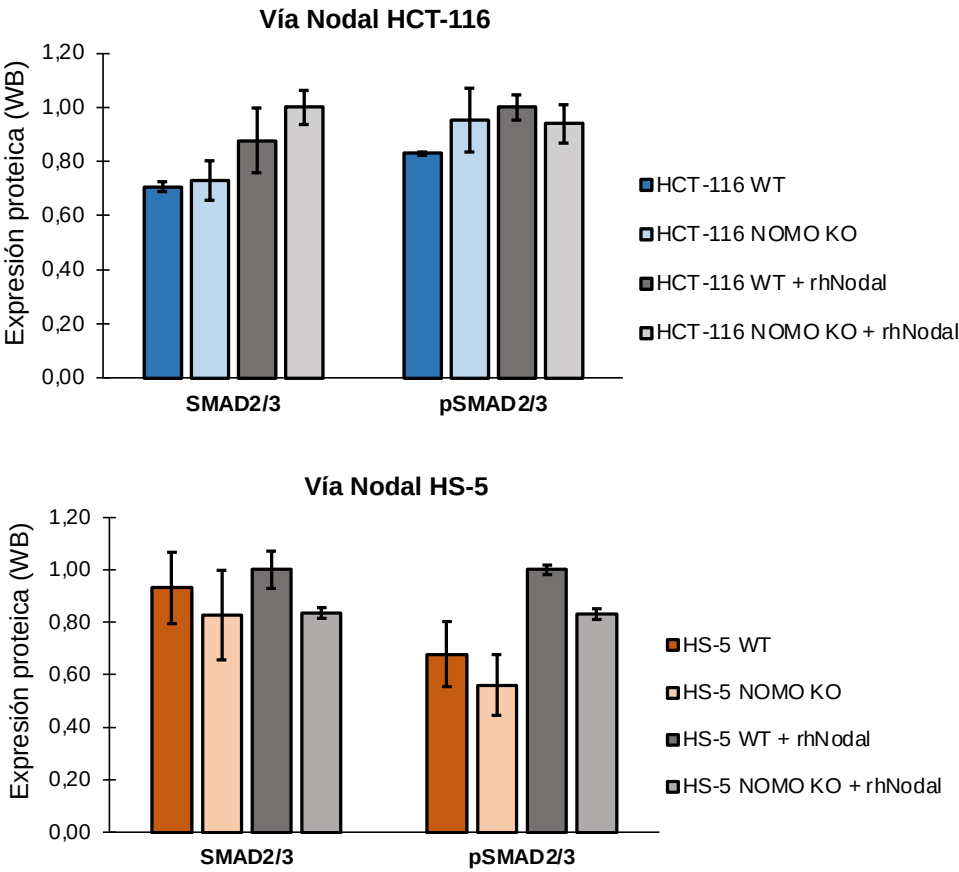
Se ha descrito en numerosas ocasiones que la actividad de la vía Nodal se ve modificada en diversos tipos tumorales. Así, una vez los receptores son activados por el ligando Nodal, éstos fosforilan a las proteínas SMAD2 y SMAD3 que se unen a un mediador común, la proteína SMAD4 (Co-Smad). Cuando SMAD4 se activa, forma un complejo transcripcional con SMAD2/3, el cual se transloca al núcleo para activar a distintos factores de transcripción. Por este motivo, quisimos estudiar si la inactivación de NOMO afectaba a la actividad de la vía Nodal alterando la expresión de las proteínas que integran la vía de las SMADs.

En primer lugar, analizamos la expresión de SMAD2 y SMAD3 a nivel endógeno y fosforilado, mediante western blot. El estado de estas dos proteínas se cuantificó tanto en condiciones basales como en condiciones de activación de la ruta. Para ello, antes de realizar la extracción de proteínas, las células se incubaron durante 24 horas con el ligando Nodal (rhNodal). Como se puede observar en la figura 35, los niveles de expresión de SMAD2 y SMAD3 fue similar en los clones NOMO-WT y NOMO-KO, en condiciones basales y tras activación de la vía con el ligando Nodal. Además, el estado de fosforilación de estas dos proteínas tampoco se vio alterado tras la inactivación de NOMO en ninguna de las dos líneas celulares.

A



B

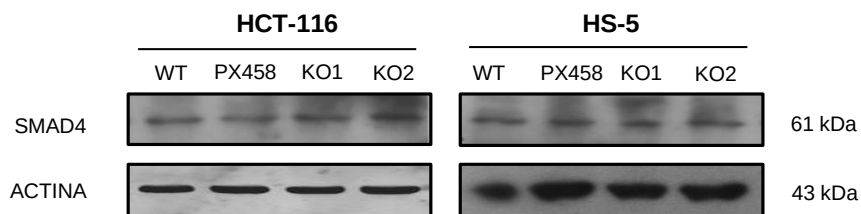


Cuantificación WB								
	HCT-116 WT	DS	HCT-116 WT + rhNodal	DS	HCT-116 NOMO KO	DS	HCT-116 NOMO KO + rhNodal	DS
SMAD2/3	0,70	0,02	0,87	0,12	0,73	0,07	1,00	0,06
pSMAD2/3	0,83	0,01	1,00	0,04	0,95	0,11	0,94	0,07
	HS-5 WT	DS	HS-5 WT + rhNodal	DS	HS-5 NOMO KO	DS	HS-5 NOMO KO + rhNodal	DS
SMAD2/3	0,93	0,14	1,00	0,07	0,83	0,17	0,84	0,02
pSMAD2/3	0,68	0,12	1,00	0,02	0,56	0,11	0,83	0,02

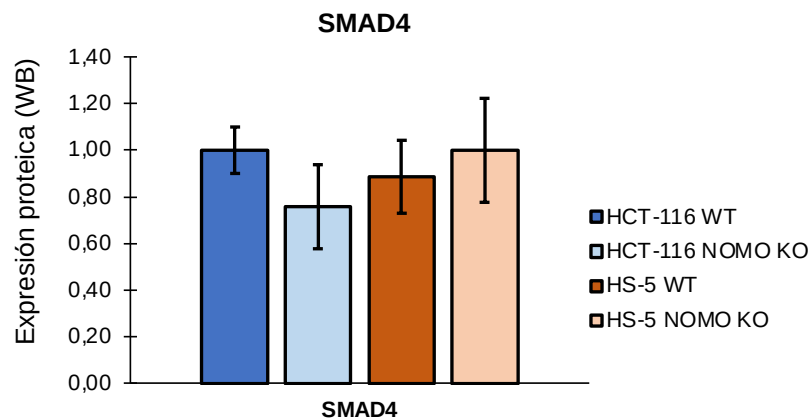
Figura 35. Niveles de expresión de las proteínas SMAD2, SMAD3, pSMAD2 y pSDMAD3 en los clones NOMO-WT y NOMO-KO de las líneas celulares HCT-116 y HS-5. A) Niveles de expresión de SMAD2/3 y pSMAD2/3 en condiciones basales y con activación de la vía tras incubación con el ligando nodal (rhNodal) durante 24 horas. B) Cuantificación y normalización de los niveles de expresión de las proteínas SMAD2/3 y pSMAD2/3. Se han cuantificado dos clones de cada condición y se representa su promedio. DS: desviación estándar.

Posteriormente, se estudió el estado de SMAD4 en condiciones basales, con el objetivo de ver si la inactivación de NOMO generaba algún cambio en su expresión. En este caso, la pérdida de NOMO tampoco parece influir en los niveles de expresión proteica de SMAD4 (Figura 36).

A



B



Cuantificación WB								
	HCT-116 WT	DS	HCT-116 NOMO KO	DS	HS-5 WT	DS	HS-5 NOMO KO	DS
SMAD4	1,00	0,10	0,76	0,18	0,89	0,16	1,00	0,22

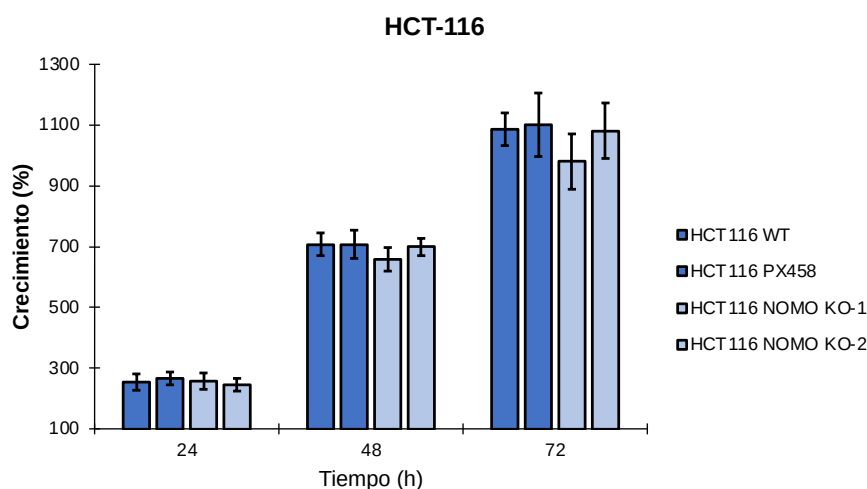
Figura 36. Niveles de expresión proteica de SMAD4 en los clones NOMO-WT y NOMO-KO de las líneas celulares HCT-116 y HS-5. A) Western blot de SMAD4. B) Cuantificación y normalización de los niveles de expresión proteica de SMAD4. Se han cuantificado dos clones de cada condición y se representa su promedio. DS: desviación estándar.

Estos resultados confirman que la actividad de la vía Nodal no se ve afectada por la pérdida de *NOMO1*. Sin embargo, el complejo formado por NCLN y NOMO si parece desestabilizarse. En este contexto, *NOMO1* podría estar desempeñando su papel a través de otra vía de señalización no conocida hasta el momento.

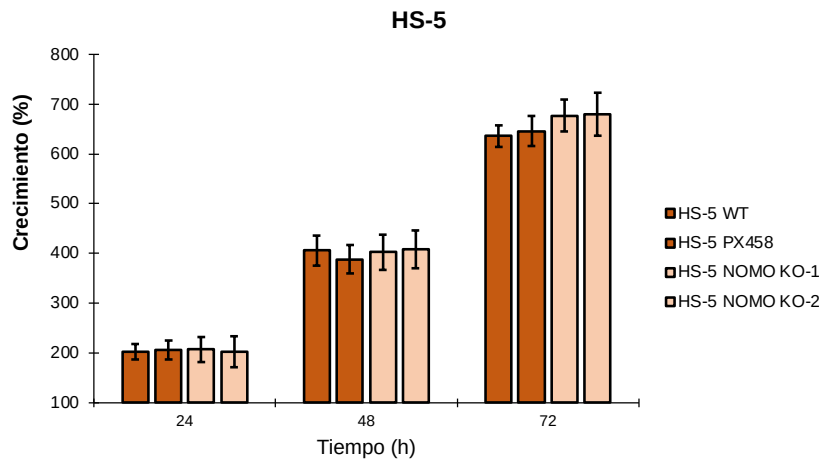
2.4. Estudio de la viabilidad celular

Independientemente de los mecanismos moleculares alterados tras la pérdida de *NOMO1*, quisimos determinar si la inactivación de este gen afectaba a la proliferación celular. Para ello, se cultivaron 2 clones NOMO-WT y 2 clones NOMO-KO de las líneas HCT-116 y HS-5, y se evaluó mediante MTT su tasa de crecimiento a 0, 24, 48 y 72 horas. Este ensayo se realizó por triplicado, donde cada una de estas réplicas biológicas constó de 3 réplicas técnicas. De esta manera, cada clon tuvo 9 mediciones a cada tiempo. Las tasas de crecimiento de los clones *knockout* no mostraron diferencias significativas (p-valor >0,05) respecto de los clones control (Figura 37). Por tanto, la inactivación de NOMO no modifica la viabilidad celular de las líneas HCT-116 y HS-5.

A



	HCT-116 WT		HCT-116 PX458		HCT-116 NOMO KO-1		HCT-116 NOMO KO-2	
Tiempo (h)	Crecimiento (%)	DS	Crecimiento (%)	DS	Crecimiento (%)	DS	Crecimiento (%)	DS
0	100,00	-	100,00	-	100,00	-	100,00	-
24	255,35	27,34	266,96	21,53	258,04	26,09	245,90	21,60
48	706,97	37,34	706,49	46,03	659,01	38,63	699,29	27,96
72	1084,94	53,17	1101,01	104,61	980,31	90,69	1080,79	91,98

B

	HS-5 WT		HS-5 PX458		HS-5 NOMO KO-1		HS-5 NOMO KO-2	
Tiempo (h)	Crecimiento (%)	DS	Crecimiento (%)	DS	Crecimiento (%)	DS	Crecimiento (%)	DS
0	100,00	-	100,00	-	100,00	-	100,00	-
24	255,35	27,34	266,96	21,53	258,04	26,09	245,90	21,60
48	706,97	37,34	706,49	46,03	659,01	38,63	699,29	27,96
72	1084,94	53,17	1101,01	104,61	980,31	90,69	1080,79	91,98

Figura 37. Efecto de la inactivación de NOMO sobre la viabilidad celular. A) MTT de la línea celular HCT-116. B) MTT de la línea celular HS-5. DS: desviación estándar.

2.5. Estudio del estado *stem cell*

Para estudiar si la inactivación de NOMO afectaba a la diferenciación celular, se realizó un análisis de expresión de los marcadores de membrana CD44 y CD24. Estas dos proteínas de superficie celular se consideran marcadores de células madre de cáncer de colon, y se asocian principalmente a procesos de adhesión celular, migración y metástasis. Para ello, se compararon 2 clones NOMO-WT y 2 clones NOMO-KO de cada línea celular mediante citometría de flujo. En total, se contabilizaron 50.000 eventos para cada clon y se tuvieron en cuenta para el análisis los porcentajes de células vivas con expresión de CD44 o CD24 (Figura 38). La expresión de dichos marcadores en las líneas celulares HCT-116 y HS-5 no presentó diferencias significativas (p -valor $>0,05$) en función del estado de NOMO.

Es decir, las células NOMO-KO no ofrecen indicios de dediferenciación como sucede en las células madre iniciadoras de la carcinogénesis colorrectal.

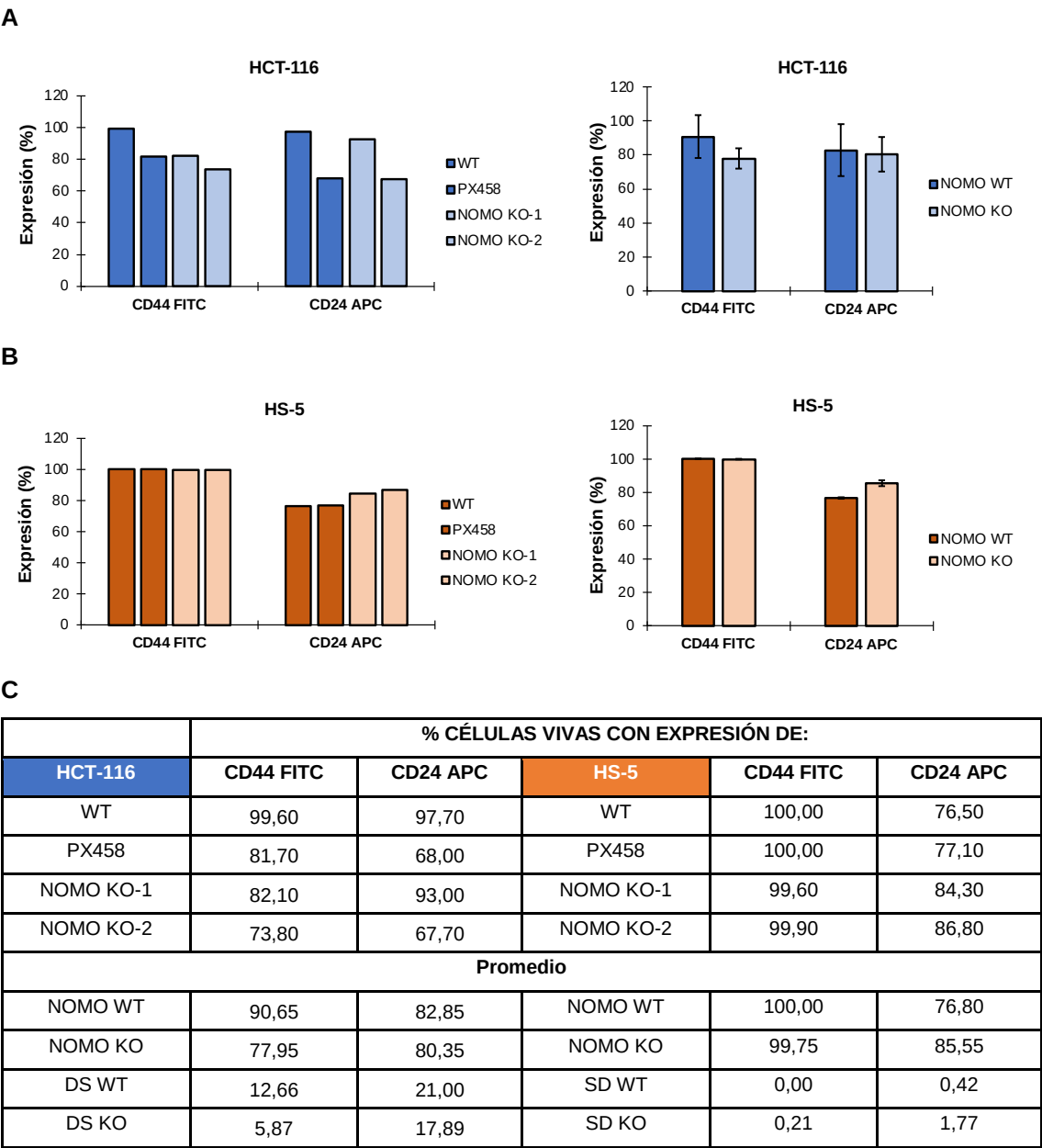


Figura 38. Expresión de los marcadores de membrana CD44 y CD24 en las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO. A) Expresión de los clones individuales y agrupados por genotipo de la línea HCT-116. B) Expresión de los clones individuales y agrupados por genotipo de la línea HS-5. C) Cuantificación normalizada de CD44 y CD24 detectada mediante citometría de flujo en las células NOMO-WT y NOMO-KO de ambas líneas celulares. DS: desviación estándar.

3. Estudio del perfil transcriptómico y proteómico de las líneas celulares carentes de *NOMO1*

Dado que los resultados obtenidos hasta el momento indicaban que la inactivación del *NOMO1* no afecta a la actividad de la vía de señalización Nodal, quisimos estudiar si la pérdida de este gen modificaba los perfiles de expresión génica y proteica. De esta manera, podríamos identificar genes y/o rutas de señalización alteradas como consecuencia de la pérdida de *NOMO1* que pudieran explicar la génesis del EOCRC.

3.1. Estudio del transcriptoma

Para estudiar el transcriptoma, se extrajo RNA total de los clones *NOMO1-KO* y *NOMO1-WT* de las dos líneas celulares y se analizó el perfil de expresión mediante microarrays y RNA-seq. Para ambos análisis, se consideró un FC superior a 1,5 para los genes sobre-expresados y un FC inferior a -1,5 para los genes infra-expresados. Los datos crudos obtenidos en ambos ensayos se han depositado en el repositorio “*Gene Expression Omnibus (GEO)*” con el identificador de acceso GSE198383.

- **Microarrays de expresión:**

El análisis del perfil transcripcional se llevó a cabo mediante microarrays de expresión, comparando 4 clones *NOMO1-KO* y 4 clones *NOMO1-WT*. Esto es, se incluyeron en el estudio 2 clones *wild type* y 2 clones *knockout* de las líneas celulares HCT-116 y HS-5. De esta manera, se eliminaría el efecto del fondo genético de cada línea, por lo que los cambios detectados en el perfil de expresión génica se deberían únicamente a la pérdida de *NOMO1*.

Tras el análisis de los datos aportados por el microarray, se pudieron identificar un total de 126 genes desregulados ($FC > 1,5$; $FC < -1,5$), debido a la inactivación de *NOMO1*, en común a las dos líneas celulares. De estos 126 genes, 81 se encontraban sobre-expresados ($FC > 1,5$) y 45 se presentaron de forma infra-expresada ($FC < -1,5$). Así, se puede confirmar que la pérdida de *NOMO1* genera un perfil transcripcional diferencial en las líneas HCT-116 y HS-5 (Figura 39).

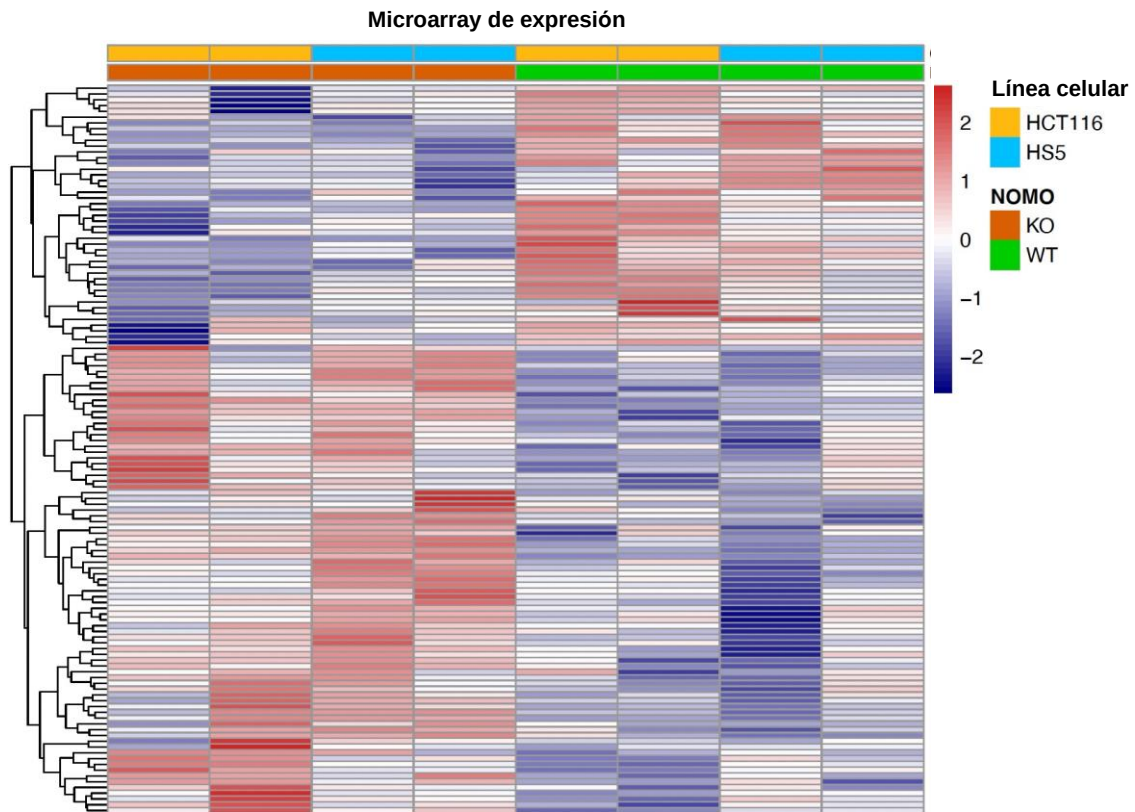


Figura 39. Perfil de expresión diferencial de las líneas celulares HCT-116 y HS-5 con inactivación de *NOMO1*. Representación mediante *heatmap* de los genes comunes desregulados en el microarray de expresión ($FC > 1,5$; $FC < -1,5$).

Posteriormente, se realizó un análisis de enriquecimiento de rutas (GSEA) utilizando la herramienta *WebGestalt*. Para ello, se introdujo en el sistema los 126 genes desregulados que se identificaron en el microarray tras la inactivación de *NOMO1*. Además, se añadió el FC de cada gen con el fin de reflejar en el análisis de enriquecimiento la magnitud de su cambio de expresión. Como resultado, se obtuvieron un total de 18 rutas de señalización afectadas por la pérdida de *NOMO1*, tras contrastar el set de genes desregulados con la base de datos “*hallmarks*” (Figura 40). El análisis de enriquecimiento con las bases de datos “*KEGG*” y “*Reactome*” se detallan en el anexo 1.

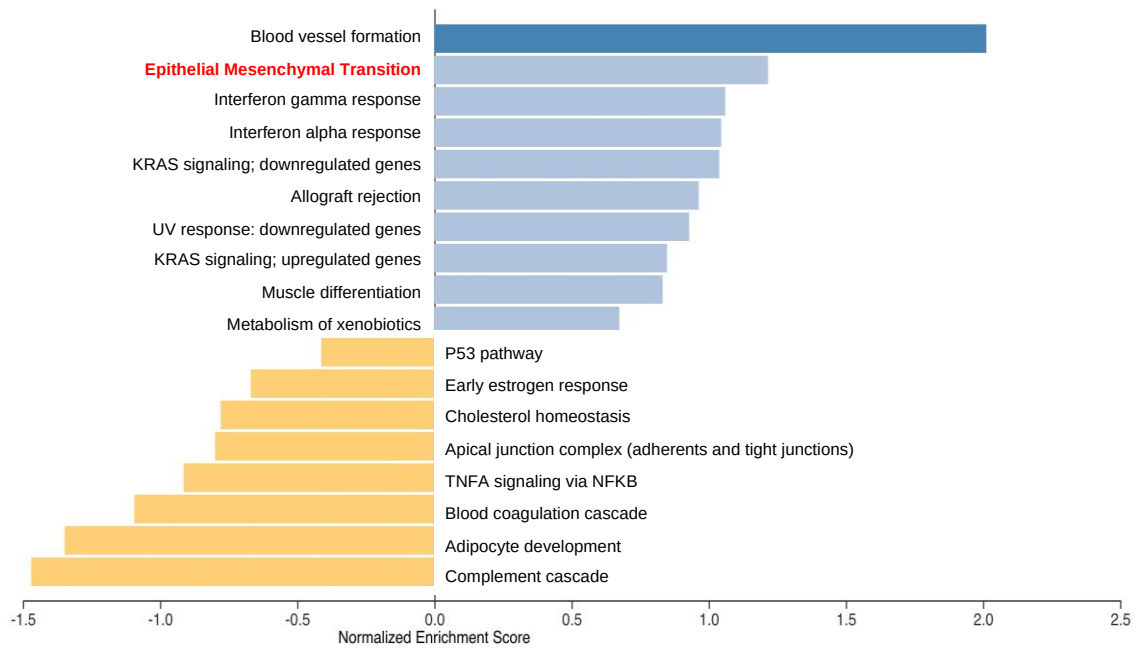


Figura 40. Procesos biológicos desregulados asociados a la pérdida de *NOMO1* en las líneas celulares HCT-116 y HS-5. Se representan los procesos alterados al introducir el set de genes desregulados en el microarray de expresión en el análisis GSEA-hallmarks ($FC > 1,5$; $FC < -1,5$).

- **Secuenciación de RNA:**

Para el análisis del perfil transcriptómico mediante RNA-seq, se compararon un total de 6 clones *NOMO1-KO* y 6 clones *NOMO1-WT*, englobando a las líneas celulares HCT-116 y HS-5. En este caso, se incluyeron en el estudio 3 clones *wild type* y 3 clones *knockout* de cada línea celular. Al igual que en el análisis de microarrays, de esta manera se eliminaría el efecto del fondo genético de cada línea, por lo que los cambios detectados en el perfil de expresión génica se podrían atribuir a la pérdida de *NOMO1*.

El análisis de los resultados del RNA-seq mostró un total de 592 genes desregulados significativamente (*False Discovery Rate*; $FDR < 0,05$) en las células carentes de *NOMO1* de las dos líneas celulares. De todos ellos, 108 genes se encontraban significativamente sobre-expresados ($FC > 1,5$) y 213 genes se encontraban infra-expresados ($FC < -1,5$). Estos datos corroboran, nuevamente, que la inactivación de *NOMO1* genera un cambio en el perfil transcripcional de las líneas HCT-116 y HS-5 (Figura 41).

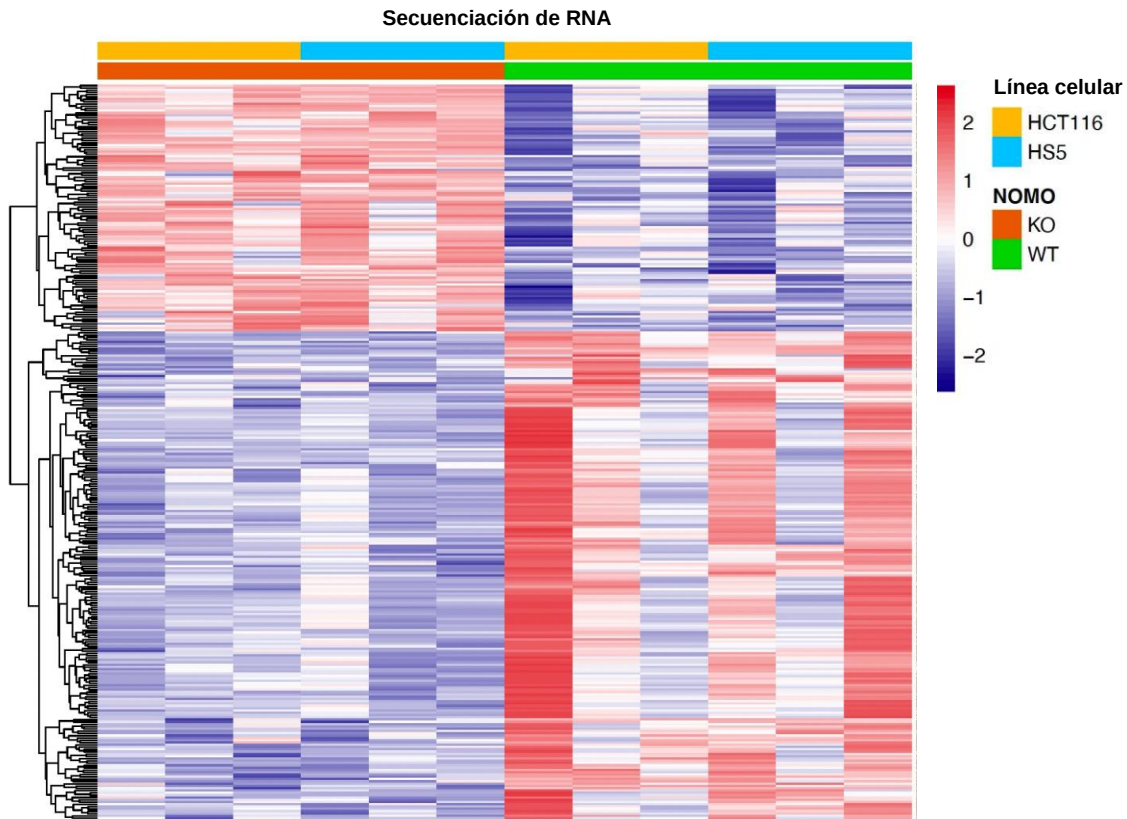


Figura 41. Perfil de expresión diferencial de las líneas celulares HCT-116 y HS-5 con inactivación de *NOMO1*. Representación mediante *heatmap* de los genes comunes desregulados en la secuenciación de RNA ($FC > 1,5$; $FC < -1,5$).

Además, el análisis de enriquecimiento de rutas (GSEA) realizado con el set de 321 genes desregulados significativamente en el RNA-seq mostró un total de 20 procesos afectados en las células carentes de *NOMO1* (Figura 42). Cabe destacar que entre los 20 procesos afectados en el RNA-seq se encontraban la mayoría de las rutas que se identificaron en el microarray de expresión. El análisis de enriquecimiento con las bases de datos “KEGG” y “Reactome” se detallan en el anexo 2.

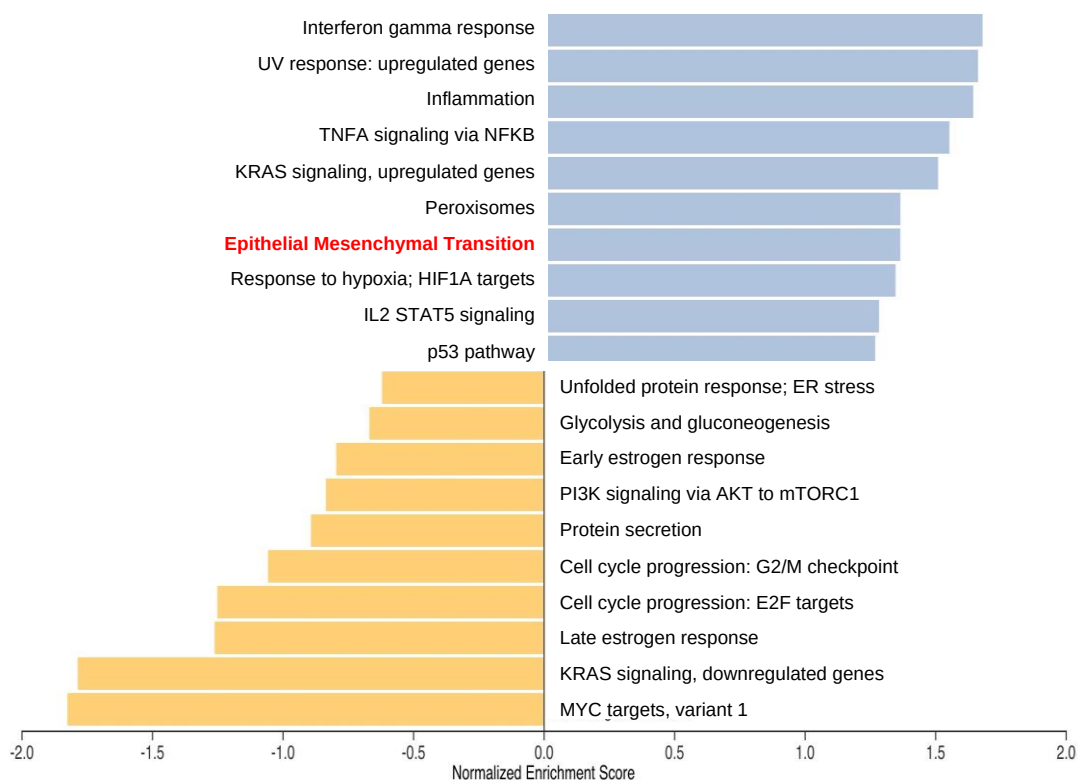


Figura 42. Procesos biológicos desregulados asociados a la pérdida de *NOMO1* en las líneas celulares HCT-116 y HS-5. Se representan los procesos alterados al introducir el set de genes desregulados en la secuenciación de RNA en el análisis GSEA-hallmarks ($FC > 1,5$; $FC < -1,5$).

Los datos obtenidos en el perfil transcripcional indican que las líneas celulares carentes de *NOMO1* presentan genes desregulados que se relacionan con la ruta de TNF α (Factor de Necrosis Tumoral α), así como con los procesos de inflamación y formación de vasos sanguíneos, entre otros. Estas vías están comúnmente alteradas durante la EII, la cual es un factor de riesgo importante para el desarrollo del CCR. Interesantemente, los genes desregulados en el microarray y en el RNA-seq generan la alteración de la transición epitelio-mesénquima, un proceso fundamental para la adquisición de la capacidad invasiva y metastásica de las células tumorales (Tabla 17).

Tabla 17. Genes asociados a la transición epitelio-mesénquima (EMT) que se encuentran desregulados en las líneas celulares HCT-116 y HS-5 tras la inactivación de *NOMO1*.

Gen	Descripción	FC	Estado	Ensayo
<i>LUM</i>	Lumican	2,51	Sobre-expresado	Microarray
<i>POSTN</i>	Periostin	2,14	Sobre-expresado	Microarray
<i>VCAN</i>	Versican	2,02	Sobre-expresado	Microarray
<i>CXCL8</i>	C-X-C motif chemokine ligand 8	1,81	Sobre-expresado	Microarray
<i>PTX3</i>	Pentraxin	1,78	Sobre-expresado	Microarray
<i>EDIL3</i>	EGF like repeats and discoidin domains 3	1,71	Sobre-expresado	Microarray
<i>LOXL1</i>	Lysyl oxidase like 1	1,64	Sobre-expresado	Microarray
<i>BDNF</i>	Brain derived neurotrophic factor	1,63	Sobre-expresado	Microarray
<i>CAPG</i>	Capping Actin Protein	1,62	Sobre-expresado	RNA-seq
<i>CCN1</i>	Cellular Communication Network Factor 1	-1,70	Infra-expresado	RNA-seq
<i>CCN2</i>	Cellular Communication Network Factor 2	-2,38	Infra-expresado	RNA-seq
<i>CD44</i>	CD44	-1,91	Infra-expresado	RNA-seq
<i>PCOLCE</i>	Procollagen Endopeptidase Enhancer	2,17	Sobre-expresado	RNA-seq
<i>QSOX1</i>	Quiescin Sulfhydryl Oxidase 1	1,51	Sobre-expresado	RNA-seq
<i>SERPINE2</i>	Serpin Family E Member 2	1,62	Sobre-expresado	RNA-seq
<i>THBS1</i>	Thrombospondin 1	-2,69	Infra-expresado	RNA-seq
<i>VEGFA</i>	Vascular Endothelial Growth Factor A	1,72	Sobre-expresado	RNA-seq

Por otro lado, se observó que los genes desregulados en las líneas celulares *NOMO1-KO* que se vinculaban con la EMT también se relacionan con la afectación de la migración celular (Figura 43).

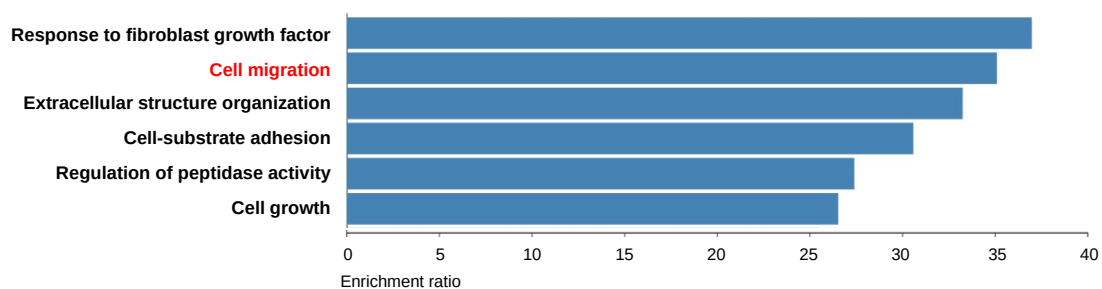
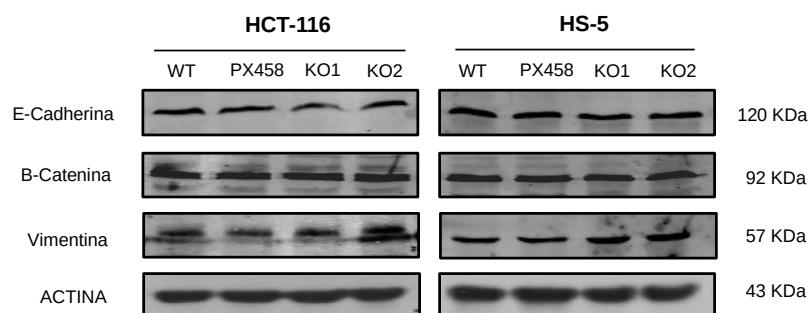


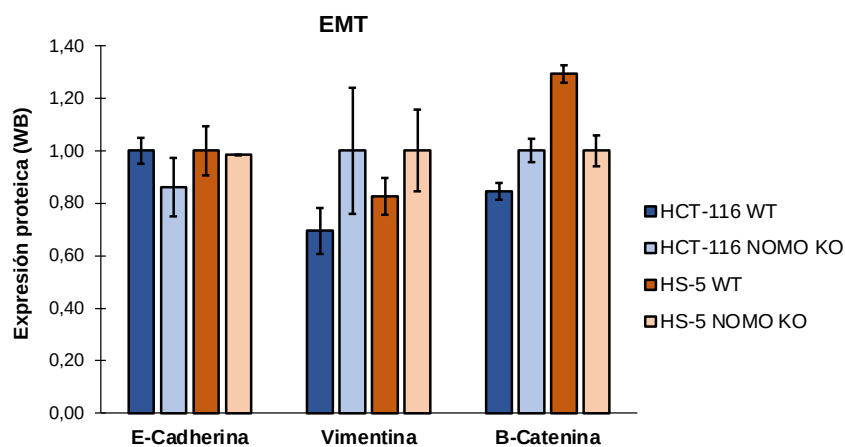
Figura 43. Procesos biológicos desregulados en las líneas celulares carentes de *NOMO1*, al incluir los genes diferencialmente expresados que mostraban asociación con la EMT en el análisis de sobre-representación de rutas.

Además, la pérdida de *NOMO1* es recurrente en EOCRC, un subgrupo de pacientes que se diagnostican fundamentalmente en estadios avanzados y que desarrollan metástasis tempranas. Así, la célula tumoral presentaría una mayor capacidad migratoria y un importante potencial para invadir tejidos. Para ello, el mecanismo de transición desde un estado epitelial a mesenquimal es fundamental. Por este motivo, nos propusimos estudiar mediante WB la expresión de E-Cadherina, Vimentina y β -Catenina, tres marcadores de EMT que también se relacionan con un aumento de la capacidad migratoria. Sin embargo, la expresión proteica de E-Cadherina, Vimentina y β -Catenina no mostró cambios significativos entre los clones *NOMO-KO* y *WT* de las dos líneas celulares (Figura 44).

A



B



	Cuantificación WB							
	HCT-116 WT	DS	HCT-116 NOMO KO	DS	HS-5 WT	DS	HS-5 NOMO KO	DS
E-Cadherina	1,00	0,05	0,86	0,11	1,00	0,09	0,99	0,00
Vimentina	0,69	0,09	1,00	0,24	0,83	0,07	1,00	0,16
β -Catenina	0,85	0,03	1,00	0,04	1,29	0,03	1,00	0,06

Figura 44. Expresión proteica de los marcadores de EMT en las líneas celulares HCT-116 y HS-5 con inactivación de *NOMO*. A) Expresión de E-Cadherina, Vimentina y β -Catenina analizada mediante western blot en los clones *NOMO-KO* y *NOMO-WT*. B) Cuantificación y normalización de los niveles de expresión proteica de E-Cadherina, Vimentina y β -Catenina. DS: desviación estándar.

En resumen, la inactivación de *NOMO1* genera un perfil transcripcional diferencial y desencadena la alteración de la EMT, entre otras vías afectadas, la cual no se puede explicar por la desregulación de E-cadherina, Vimentina y β -Catenina. Por tanto, sería necesario identificar otras proteínas desreguladas tras la pérdida de *NOMO1* que pudieran explicar la afectación de la EMT y que pudieran relacionarse con los procesos de migración, invasión y metástasis.

3.2. Estudio del proteoma

Para dilucidar los cambios en los perfiles de expresión proteica tras la inactivación de *NOMO*, se llevó a cabo un análisis mediante espectrometría de masas (LC-IMS/MS) de las líneas celulares HCT-116 y HS-5. Para ello, se compararon 4 réplicas de los clones *NOMO*-WT y *NOMO*-KO empleando 0,5 μ g de proteína de cada muestra. Los datos crudos obtenidos en el ensayo de LC-IMS/MS se han depositado en “*ProteomeXchange Consortium*” (<http://proteome-central.proteomexchange.org>), a través del repositorio PRIDE, con el identificador de acceso PXD033636.

La cuantificación del proteoma incluyó un total de 3227 proteínas, las cuales se clasificaron como sobre-expresadas o infra-expresadas cuando el ratio *KOMT* fue mayor o menor a 1, respectivamente. Así, la inactivación de *NOMO* generó un total de 357 proteínas desreguladas significativamente (p -valor $<0,05$) en la línea HCT-116, de las cuales 205 proteínas se encontraban sobre-expresadas y 152 infra-expresadas. Por su parte, en la línea HS-5 *NOMO*-KO se identificaron 486 proteínas desreguladas (p -valor $<0,05$), de las cuales 182 se presentaban de forma sobre-expresada y 304 de forma infra-expresada. Cuando se cruzaron las proteínas desreguladas en los clones *knockout* de ambas líneas, se encontraron 16 proteínas infra-expresadas y 12 sobre-expresadas en común (Figura 45, Tabla 18).

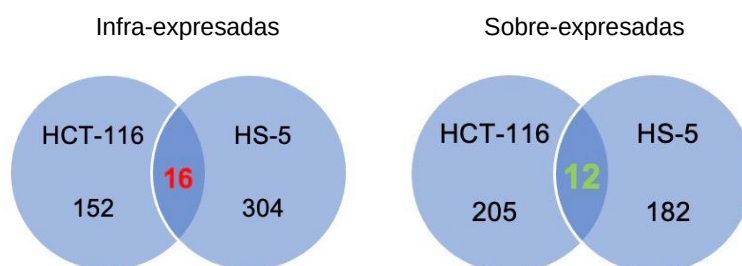


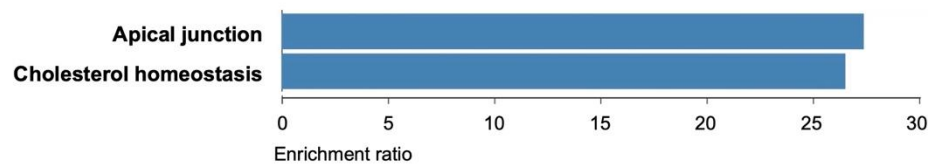
Figura 45. Representación mediante diagramas de Venn de las proteínas infra-expresadas y sobre-expresadas identificadas mediante LC-IMS/MS de forma común en las líneas celulares HCT-116 y HS-5 con inactivación de *NOMO*.

Tabla 18. Proteínas desreguladas tras la pérdida de NOMO en las líneas HCT-116 y HS-5, identificadas mediante LC-IMS/MS.

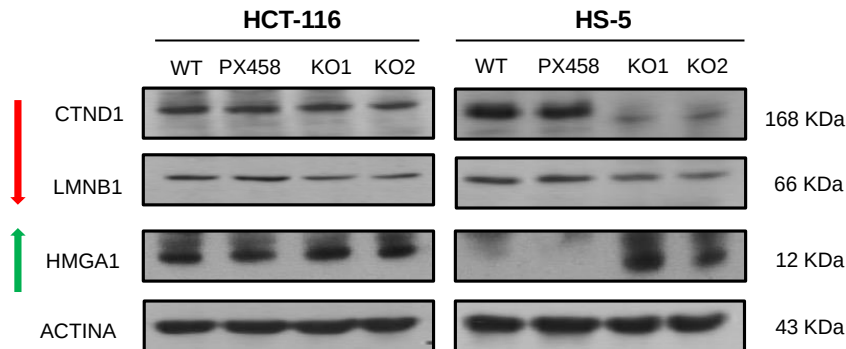
Proteína	Descripción	Expresión KO/WT		Estado
		HCT-116	HS-5	
K2013	Uncharacterized protein	2,71	4,59	Sobre-expresado
GFPT2	Glutamine--fructose-6-phosphate aminotransferase	2,11	3,68	
S10A6	Protein S100-A6	2,01	3,47	
FDFT	Squalene synthase	1,99	2,86	
HMGA1	High mobility group protein A1	1,98	2,65	
SPB4	Serpin B4	1,81	2,55	
TMM33	Transmembrane protein 33	1,79	1,87	
RRAS	Ras-related protein	1,47	1,79	
NSDHL	Sterol-4-alpha-carboxylate 3-dehydrogenase	1,43	1,61	
PRDX1	Peroxiredoxin-1	1,31	1,44	
CUL2	Cullin-2	1,26	1,40	
DHE3	Glutamate dehydrogenase 1	1,14	1,35	
ACTN4	Alpha-actinin-4	0,89	0,78	Infra-expresado
CTND1	Catenin delta-1	0,84	0,58	
HS90B	Heat shock protein	0,91	0,79	
HMGB1	High mobility group protein B1	0,87	0,67	
ANXA3	Annexin A3	0,66	0,20	
LMNB1	Lamin-B1	0,76	0,67	
RIR1	Ribonucleoside-diphosphate reductase	0,78	0,49	
ITA3	Integrin alpha-3	0,52	0,43	
UBP5	Ubiquitin carboxyl-terminal hydrolase 5	0,79	0,66	
NOMO	Nodal modulator	0,02	0,02	
TDIF2	Deoxynucleotidyltransferase terminal-interacting protein 2	0,56	0,49	
H1X	Histone H1.10	0,49	0,64	
NCLN	Nicalin	0,06	0,08	
UN45A	Protein unc-45 homolog A	0,61	0,31	
MYOF	Myoferlin	0,68	0,70	
SNX9	Sorting nexin-9	0,72	0,72	

Una vez identificada la presencia de un perfil proteómico diferencial en las líneas NOMO-KO, se incluyeron las 28 proteínas desreguladas en común a las dos líneas celulares en un análisis de enriquecimiento de rutas (GSEA). En este caso, se observó que las proteínas desreguladas tras la pérdida de NOMO facilitan la alteración de procesos biológicos como la homeostasis del colesterol o las uniones apicales. En consonancia con lo observado a nivel transcriptómico, los defectos en las uniones apicales de las células se identifican normalmente en células con una mayor capacidad migratoria, invasiva y metastásica. También, entre las 28 proteínas desreguladas en el proteoma se encontraban las proteínas CTND1, LMNB1 y HMGA1, cuyos defectos se asocian con un aumento de la capacidad migratoria y de la EMT. Por este motivo, decidimos confirmar la expresión de estas proteínas mediante western blot (Figura 46).

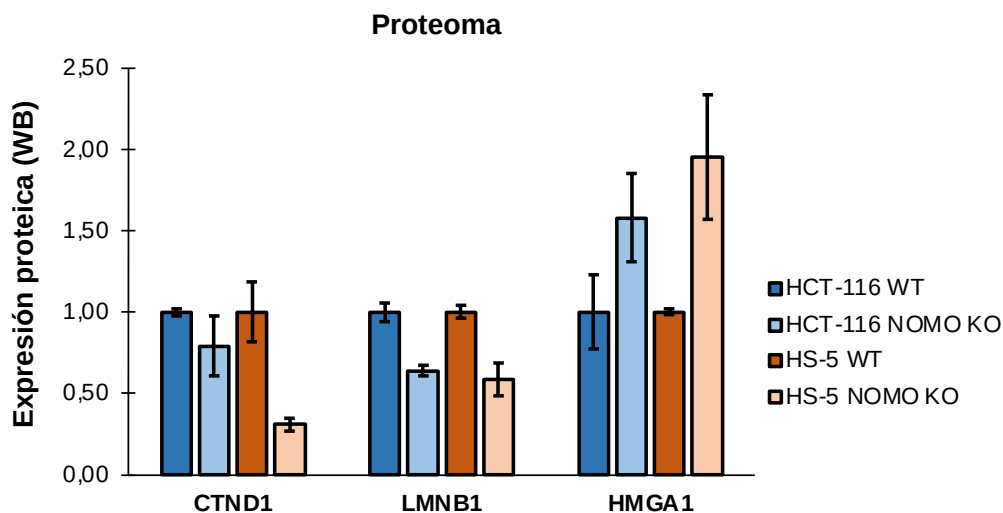
A



B



C



	Cuantificación WB							
	HCT-116 WT	SD	HCT-116 NOMO KO	SD	HS-5 WT	SD	HS-5 NOMO KO	SD
CTND1	1,00	0,02	0,79	0,18	1,00	0,18	0,31	0,04
LMNB1	1,00	0,06	0,64	0,03	1,00	0,04	0,59	0,10
HMGA1	1,00	0,23	1,58	0,27	1,00	0,02	1,95	0,38

Figura 46. Expresión de las proteínas asociadas a migración y EMT que se muestran desreguladas en el proteoma de las líneas celulares HCT-116 y HS-5 carentes de NOMO. A) Procesos biológicos desregulados en las líneas celulares NOMO-KO, incluyendo las proteínas diferencialmente expresadas detectadas mediante LC-IMS/MS. B) Expresión de las proteínas CTND1, LMNB1 y HMGA1 en los clones NOMO-WT y NOMO-KO mediante western blot. C) Cuantificación y normalización de los niveles de expresión proteica de CTND1, LMNB1 y HMGA1. DS: desviación estándar.

Por tanto, nuestros resultados confirman los cambios de expresión de CTND1, LMNB1 y HMGA1 observados en las líneas celulares carentes de NOMO. Esto sugiere que la pérdida de *NOMO1* podría inducir un aumento de la capacidad migratoria, invasiva y metastásica en EOCRC a través de la alteración de estas proteínas.

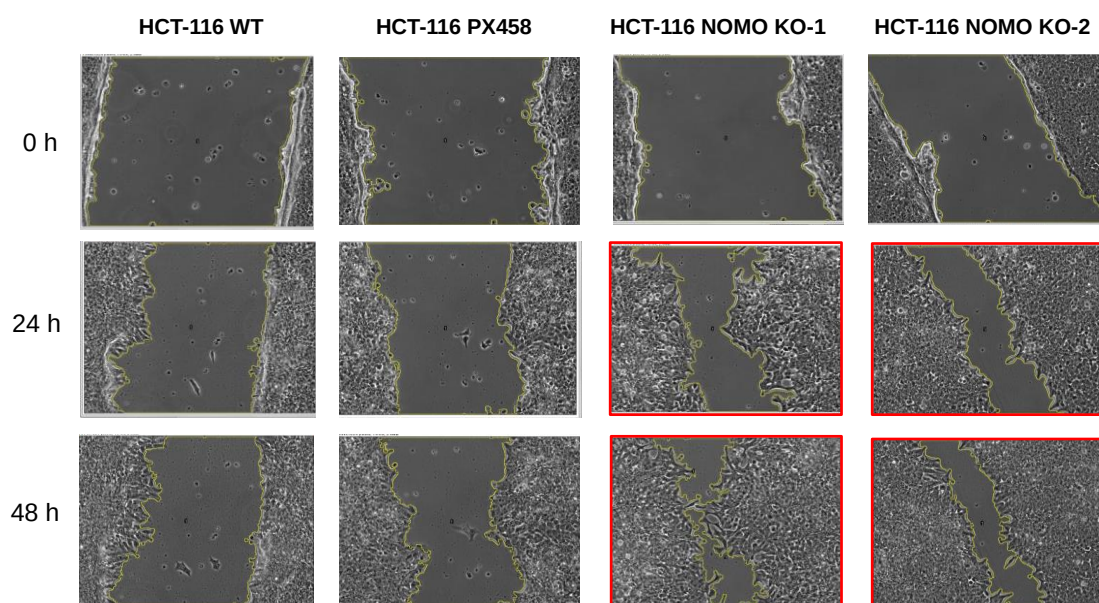
4. Análisis de la capacidad migratoria

Para determinar si la pérdida de *NOMO1* modificaba la migración celular, nos propusimos estudiar la capacidad migratoria de los clones NOMO-WT y NOMO-KO de las líneas celulares HCT-116 y HS-5. Para ello, realizamos un ensayo de *cierre de herida* y un ensayo de migración celular mediante *transwell*. Cada ensayo se replicó tres veces, y para cada una de estas réplicas se analizaron tres heridas de cada clon.

- **Estudio de migración celular mediante ensayo de *cierre de herida*:**

En el ensayo de “*cierre de herida*”, el área ocupada por las células que migraron se calculó a las 0, 12, 24, 36 y 48 horas para la línea HCT-116. Para la línea HS-5, el área ocupada se calculó hasta las 36 horas, ya que en ese momento todas las heridas estaban cerradas (Figura 47). En este caso, las células con inactivación de NOMO mostraron una mayor capacidad migratoria a todos los tiempos, especialmente a las 24 y 48 horas, dado que el porcentaje de área ocupada era significativamente mayor que la condición control (p-valor <0,001). Los porcentajes de área ocupada para cada réplica y cada tiempo de los clones *knockout* y *wild type* de las dos líneas celulares se detallan en el anexo 3.

A



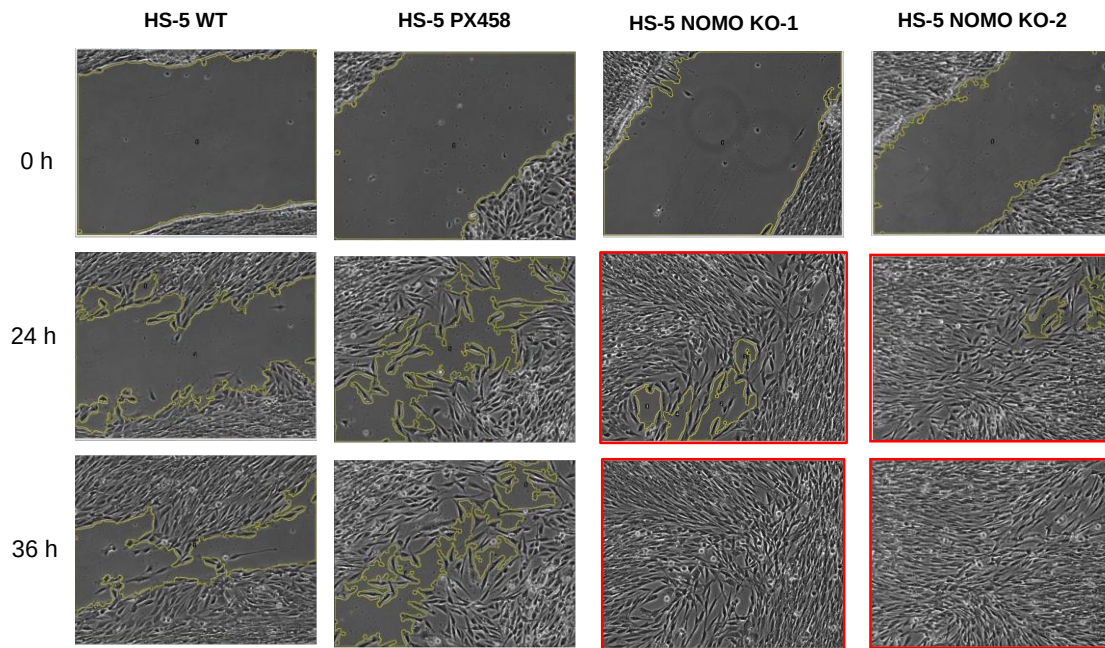
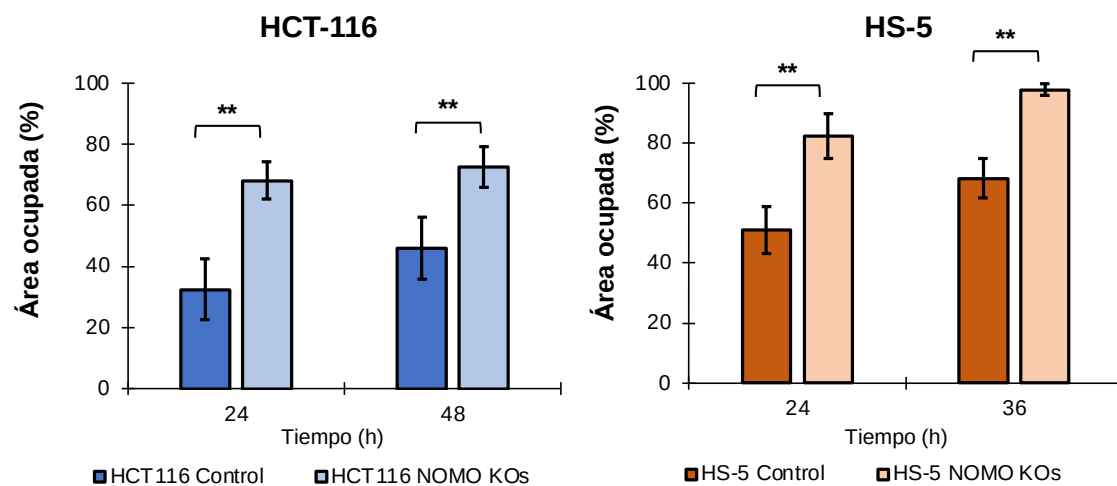
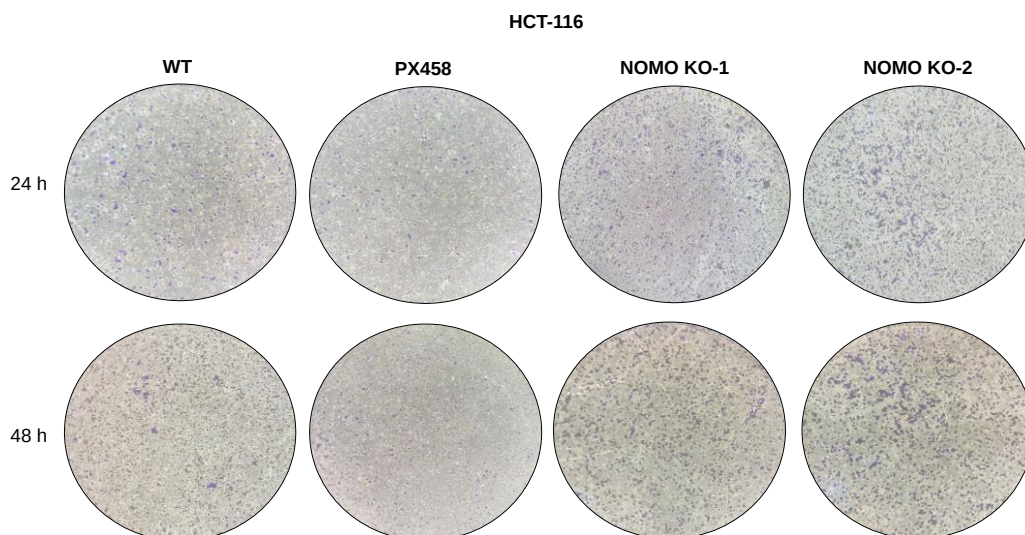
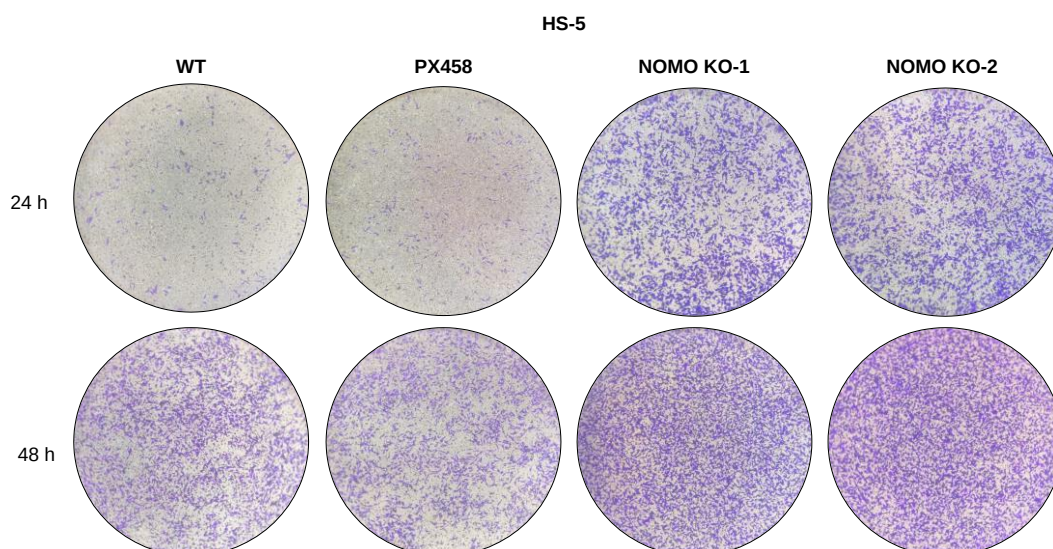
B**C**

Figura 47. Análisis mediante ensayo de *cierre de herida* de la capacidad migratoria de dos clones NOMO-KO y dos clones *wild type* de las líneas celulares HCT-116 y HS-5. A) Imágenes de microscopía tomadas a las 0, 24 y 48 horas de la línea HCT-116. B) Imágenes de microscopía tomadas a las 0, 24 y 36 horas de la línea HS-5. C) Normalización y cuantificación del área ocupada por las células WT y NOMO-KO con capacidad migratoria para cada línea celular.

**: $p < 0,001$.

- **Estudio de migración celular mediante ensayo de *transwell*:**

Posteriormente, quisimos validar este incremento de la capacidad migratoria mediante un ensayo de *transwell*. En este caso, las células que fueron capaces de migrar a través de una membrana porosa fueron teñidas con cristal violeta a las 24 y 48 horas, y finalmente cuantificadas con el programa ImageJ®. Como se puede observar en la figura 48, el número de células que consiguieron atravesar la membrana porosa fue significativamente mayor en los clones NOMO-KO que en los clones NOMO-WT de las líneas HCT-116 y HS-5 (p-valor <0,001). Por tanto, con este ensayo se corrobora que la inactivación de *NOMO1* produce un aumento de la migración celular. Las cuantificaciones del número de partículas positivas (células que han migrado) para cada réplica y cada tiempo de los clones *knockout* y *wild type* de las dos líneas celulares se detallan en el anexo 4.

A**B**

C

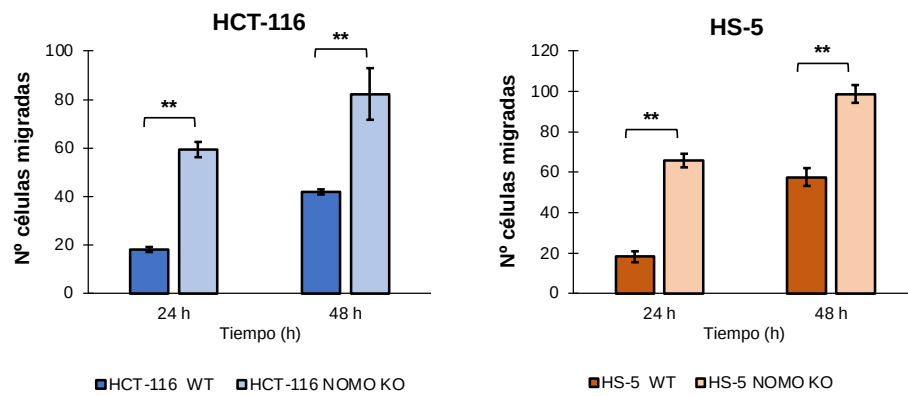


Figura 48. Análisis de la capacidad migratoria de las líneas HCT-116 y HS-5 mediante ensayo de *transwell*. Se han comparado dos clones NOMO-KO y dos clones NOMO-WT de cada línea celular. A) Imágenes al microscopio óptico de las células migradas de la línea HCT-116, tomadas con el objetivo 10X. B) Imágenes al microscopio óptico de las células migradas de la línea HS-5, tomadas con el objetivo 10X. C) Normalización y cuantificación de las células con capacidad migratoria (nº células migradas). **: $p < 0,001$.

De forma consistente en todas las líneas celulares, nuestros resultados muestran que la capacidad migratoria de las células NOMO-KO aumenta de forma significativa en las líneas celulares HCT-116 y HS-5. Por ello, concluimos que la inactivación de *NOMO1* promueve la migración celular en las líneas celulares incluidas en nuestro trabajo.

CAPÍTULO II. CARACTERIZACIÓN DE LA PÉRDIDA DE *NOMO1* EN UN MODELO MURINO CONDICIONAL

1. Generación de un modelo murino condicional para *Nomo1* mediante el sistema de edición génica CRISPR/cas9

Además del estudio en líneas celulares, quisimos reproducir la situación de pérdida de función de NOMO en otros modelos de experimentación, con el fin de definir su papel en el desarrollo del EOCRC. Para ver el efecto de su delección somática, generamos un modelo murino condicional para *Nomo1* con el objetivo de inducir su delección en un momento y contexto celular específico. La elección de un modelo murino condicional se debe a que la pérdida de *Nomo1* en estado embrionario es letal.

Mediante la tecnología de edición génica CRISPR/cas9 se generó una rotura de doble hebra en el locus de *Nomo1* en ratones de la cepa C57BL/6J, concretamente en una región del exón 3-intrón 4 del gen. En esa región, se insertaron dos secuencias *loxP* que flanquean el exón 3 y que son reconocidas por la recombinasa Cre. Esta recombinasa provoca la recombinación de las dos secuencias *loxP*, de manera que escinde el exón 3 de *Nomo1*, rompiéndose el marco de lectura y generando su inactivación.

La generación de este modelo murino se ha descrito en un trabajo previamente publicado por nuestro grupo²⁰². En resumen, tras generar la descendencia, se genotiparon todos los ratones con el fin de identificar aquellos alelos modificados correctamente. Tan solo en uno de los animales analizados se había producido la recombinación homóloga en toda la extensión del DNA molde insertado, y por lo tanto incorporó las modificaciones genómicas deseadas. Se trataba de un ratón mosaico con 3 alelos diferentes: uno de ellos con una delección de 4 pb dentro de la secuencia del exón 3, otro con una delección mayor de 47 pb, y un último alelo con las modificaciones esperadas (alelo *Flox*).

Una vez caracterizados los alelos, se estudió si el alelo de interés se transmitía a la línea germinal para poder segregarlo a la descendencia. Así, se puso en cruce dicho ratón fundador con dos hembras con alelos *Nomo1 wild type*. Su descendencia, un total de 9 ratones, se genotiparon mediante PCR y se vio que 4 de las crías poseían el alelo *flox* y un alelo silvestre. Es decir, se generaron eficientemente 4 ratones condicionales para *Nomo1* en heterocigosis (*Nomo1^{flox/wt}*). Finalmente, se pusieron en cruce estos animales heterocigotos para generar ratones con el gen *Nomo1* mutado en homocigosis (*Nomo1^{flox/flox}*), en heterocigosis (*Nomo1^{flox/wt}*) y ratones *wild type*.

1.1. Validación del sistema cre-loxp *in vitro*

Para comprobar el funcionamiento del sistema cre-loxp *in vitro*, se infectó un cultivo de MEFs, obtenidos de embriones que procedían del cruce entre ratones con *Nomo1* mutado en heterocigosis. Estos MEFs fueron extraídos entre los días 12,5 y 13,5 de gestación. Posteriormente, se mantuvieron en cultivo hasta ser infectados con un adenovirus que porta las secuencias codificantes de la recombinasa Cre y del marcador de selección GFP (AD5CMVCre-GFP) (Figura 49).

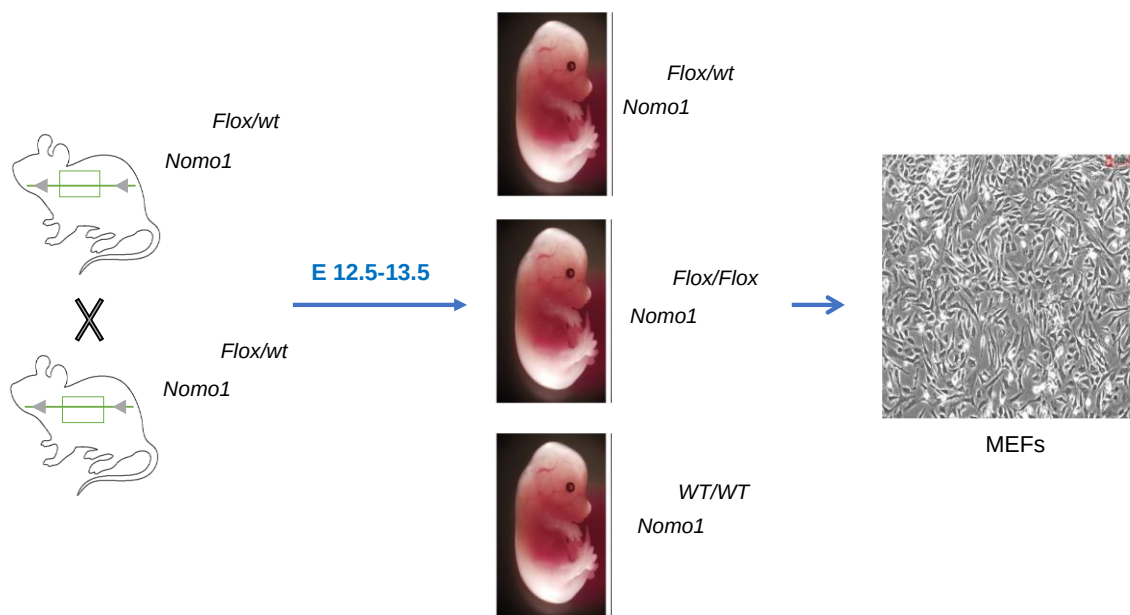


Figura 49. Extracción de fibroblastos embrionarios de ratón (MEFs) para la validación *in vitro* del sistema cre-loxp.

En este caso, se infectaron 100.000 MEFs procedentes de ratones con *Nomo1* mutado en homocigosis, heterocigosis y *wild type*. Tras 48 horas de la infección, se testó su eficiencia en un microscopio de fluorescencia, y se extrajo DNA y RNA para validar la edición génica mediante PCR, secuenciación Sanger y RT-qPCR.

En primer lugar, los MEFs infectados con el adenovirus mostraron expresión de GFP, indicando que el adenovirus se había introducido en las células y, por tanto, la recombinasa CRE se estaría expresando. En segundo lugar, la validación en DNA genómico mediante PCR mostró un fragmento de 482 pb como resultado de la escisión del exón 3 de *Nomo1*. Los ratones *Nomo1^{flox/flox}* mostraron un fragmento de 482 pb, los ratones *Nomo1^{flox/wt}* un fragmento de 482 pb y otro de 786 pb, y los ratones *wild type* un único fragmento de 786 pb. En tercer lugar, se analizó mediante RT-qPCR los niveles de expresión de *Nomo1* en los MEFs infectados y sin infectar. En este caso, se observó una disminución significativa de la

expresión de *Nomo1* en los embriones condicionales en homocigosis y heterocigosis que se habían editado correctamente (Figura 50). Cabe destacar que la expresión no fue nula en ningún caso ya que no todas las células consiguieron introducir el adenovirus, por lo que no se produciría la escisión del exón 3 de *Nomo1* en el 100% del cultivo.

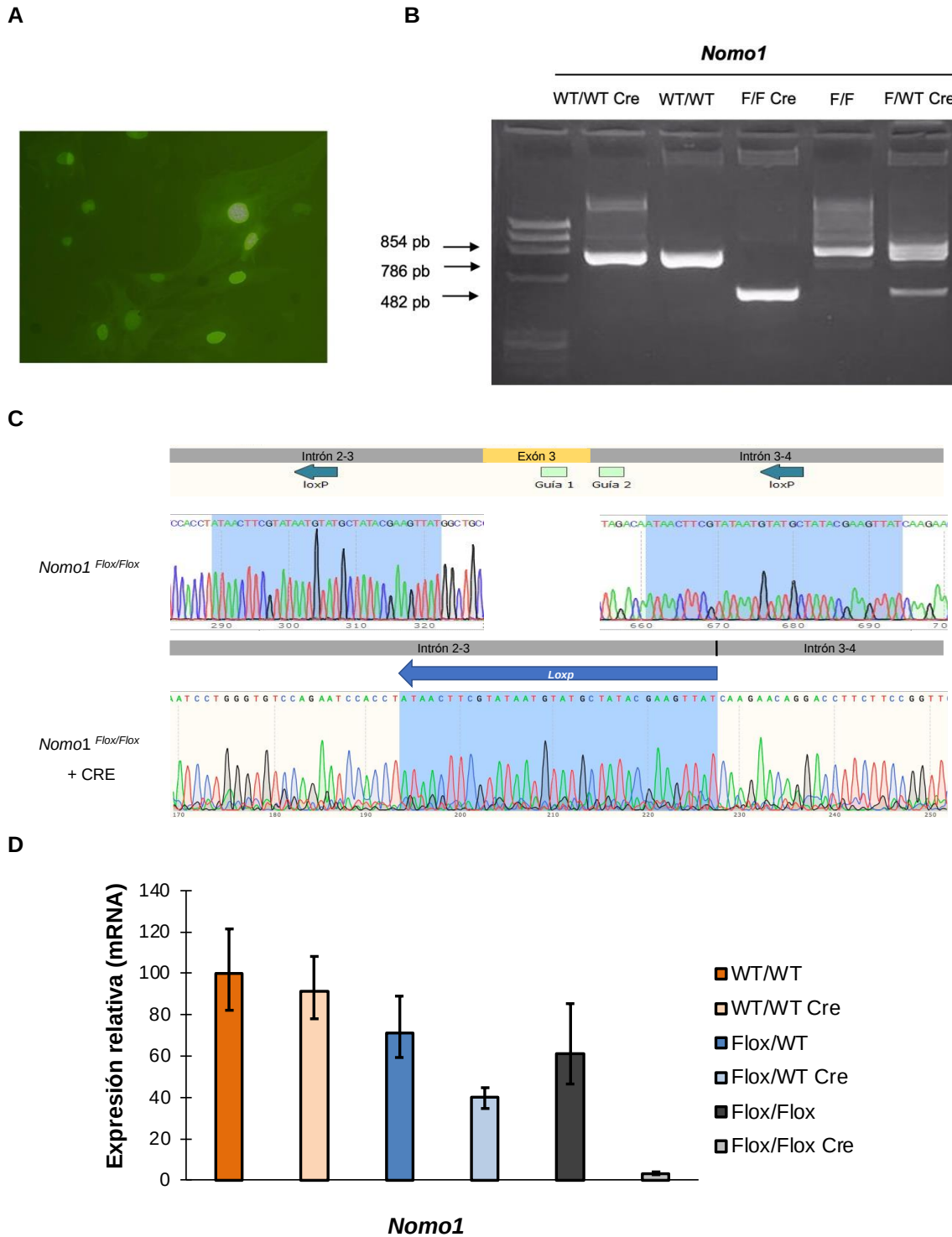


Figura 50. Validación del sistema cre-loxp *in vitro*. Escisión del exón 3 de *Nomo1* en fibroblastos embrionarios de ratón (MEFs) mediante infección con AD5CMVCre-GFP. A) Imagen de microscopía de fluorescencia donde se detecta la expresión de GFP, como control de eficiencia

de la infección, en MEFs. B) Validación mediante PCR del sistema cre-loxp en DNA genómico de MEFs infectados y sin infectar. C) Secuencia de nucleótidos del exón 3 de *Nomo1* en MEFs que presentan la delección en homocigosis. D) Niveles normalizados de mRNA del exón 3 de *Nomo1* detectados mediante RT-qPCR.

1.2. Validación del sistema cre-loxp *in vivo*

Para ver el efecto de la delección somática de *NOMO1* en EOCRC, y considerando este gen como un posible supresor tumoral, quisimos silenciar la expresión de *Nomo1* en las células del colon de los ratones condicionales generados previamente.

Para ello, se inyectó el adenovirus que porta la recombinasa CRE en el colon de 9 ratones de 8 semanas de edad: 6 *Nomo1*^{flox/flox}, 2 *Nomo1*^{flox/wt} y 1 *Nomo1*^{wt/wt}. Como grupo control, se utilizaron 2 ratones *Nomo1*^{flox/flox} y un ratón *Nomo1*^{flox/wt} sin tratar con el adenovirus. Tras 20 meses de seguimiento, todos los ratones se sacrificaron por dislocación cervical y se extrajo DNA del intestino para comprobar el funcionamiento del sistema cre-loxp *in vivo*.

Tras evaluar mediante PCR el estado del exón 3 de *Nomo1*, se pudo observar que el sistema cre-loxp había funcionado eficientemente, ya que fue capaz de generar la escisión del exón 3 en el intestino. Sin embargo, se pudo comprobar que las células que incorporaron el adenovirus habían sido editadas, mientras que las células que no consiguieron infectarse mantuvieron el estado de *Nomo1* *wild type*. Por tanto, los ratones *Nomo1*^{flox/flox} y *Nomo1*^{flox/wt} infectados amplificaron un fragmento de 854 pb (alelo flox) o 786 pb (alelo wt) y otro de 482 pb (alelo *delta*). Por su parte, todos los ratones del grupo control mostraron un único fragmento de 854 pb (alelo flox) o 786 pb (alelo wt), según correspondiese (Figura 51).

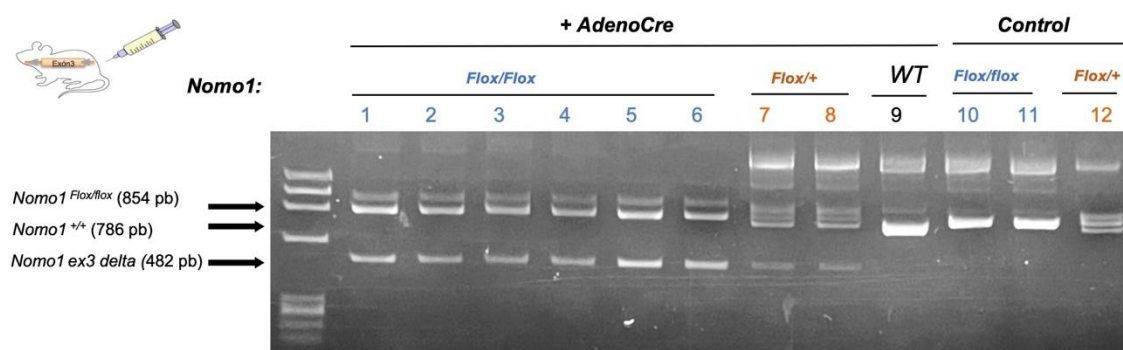


Figura 51. Validación del sistema cre-loxp *in vivo*. Estudio mediante PCR del estado del exón 3 de *Nomo1* en el intestino de ratones condicionales infectados con el adenovirus-Cre y ratones control.

2. Análisis anatomopatológico de los modelos murinos con inactivación de *Nomo1*

2.1. Inactivación de *Nomo1* mediante inyección con adenovirus portador de la recombinasa Cre

Tras la activación de la Cre, los animales se monitorizaron durante 20 meses y se sacrificaron por dislocación cervical, momento en el que se llevó a cabo una inspección macroscópica. Posteriormente, se realizó un análisis histológico del tracto gastrointestinal y de otros tejidos afectados de cada animal en el Servicio de Anatomía Patológica del Hospital Universitario de Salamanca. El estudio anatomopatológico confirmó que ni los ratones con inactivación de *Nomo1* ni los ratones del grupo control mostraban lesiones en el intestino compatibles con el desarrollo de CCR (Tabla 19).

Tabla 19. Análisis anatomopatológico de los ratones condicionales para *Nomo1* infectados con el adenovirus portador de la recombinasa Cre.

ID	Genotipo: <i>Nomo1</i>	Sexo	AdenoCRE	Anatomía Patológica
1	FLOX/FLOX	Hembra	Si	Normal
2	FLOX/FLOX	Macho	Si	Normal
3	FLOX/FLOX	Hembra	Si	Normal
4	FLOX/FLOX	Hembra	Si	Normal
5	FLOX/FLOX	Hembra	Si	Displasia intestino delgado
6	FLOX/FLOX	Hembra	Si	Normal
7	FLOX/WT	Hembra	Si	Normal
8	FLOX/WT	Macho	Si	Normal
9	WT/WT	Hembra	Si	Normal
10	FLOX/FLOX	Macho	NO	Normal
11	FLOX/FLOX	Hembra	NO	Intestino normal. Sarcoma de mama.
12	FLOX/WT	Hembra	NO	Normal

De manera excepcional, se observó que un animal con delección de *Nomo1* en homocigosis presentaban una lesión displásica en el intestino delgado (Figura 52). Además, un ratón del grupo control presentó un sarcoma de mama.

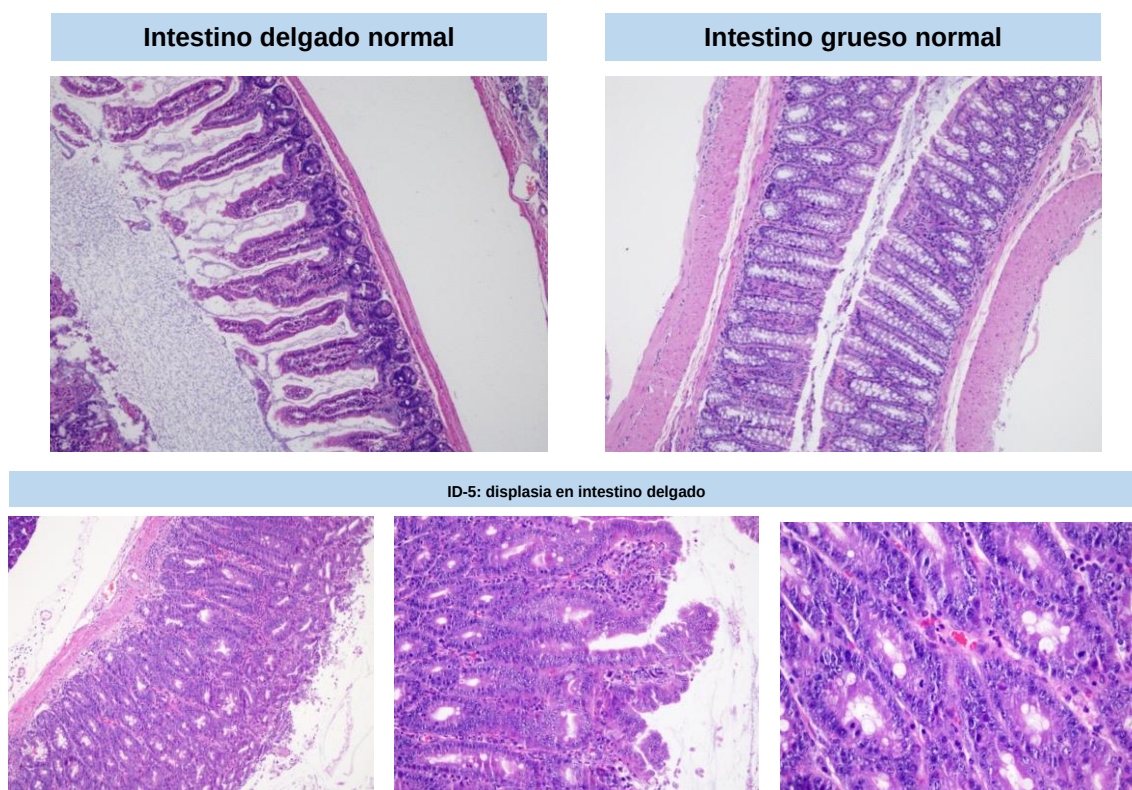


Figura 52. Histología del intestino delgado y grueso normal. Se representa el análisis histológico de un animal con inactivación de *Nomo1* y displasia en el intestino delgado.

Estos resultados indican que la pérdida de *Nomo1* podría ser una mutación *passenger* en el desarrollo del cáncer colorrectal.

2.2. Inactivación de *Nomo1* en ratones *Msi1-Cre*

Además de inducir la pérdida de *Nomo1* de manera local en el intestino, nos propusimos estudiar si la delección en heterocigosis de *Nomo1* desde fases embrionarias promovía la aparición de tumores en el colon. Cabe destacar que este modelo solo pudo llevarse a cabo con la inactivación de un alelo del gen ya que la pérdida en homocigosis desde estadios embrionarios es letal. Para ello, se pusieron en cruce ratones *Nomo1*^{flox/wt} con ratones que expresan la recombinasa Cre de manera constitutiva. La expresión de la Cre en estos animales está controlada por el promotor *Msi1*, un marcador de células *stem* identificado justo encima de las células de Paneth del intestino delgado. La expresión de dicha recombinasa ha sido identificada en varios tejidos como cerebro, pulmón, piel y criptas intestinales. Así, su expresión en dichos tejidos promueve la escisión del exón 3 de *Nomo1* por recombinación de las dos secuencias *loxP*.

Se generaron un total de 13 ratones *Nomo1*^{flox/wt}; *Msi1-Cre*⁺ (delección heterocigota de *Nomo1*) y se utilizaron como control 12 ratones *Nomo1*^{flox/wt} (condicionales sin delección).

Se realizó un seguimiento de todos los animales durante al menos un año para ver si desarrollaban alguna patología. Al cabo de 14 meses, se sacrificaron mediante dislocación cervical 3 ratones con la pérdida en heterocigosis y 1 ratón del grupo control, dado que visualmente ningún animal presentaba signos de enfermedad (Tabla 20).

Tabla 20. Análisis anatomopatológico de los ratones *Nomo1*^{flox/wt} con expresión constitutiva de la recombinasa Cre.

ID	Genotipo: <i>Nomo1</i>	Sexo	Msi1-Cre	Anatomía Patológica
13	FLOX/WT	Macho	Si	Normal
14	FLOX/WT	Macho	Si	Angiosarcoma rectal
15	FLOX/WT	Hembra	Si	Normal
16	FLOX/WT	Macho	No	Normal

El estudio histológico del tracto gastrointestinal de todos los ratones no mostró ninguna evidencia de neoplasia intestinal. De forma anecdótica, un animal *Nomo1*^{flox/wt}; Msi1-Cre⁺ presentó un angiosarcoma en el recto (Figura 53). Estos datos corroboran que la pérdida de *Nomo1* en el colon podría no ser una mutación *driver* del cáncer colorrectal.

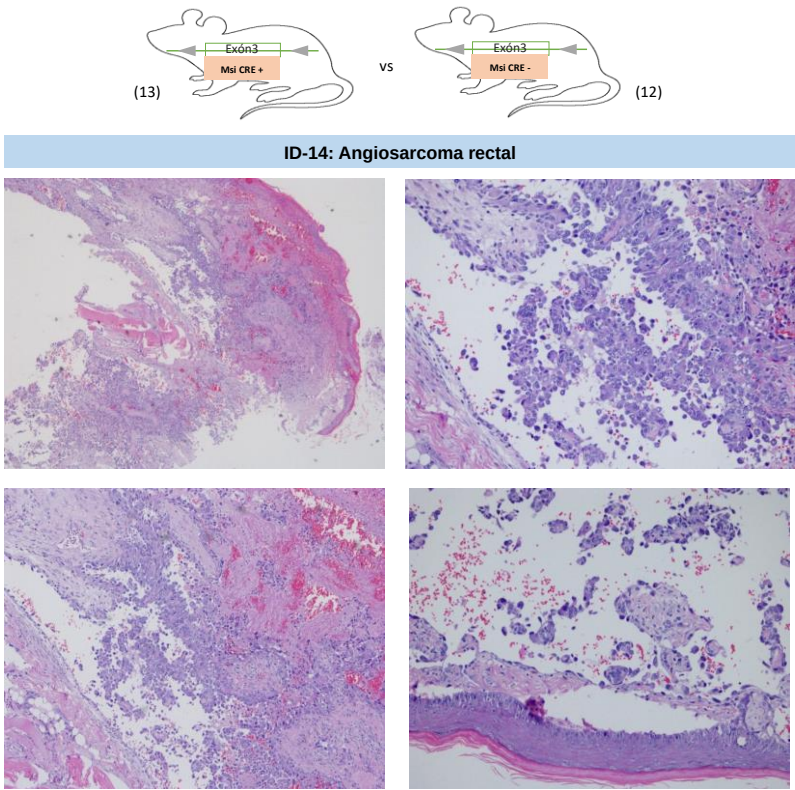


Figura 53. Angiosarcoma rectal de un animal *Nomo1*^{flox/wt}; Msi1-Cre⁺.

2.3. Modelo de carcinogénesis colorrectal mediante administración de 1,2-dimetilhidrazina (DMH)

Una vez confirmado que la inactivación de *Nomo1* no es una mutación *driver* en EOCRC, nos planteamos la posibilidad de que tuviese un papel importante como mutación *passenger* en esta enfermedad. En este sentido, cabría esperar que la pérdida de *Nomo1* acelerase la génesis del EOCRC. Por este motivo, decidimos generar un modelo de carcinogénesis colorrectal mediante la administración de un agente metilante del DNA, el DMH. El tratamiento con este fármaco promueve la aparición desde lesiones benignas hasta adenocarcinomas en el colon.

Para ello, se administró el DMH a un grupo de 4 ratones con delección heterocigota del gen (*Nomo1*^{flox/wt}; *Msi1*-Cre+), a un grupo de 6 ratones condicionales como grupo control 1 (*Nomo1*^{flox/wt}) y a un grupo de 5 ratones *wild type* como grupo control 2. Se administraron 20 mg/kg una vez por semana, durante 20 semanas. A la semana 22, se sacrificaron todos los animales por dislocación cervical y se realizó un estudio anatomopatológico del tracto gastrointestinal u otros tejidos con signos de patología. Se realizó el análisis histológico de todos los ratones con delección heterocigota de *Nomo1*, de 3 ratones del grupo control 1 y de 2 ratones del grupo control 2. En este caso, se puede confirmar que el modelo de carcinogénesis había funcionado correctamente, ya que la mayoría de los animales habían desarrollado adenomas tras 20 semanas de tratamiento (Tabla 21).

Tabla 21. Análisis anatomopatológico de los ratones *Nomo1*^{flox/wt} con expresión constitutiva de la recombinasa Cre y de los ratones control en un modelo de carcinogénesis colorrectal inducida por DMH.

ID	Genotipo: <i>Nomo1</i>	Sexo	<i>Msi1</i> -Cre	Anatomía Patológica
17	FLOX/WT + DMH	Macho	Si	Normal
18	FLOX/WT + DMH	Hembra	Si	Normal
19	FLOX/WT + DMH	Macho	Si	Intestino normal Adenocarcinoma de pulmón
20	FLOX/WT + DMH	Hembra	Si	Adenocarcinoma en colon
21	FLOX/WT + DMH	Hembra	No	Adenoma en colon
22	FLOX/WT + DMH	Macho	No	Adenoma en colon
23	FLOX/WT + DMH	Hembra	No	Normal
24	WT/WT + DMH	Macho	No	Normal
25	WT/WT + DMH	Hembra	No	Adenoma en intestino

Además, en dos animales con pérdida de *Nomo1* en heterocigosis se observaron lesiones malignas. En el primero, se identificó un adenocarcinoma de pulmón con patrón papilar. En el segundo, se detectó un adenocarcinoma *in situ* intramucoso en el colon (Figura 54).

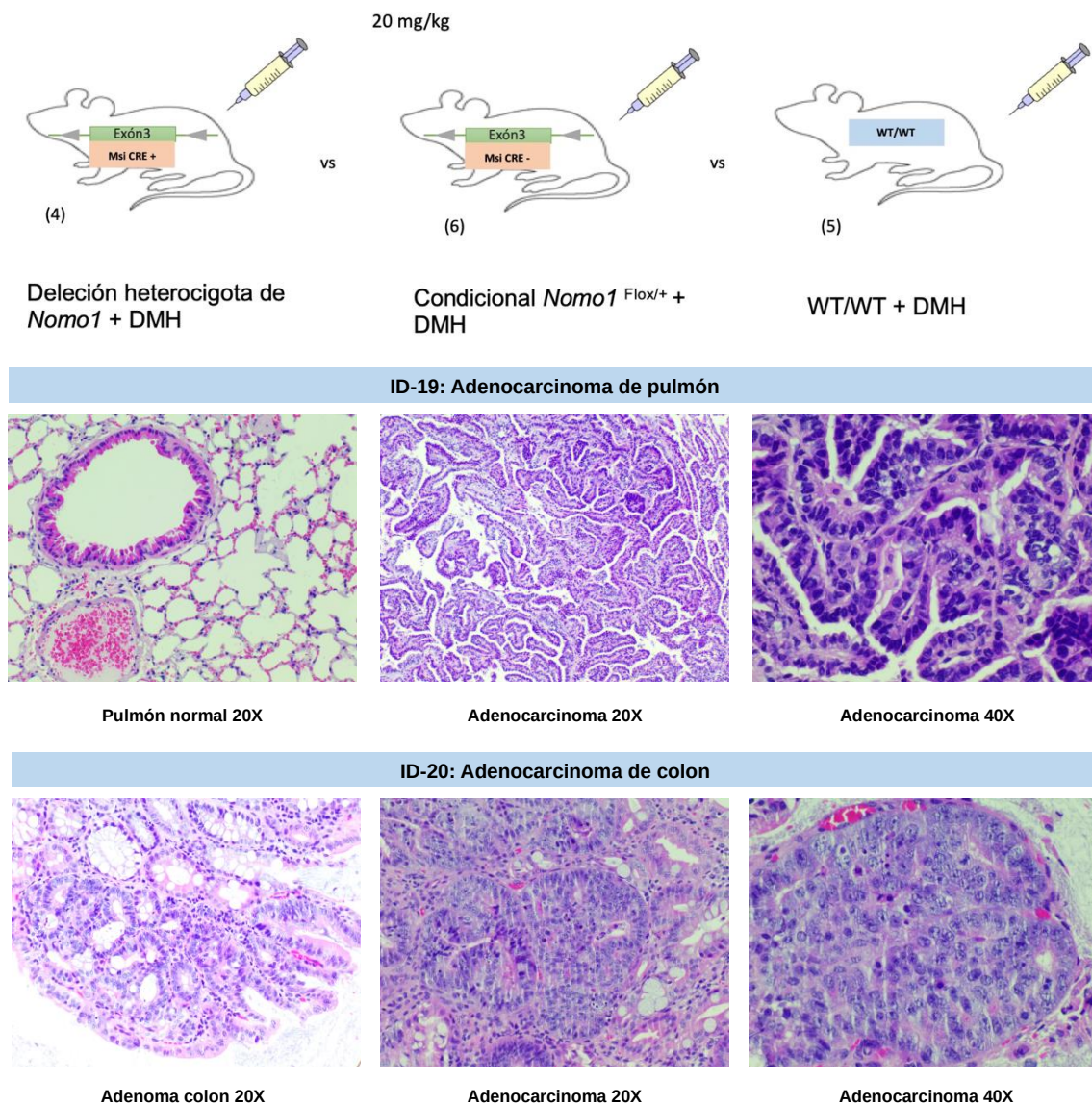


Figura 54. Histología del tejido pulmonar e intestinal de animales con inactivación de *Nomo1* en heterocigosis tratados con DMH. Se representa un adenocarcinoma de pulmón y un adenocarcinoma de colon.

CAPÍTULO III. ESTUDIO DE LA RESPUESTA AL TRATAMIENTO CON AGENTES QUIMIOTERÁPICOS

1. Estudio de la respuesta al tratamiento con 5-fluorouracilo, irinotecán, oxaliplatino y cisplatino de las líneas celulares carentes de *NOMO1*

Finalmente, independientemente de los mecanismos moleculares alterados tras la inactivación de *NOMO1* y su condición de mutación *passenger* en CCR, quisimos estudiar si la pérdida de este gen modificaba la respuesta al tratamiento. De esta manera, podríamos identificar si los pacientes menores de 50 años con cáncer colorrectal y delección de *NOMO1* se pueden beneficiar de un tratamiento particular. Para ello, analizamos la viabilidad celular de las líneas celulares HCT-116 y HS-5, con y sin inactivación de *NOMO1*, a través de los ensayos de MTT y de Anexina V/Yoduro de propidio.

1.1. Análisis de viabilidad celular mediante MTT

Para estudiar la respuesta al tratamiento en las dos líneas celulares mediante MTT, se trataron las células NOMO-KO y NOMO-WT con dosis crecientes de 5-FU, irinotecán, oxaliplatino y cisplatino. Para cada línea celular, se compararon 3 clones NOMO-KO y 3 clones NOMO-WT. La medición del número de células viables se realizó tras 24, 48 y 72 horas de incubación con el fármaco.

Como se puede observar en la figura 55, existe una respuesta dosis-dependiente para cada uno de los fármacos empleados en este trabajo, independientemente del tiempo de incubación. En este caso, no se encontraron diferencias significativas (p -valor $>0,05$) en la respuesta al tratamiento con 5-FU, irinotecán y oxaliplatino en función del estado de *NOMO* en ambas líneas celulares. Sin embargo, si se observó una mayor resistencia al cisplatino de las células NOMO-KO, si bien es cierto que este fármaco no está indicado para el tratamiento del CCR. Así, el IC_{50} para cada fármaco fue el siguiente:

- **5-fluorouracilo:**

IC_{50} HCT-116 NOMO-WT y KO: 4,25 y 5,85.

IC_{50} HS-5 NOMO-WT y KO: 20,38 y 19,16.

- **Irinotecán:**

IC_{50} HCT-116 NOMO-WT y KO: 6,66 y 7,71.

IC_{50} HS-5 NOMO-WT y KO: 5,53 y 3,91.

- **Oxaliplatino:**

IC_{50} HCT-116 NOMO-WT y KO: 1,58 y 1,91.

IC₅₀ HS-5 NOMO-WT y KO: 3,79 y 4,88.

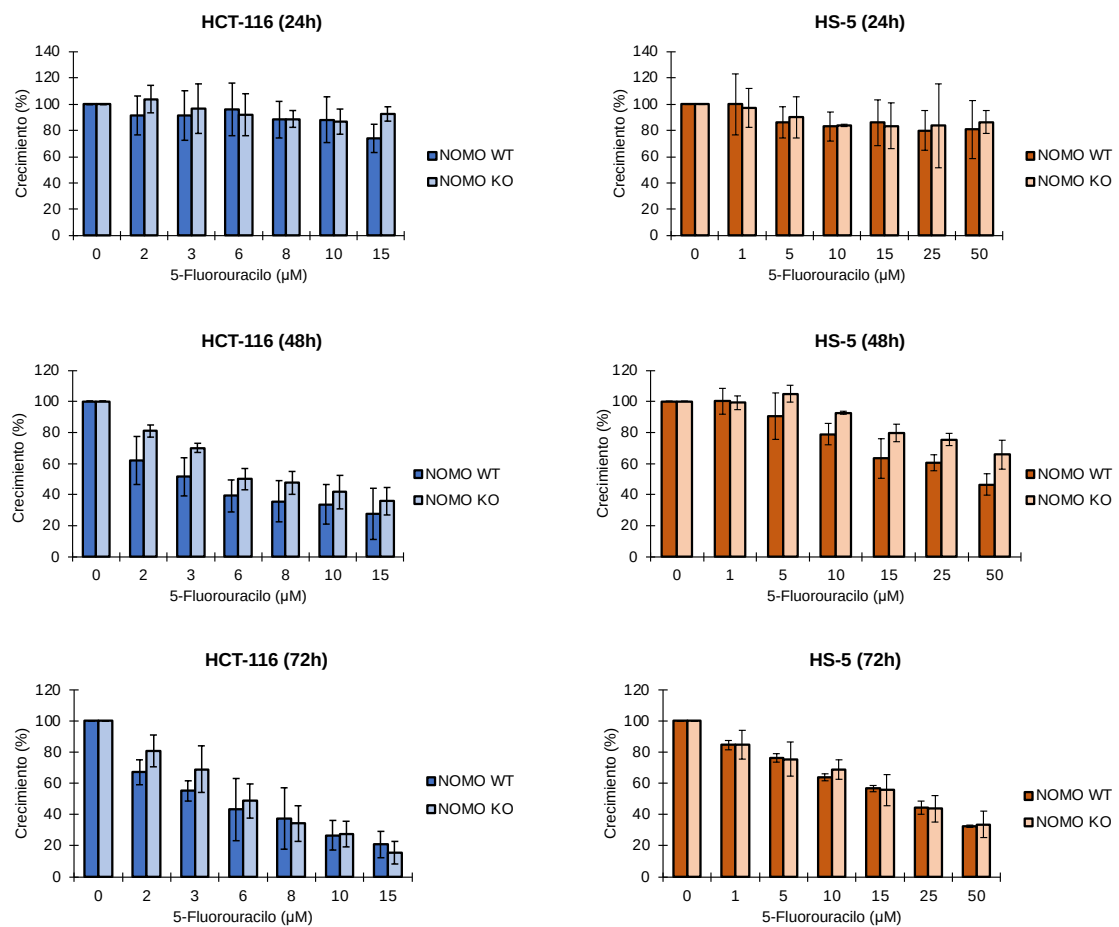
- **Cisplatino:**

IC₅₀ HCT-116 NOMO-WT y KO: 4,53 y 12,68.

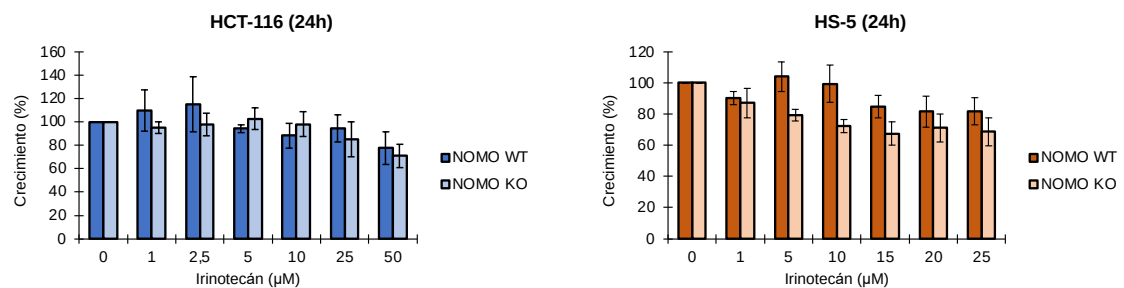
IC₅₀ HS-5 NOMO-WT y KO: 0,25 y 0,77.

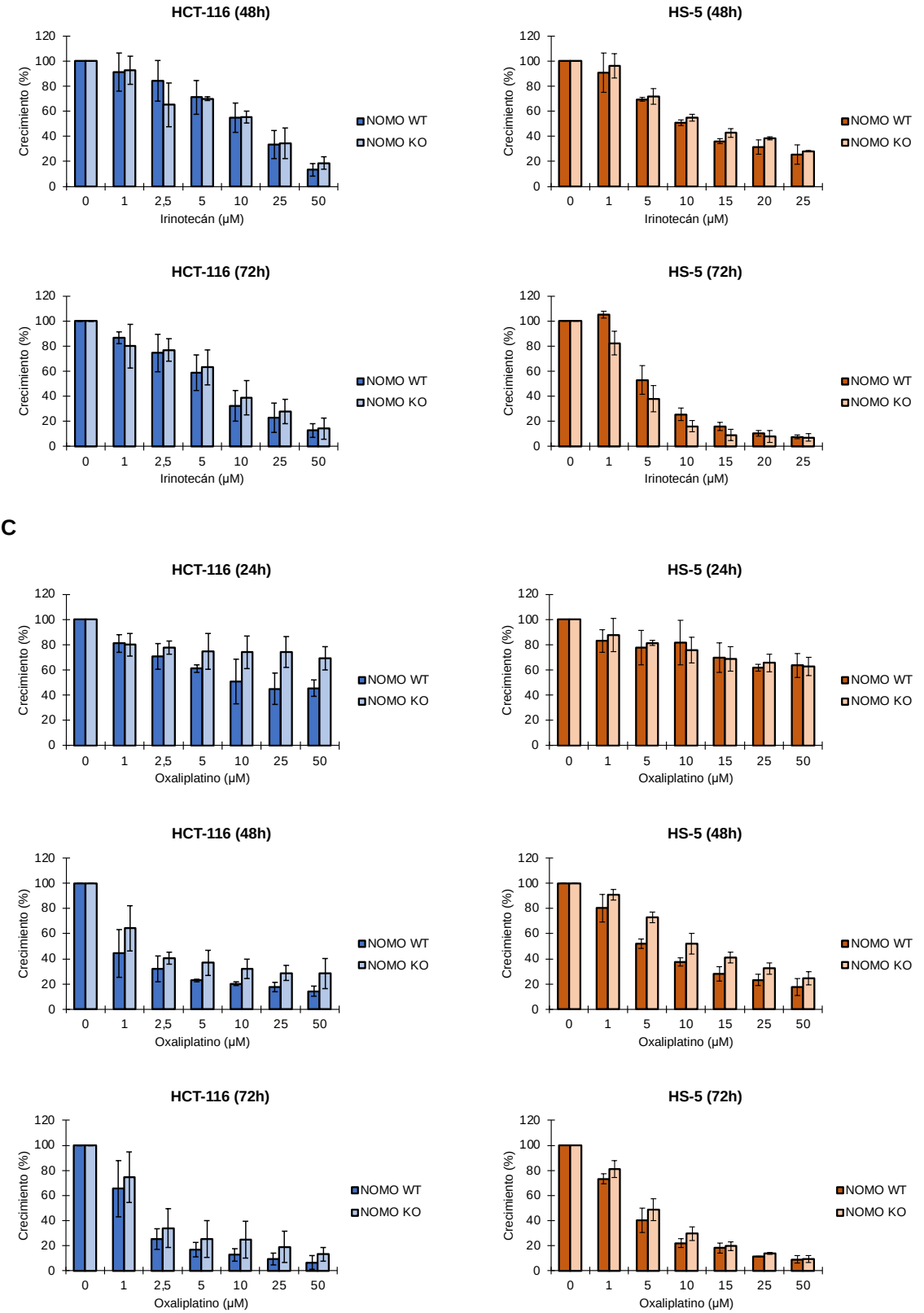
Con todo esto, la pérdida de *NOMO1* no modifica la respuesta al tratamiento con 5-fluorouracilo, irinotecán y oxaliplatino en EOCRC, pero si ofrece resistencia al cisplatino.

A



B





D

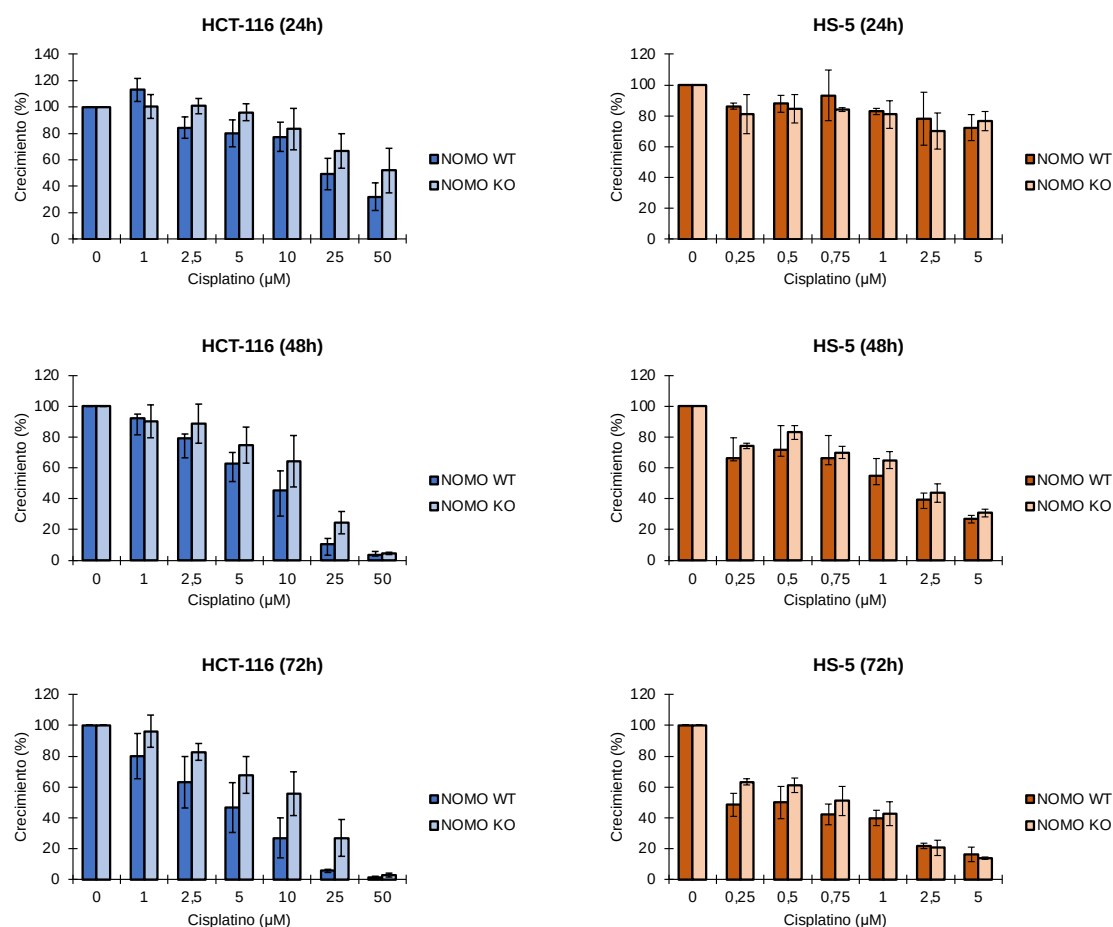


Figura 55. Análisis de viabilidad celular mediante MTT de las líneas celulares HCT-116 y HS-5, con y sin inactivación de NOMO, tras el tratamiento con agentes quimioterápicos. Se representan los porcentajes de crecimiento promedios de 3 clones NOMO-KO y 3 clones NOMO-WT de cada línea celular. A) Respuesta al tratamiento con 5-FU. B) Respuesta al tratamiento con irinotecán. C) Respuesta al tratamiento con oxaliplatino. D) Respuesta al tratamiento con cisplatino.

1.2. Análisis de viabilidad celular mediante el ensayo Anexina V/Yoduro de Propidio

Por último, se trataron de forma similar las células NOMO-KO y NOMO-WT de las dos líneas celulares con dosis crecientes de los mismos agentes quimioterápicos. En este caso, las concentraciones se seleccionaron en función de la respuesta detectada mediante MTT. Al igual que en el caso anterior, se compararon 3 clones NOMO-KO y 3 clones NOMO-WT de las líneas HCT-116 y HS-5 a las 72 horas tras el marcaje con Anexina V/Yoduro de Propidio. De esta manera, no solo podríamos contabilizar por citometría de flujo las células vivas, sino también las células que se encuentran en apoptosis, necrosis y necro-apoptosis.

Como se muestra en la figura 56, la viabilidad celular de las líneas HCT-116 y HS-5, con y sin inactivación de NOMO, no mostró diferencias significativas (p -valor $>0,05$) en la respuesta al tratamiento con 5-FU, irinotecán y oxaliplatino. Al igual que en el caso anterior, la línea HCT-116 NOMO-KO ofreció una mayor resistencia al cisplatino. Por otro lado, se observó un efecto dosis-dependiente para todos los fármacos. Así, el IC_{50} para cada uno de ellos fue el siguiente:

- **Irinotecán:**

IC_{50} HCT-116 NOMO-WT y KO: 4,08 y 5,59.

IC_{50} HS-5 NOMO-WT y KO: 13,57 y 14,60.

- **Oxaliplatino:**

IC_{50} HCT-116 NOMO-WT y KO: 4,14 y 3,97.

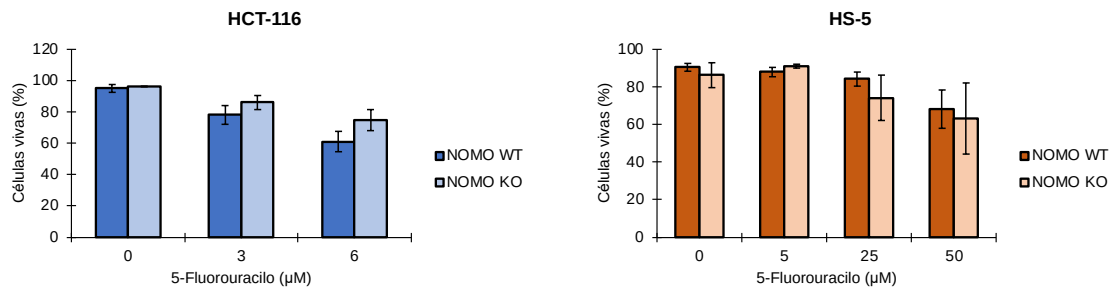
IC_{50} HS-5 NOMO-WT y KO: 6,05 y 11,11.

- **Cisplatino:**

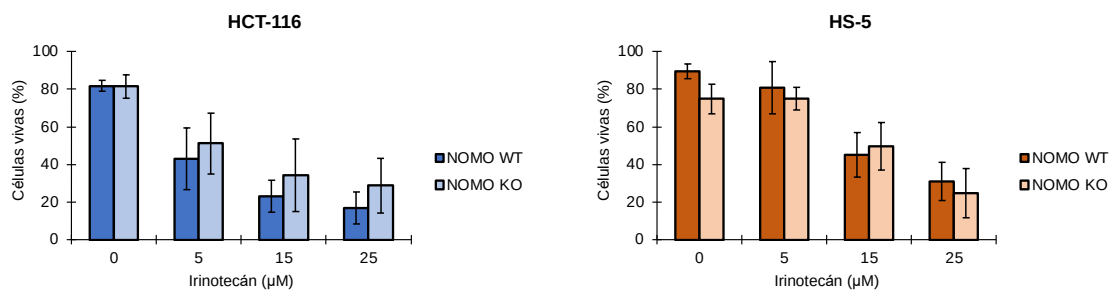
IC_{50} HCT-116 NOMO-WT y KO: 6,00 y 13,33.

IC_{50} HS-5 NOMO-WT y KO: 3,73 y 3,40.

A



B



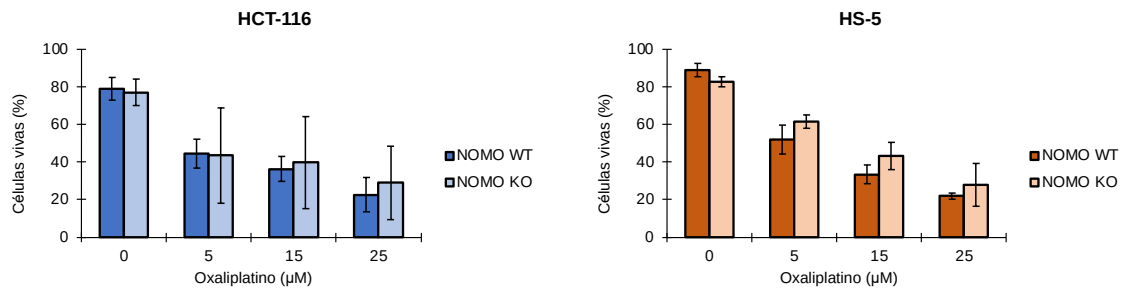
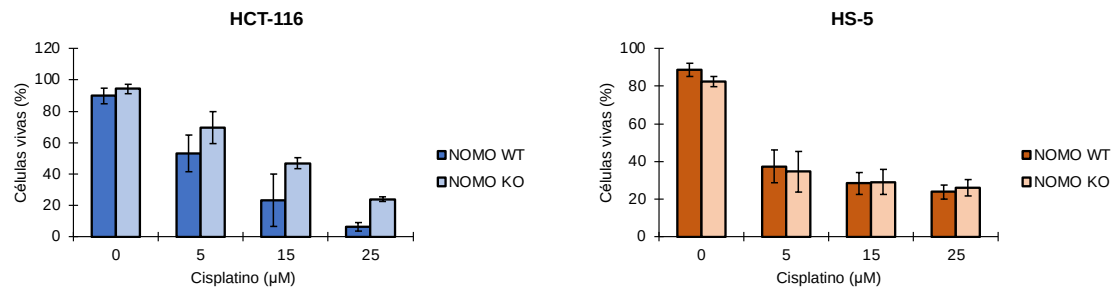
C**D**

Figura 56. Análisis de viabilidad celular mediante citometría de flujo de las líneas celulares HCT-116 y HS-5, con y sin inactivación de *NOMO*, tras el tratamiento con agentes quimioterápicos. Se representan los porcentajes promedios de 3 clones *NOMO-KO* y 3 clones *NOMO-WT* de cada línea celular. Marcaje con Anexina V/Yoduro de propidio a las 72 horas de tratamiento. A) Respuesta al tratamiento con 5-FU. B) Respuesta al tratamiento con irinotecán. C) Respuesta al tratamiento con oxaliplatino. D) Respuesta al tratamiento con cisplatino.

En conclusión, la pérdida de *NOMO1* no modifica la respuesta al tratamiento con 5-FU, irinotecán y oxaliplatino en las líneas celulares HCT-116 y HS-5. Por ello, no podemos realizar una indicación específica para este subgrupo de pacientes. Sin embargo, existe una resistencia al tratamiento con cisplatino en las células con inactivación del gen *NOMO1*.

CAPÍTULO IV. ESTUDIO DEL ACORTAMIENTO TELOMÉRICO EN PACIENTES CON CÁNCER COLORRECTAL DE APARICIÓN TEMPRANA

Dada la relación existente entre el acortamiento telomérico y el riesgo de desarrollar cáncer, nos propusimos analizar la LT de pacientes con EOCRC esporádico, con el fin de analizar la LT como un posible marcador predictivo de esta enfermedad, ya que esto permitiría modificar las estrategias de prevención dirigidas a la población de riesgo menor de 50 años.

En primer lugar, quisimos analizar si la variabilidad en la LT podría estar asociada con el desarrollo del EOCRC. Para ello, se cuantificó mediante RT-qPCR la LT absoluta de los leucocitos procedentes de SP de 87 pacientes con EOCRC esporádico y de 109 individuos sanos (grupo control), todos ellos menores de 50 años. Como se muestra en la figura 57, los telómeros de los pacientes con EOCRC eran significativamente más cortos ($p < 0,001$) que en el grupo control (LT media EOCRC: 122 kb; LT media control: 296 kb). Por lo tanto, el acortamiento telomérico podría predisponer al desarrollo del EOCRC.

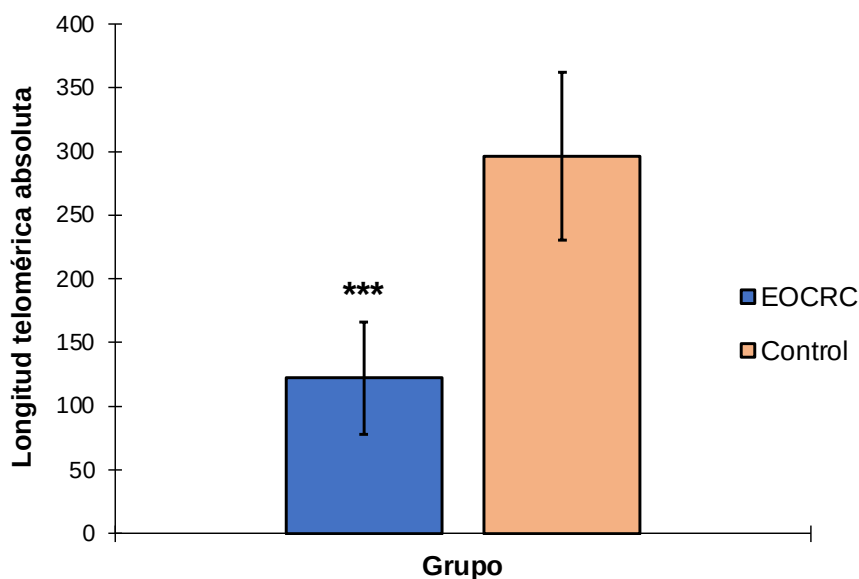


Figura 57. Cuantificación mediante RT-qPCR de la longitud telomérica absoluta en DNA procedente de sangre periférica de pacientes con EOCRC e individuos sanos. Se representan las longitudes teloméricas absolutas por célula diploide.

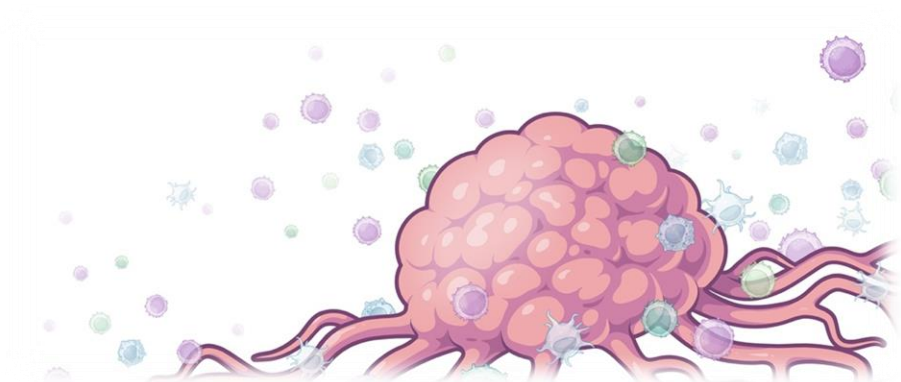
Dado que se han descrito polimorfismos en el gen *hTERT* que modulan sus niveles de transcripción en células mononucleadas de sangre periférica, nos propusimos estudiar la distribución genotípica del polimorfismo rs2853669 (A>G) en EOCRC, mediante discriminación alélica por PCR con sondas TaqMan®. Este SNP se localiza en la región promotora del gen *hTERT*, concretamente en la posición -245, y ya ha sido relacionado con el acortamiento telomérico secundario a mal funcionamiento de la telomerasa²¹⁵.

Como se muestra en la tabla 22, en nuestra cohorte, el estado del alelo minoritario (G) se asoció con un mayor riesgo de desarrollar EOCRC ($p=0,028$; OR: 1,326 (IC 95%: 1,018-1,728). En este caso, observamos que el alelo G era menos frecuente en EOCRC (26%) frente al grupo control (36%). Estos datos sugieren que el alelo minoritario podría desempeñar un papel protector frente al desarrollo del EOCRC. Por lo tanto, la pérdida de este factor protector podría aumentar el riesgo de desarrollar EOCRC a través de un mal funcionamiento de la telomerasa.

Tabla 22. Distribución genotípica del polimorfismo rs2853669 del gen *hTERT* en EOCRC y en la población control.

SNP	Genotipo	Controles	EOCRC esporádico	P-valor	OR (95% CI)
<i>hTERT</i> (rs2853669)	AA	43 (39%)	47 (54%)	0,082	
	AG	53 (49%)	35 (40%)		
	GG	13 (12%)	5 (6%)		
	AA+AG	96 (88%)	82 (94%)	0,137	
	GG	13 (12%)	5 (6%)		
	GG+GA	66 (61%)	40 (46%)	0,042	1,803 (1,020-3,190)
	AA	43 (39%)	47 (54%)		
	A	139 (64%)	129 (74%)	0,028	1,326 (1,018-1,728)
	G	79 (36%)	45 (26%)		

Discusión



Discusión

El cáncer colorrectal es el tercer tumor más frecuente y la segunda causa de muerte por cáncer en el mundo. En el momento actual, su incidencia se está incrementando debido al envejecimiento poblacional, aunque su mortalidad está disminuyendo. Esto se debe a la detección precoz de la enfermedad gracias a la implementación y mejora de los programas de cribado que se destinan a las personas mayores de 50 años, así como a la aparición de estrategias terapéuticas más eficaces.

Sin embargo, la incidencia del CCR en pacientes menores de 50 años (EOCRC) ha aumentado de forma importante en las últimas décadas, representando en el momento actual el 10% de todos los casos de CCR. En este sentido, el aumento de la incidencia del EOCRC de etiología desconocida es una epidemia mundial, sobre todo teniendo en cuenta que este subgrupo de pacientes se caracteriza por un diagnóstico de la enfermedad en estadio avanzado, con peor pronóstico y con fenotipos más agresivos del tumor²¹⁶. Además, el EOCRC de tipo esporádico abarca casi el 80% de todos los casos de EOCRC y presenta limitaciones para la identificación de biomarcadores específicos que faciliten su diagnóstico precoz. Los criterios de edad definidos en los programas de cribado del CCR hacen que este subgrupo de pacientes no pueda diagnosticarse en estadios tempranos de la enfermedad. Esto justifica que en los últimos años la mortalidad de este tipo de tumor se haya incrementado considerablemente⁶³.

A pesar del incremento en la incidencia del EOCRC, son pocos los estudios que han explorado las bases genéticas y los mecanismos moleculares implicados su génesis^{141,217,218}. Por tanto, los estudios sobre la etiología del EOCRC son imprescindibles para mejorar la prevención, el diagnóstico y el tratamiento del CCR de inicio temprano.

En un estudio previo realizado por nuestro grupo, se identificó una delección recurrente de la región cromosómica 16p13.12-p13.11 en pacientes menores de 50 años con CCR esporádico. Además, se demostró que más del 80% de los tumores EOCRCs examinados tenían una delección somática en homocigosis del gen *NOMO1*, localizado en dicha región cromosómica, y presentaban estabilidad de microsatélites¹⁵⁶. Además, en un análisis sobre el perfil mutacional del gen *NOMO1* en EOCRC, se encontró que el 50% de los tumores con delección heterocigota presentaban mutación patogénica en el otro alelo. Esto pone de manifiesto que *NOMO1* podría ser inactivado, no sólo por delección, sino también por mutación en los tumores EOCRC. Estos datos mostraban, por primera vez, la existencia de mutaciones patogénicas en el gen *NOMO1* asociadas al EOCRC, y sugerían que este gen podría actuar como un supresor tumoral.

1. CRISPR/Cas9 genera eficientemente la inactivación de *NOMO1* en líneas celulares

En este trabajo se presenta un estudio sobre el posible papel que desempeña el gen *NOMO1* en la carcinogénesis colorrectal en individuos jóvenes. Para ello, se reprodujo la situación de pérdida de función de *NOMO1* en líneas celulares y modelos animales utilizando la tecnología de edición génica CRISPR/cas9.

En primer lugar, se inactivó el gen *NOMO1* en las líneas celulares HCT-116 (EOCRC) y HS-5 (no tumoral) mediante el sistema de edición génica CRISPR/cas9. La elección de una línea celular de CCR y una línea no tumoral no solo nos ha permitido testar la reproducibilidad de nuestros resultados, sino también vincular los cambios fenotípicos observados a la pérdida de *NOMO1*, ya que el fondo genético de cada línea presenta importantes diferencias.

CRISPR/cas9 es una herramienta de edición genómica que puede programarse para cumplir distintas funciones según la modificación genética que se pretenda reproducir en la célula. Esta técnica está basada en la creación de roturas de doble hebra en el DNA que deben ser reparados por la célula. Por un lado, las DSB se reparan preferentemente a través de la unión de extremos no homólogos, un sistema de reparación propenso a errores que da lugar a inserciones o deleciones en la región diana (InDels). Estas modificaciones se caracterizan por interrumpir el marco de lectura y facilitar la formación de codones de parada prematuros, por lo que este sistema se considera el más efectivo para generar *knockouts* de genes²¹⁹. Por este motivo, hemos usado esta estrategia para generar la delección del exón 3 del gen *NOMO1*. Alternativamente, CRISPR/cas9 puede emplearse para insertar secuencias de interés en una región diana del gen. Para ello, es necesario añadir una secuencia de DNA donadora, la cual se introduce en la región de corte mediante el sistema de reparación por recombinación homóloga. Se trata de un sistema de edición génica donde se busca activar o reprimir la expresión génica, así como modificar la función de la proteína²¹⁹.

Para generar las líneas *knockout* del gen *NOMO1* se seleccionó el exón 3 como secuencia diana, dado que el último nucleótido de este exón y los dos primeros del exón 4 forman un mismo codón y codifican un mismo aminoácido. Por tanto, al producirse el corte de doble hebra del DNA mediado por la cas9 y repararse mediante NHEJ, se produce un cambio en el marco de lectura que facilita la formación de un codón de parada prematuro, dando lugar a una proteína truncada.

Para generar la inactivación de *NOMO1* se diseñaron tres sgRNAs distintos, localizados en el exón 3 y sus regiones intrónicas adyacentes. Cada uno de los sgRNAs fue clonado individualmente en un vector que portaba la secuencia codificante de la cas9. Así, al introducir el plásmido en la célula, cada sgRNA marcaría el punto de corte específico en la secuencia del gen, que sería reconocido por la cas9 para generar el corte de doble hebra en el DNA. Para producir la delección, se transfectaron las líneas celulares con dos combinaciones de guías: sgRNA1+sgRNA2, sgRNA1+sgRNA3 y sgRNA2+sgRNA3. De esta manera, se pudo aumentar las probabilidades de éxito ya que al menos una de las guías sería exónica, y un corte en esa región reparado defectuosamente por la célula podría afectar a la funcionalidad de la proteína. Además, al transfectar con dos guías prácticamente consecutivas se facilitó la pérdida del fragmento comprendido entre ellas sin que el sistema de reparación de la célula reparase el daño. Por último, dirigir distintos sgRNAs hacia una misma región de *NOMO1* aumenta la especificidad del corte y disminuye la frecuencia con la que aparecen las alteraciones fuera de la secuencia diana (modificaciones *off target*). En este sentido, la modificación continua del genoma aumenta la probabilidad de que se produzca la escisión fuera de la región diana, lo que reduce la selectividad de la edición provocando mutaciones no deseadas que pueden ser letales.

A pesar de que se trata de una técnica útil, sencilla y económica, la eficiencia del sistema CRISPR/cas9 es normalmente baja. En primer lugar, la eficiencia de inserción del plásmido en la línea celular HCT-116 fue muy baja (1-2% de células GFP+), mientras que en la línea HS-5 aproximadamente el 20% de las células fueron transfectadas correctamente. Esto puede deberse al método de transfección, ya que en la línea HCT-116 se utilizó la electroporación y en la línea HS-5 se empleó la lipofección, un método que genera menos letalidad. Sin embargo, la eficacia para generar clones *knockout* a partir de las células individualizadas mediante *single cell* fue similar en las dos líneas celulares. Tanto para la línea HCT-116 como HS-5 se consiguieron generar 3 clones *knockout* para *NOMO1*. Esto representa un 1% del total de células GFP+ separadas. Por lo tanto, en las líneas celulares utilizadas en este trabajo, CRISPR/cas9 ha mostrado ser un sistema de edición génica eficiente para generar la inactivación de *NOMO1*.

2. La desestabilización del complejo NOMO/NCLN no afecta a los niveles de expresión de TMEM147

NOMO, junto con NCLN y TMEM147, forman un complejo proteico localizado en el retículo endoplásmico que inhibe la vía de señalización Nodal, una vía de transducción de señales que mantiene la pluripotencia en las células madre embrionarias humanas^{163,164}.

Varios estudios han demostrado que la formación del complejo proteico NOMO/NCLN/TMEM147 está limitada por un mecanismo de regulación postranscripcional que determina el ensamblaje y formación del complejo¹⁶⁴. De hecho, la incorporación de estos componentes monoméricos en el complejo proteico facilita su estabilización, lo que conlleva un aumento significativo de su vida media¹⁶³. Sin embargo, aunque el modo de ensamblaje del complejo no se conoce bien, algunos autores han afirmado que NCLN juega un papel crítico en la formación del complejo proteico a través del control de los niveles intracelulares de NOMO y TMEM147, ya que estas formas monoméricas están sometidas a una rápida degradación proteolítica si se sintetizan en exceso¹⁶⁴.

Se ha descrito que la inactivación de cualquiera de las tres proteínas da lugar a una fuerte reducción de las otras dos. Sin embargo, la sobreexpresión de NOMO y TMEM147, solas o en combinación, no genera un aumento de los niveles de expresión de NCLN, lo que sugiere que NCLN puede ser el factor limitante en la formación de este complejo proteico^{163,164}.

En este trabajo, observamos que la inactivación de NOMO redujo fuertemente la expresión de NCLN a nivel proteico. Sin embargo, los niveles de TMEM147 no se modificaron en ninguna de las líneas celulares NOMO-KO estudiadas. Estos resultados confirman la desestabilización del complejo NOMO/NCLN, pero ponen de manifiesto que TMEM147 no es un factor esencial para su estabilización. Por tanto, nuestros resultados no apoyan la reducción de los niveles de TMEM147 tras la inactivación de NOMO, lo que podría explicarse por el cambio de contexto celular, ya que este hecho sólo está reportado durante el desarrollo embrionario¹⁶⁴. Nuestros resultados sugieren que, si bien el complejo NOMO/NCLN no consigue ensamblarse, TMEM147 no se une, pero tampoco es degradada por el proteasoma debido a la estabilidad que posee su forma monomérica en estas líneas celulares.

Además, encontramos que la desestabilización de NCLN en ausencia de NOMO no fue acompañada por una reducción de sus niveles de mRNA, como indican los resultados del microarrays de expresión, secuenciación de RNA y el análisis por qRT-PCR. Por tanto,

nuestros resultados muestran que la reducción de la proteína NCLN es causada por un mecanismo postranscripcional, lo que concuerda con los datos publicados por Dettmer *et al*¹⁶⁴. En este caso, existiría un mecanismo de regulación mediante el cual NOMO se une a NCLN a través de su dominio luminal y, seguidamente, se une a TMEM147 por su dominio citoplasmático. Así, la ausencia de NOMO estaría impidiendo el reclutamiento de NCLN para formar un complejo proteico estable, y estaría facilitando su degradación por el proteasoma, dada su inestabilidad en forma monomérica. Por su parte, TMEM147 no se reclutaría hasta que el complejo NOMO/NCLN no estuviese formado, pero no sería degradada. Esto sugiere que TMEM147 podría quedar secuestrada en algún compartimiento celular, o bien ser funcional en su forma monomérica. Son necesarios nuevos ensayos funcionales y de localización, con el fin de esclarecer la relación entre NOMO y TMEM147 en células adultas. Por tanto, el papel de *NOMO1* se asociaría con la estabilización del complejo NOMO/NCLN, sin depender del estado de TMEM147.

3. La inactivación de NOMO no modifica la actividad de la vía Nodal

Una vez dilucidado el efecto de la pérdida de *NOMO1* en la formación y estabilidad del complejo NOMO/NCLN/TMEM147, evaluamos su papel en la activación de la vía de señalización Nodal, una ruta esencial para la diferenciación celular durante el desarrollo embrionario¹⁶⁷. Aunque el papel que desempeña *NOMO1* en carcinogénesis no está claro, existen estudios que lo posicionan como un supresor tumoral, ya que el complejo NOMO/NCLN/TMEM147 actúa como antagonista de la vía de señalización Nodal durante el desarrollo embrionario.

La vía de señalización Nodal, un miembro de la familia TGF- β , está directamente relacionada con la plasticidad de las células tumorales y las *cancer stem cells*. Por ello, además de ser una ruta vital durante el desarrollo embrionario, también ejerce un papel importante en tumorigénesis^{170,220}.

La sobreexpresión de esta vía de señalización se ha vinculado con la aparición de distintos tipos de tumores¹⁷⁶, donde se ha relacionado con un aumento de la capacidad de proliferación, invasión y metástasis^{179,221}. Por ejemplo, la sobreexpresión de la vía Nodal está implicada en cáncer de mama, próstata, páncreas, melanoma, CCR, glioblastoma, carcinoma hepatocelular, cáncer de endometrio, carcinoma de células renales o cáncer de tiroides. En estos tipos tumorales, se ha descrito una proporcionalidad entre los niveles de expresión de *NOMO1* y el grado del tumor^{160,220}. Además, los pacientes cuyos tumores presentan una alta actividad de la vía Nodal tienen menor supervivencia^{179,222}. En un estudio realizado en líneas celulares de melanoma se ha demostrado que la regulación negativa de

la vía Nodal suprime el fenotipo de plasticidad y disminuye la capacidad invasiva y proliferativa de las células²²⁰. También, en un trabajo desarrollado en líneas celulares de cáncer de mama, el *knockdown* del ligando nodal disminuye la viabilidad celular, aumenta la apoptosis y provoca la pérdida de la capacidad de autorrenovación celular²²³.

Por otro lado, Cripto-1 es un correceptor fundamental para la activación de la vía Nodal, cuyo bloqueo se ha postulado como una diana potencial para inhibir la activación de la vía. Cripto-1 se sobre expresa en diferentes tipos tumorales, como en CCR, de mama, gástrico, ovario y pulmón¹⁷⁰. Además, se ha visto que antagonizar la vía Nodal a través de su inhibidor Lefty o de anticuerpos monoclonales reduce el potencial tumorigénico de las células²²⁰. Por tanto, la inhibición de NOMO podría sobre activar la vía de señalización Nodal, lo que lo sitúa como un posible supresor tumoral.

Debido al importante papel que desempeña la vía Nodal en CCR, la pérdida recurrente de *NOMO1* en EOCRC descrita por nuestro grupo, así como el papel de supresor tumoral que podría desempeñar este gen, nos propusimos estudiar el estado de todas las proteínas que integran la vía Nodal en las líneas celulares HCT-116 y HS-5, con y sin inactivación de *NOMO1*. En primer lugar, encontramos que la inactivación de *NOMO1* no afectó a la expresión de los receptores ACTRII-B y ALK4, ni del correceptor Cripto-1. Para estudiar el efecto en la actividad de la vía, analizamos la expresión de SMAD2 y SMAD3, tanto en su estado basal como fosforilado. Para ello, analizamos el estado de estas proteínas incubando previamente las células con el ligando Nodal con el fin de inducir la activación de la ruta. En este caso, la desestabilización del complejo antagonista formado por NOMO, NCLN y TMEM147 permitiría que el estado fosforilado de SMAD2 y SMAD3 se elevase. Sin embargo, no observamos cambios en la expresión de las proteínas SMAD2/3 y pSMAD2/3. Tampoco se identificaron cambios en el estado de SMAD4. Por tanto, la inactivación de NOMO no induce una mayor translocación de SMAD4 hacia el núcleo para activar los factores de transcripción responsables del fenotipo típico de los tumores con sobre activación de la vía Nodal.

Estos resultados sugieren que el posible efecto carcinogénico de la pérdida del gen *NOMO1* no está relacionado con su papel descrito como inhibidor de la vía Nodal. Por tanto, es esencial dilucidar otras vías biológicas o mecanismos moleculares que podrían estar regulados por la inactivación de *NOMO1* en EOCRC.

4. La inactivación de NOMO no modifica el estado *stem cell* ni la viabilidad celular en respuesta al tratamiento

Debido al posible papel que pudiera tener el gen *NOMO1* en la carcinogénesis colorrectal, quisimos estudiar si su pérdida afectaba al estado de diferenciación celular y a la viabilidad de las líneas celulares WT y *NOMO-KO*, tanto en condiciones basales como en respuesta al tratamiento quimioterápico.

En primer lugar, quisimos comprobar si la pérdida del gen *NOMO1* afectaba al estado de diferenciación celular. Para ello, analizamos mediante citometría de flujo la expresión de los marcadores de membrana CD24 y CD44. Estas dos proteínas se consideran marcadores de células madre de cáncer de colon (CCSC) y se asocian principalmente a los procesos de adhesión celular, migración y metástasis^{191,192}. Al igual que ocurre con la viabilidad celular, las células carentes de *NOMO* no ofrecieron cambios en la expresión de estos marcadores, por lo que el estado de diferenciación de las líneas *NOMO-KO* no habría cambiado. Sin embargo, los datos reportados por el RNA-seq indican que el marcador CD44 se encuentra sobreexpresado en las células carentes del gen *NOMO1*. Esta discrepancia entre los datos obtenidos por citometría de flujo y RNA-seq podría atribuirse a que el marcaje utilizado para la citometría se realiza frente a la proteína global, mientras que los datos de expresión de mRNA podrían verse afectados por la sobreexpresión de algún transcrito concreto. De hecho, se ha descrito recientemente cómo los defectos en el *splicing* facilitan el aumento de la isoforma CD44s, la cual es responsable de la EMT que se da en las células tumorales, así como de su mayor invasividad²²⁴.

En segundo lugar, se analizó la viabilidad celular de los clones WT y *NOMO-KO* de cada línea celular, previa incubación con dosis crecientes de 5-fluorouracilo, irinotecán, oxaliplatino y cisplatino. Estos fármacos, a excepción del cisplatino, forman parte de la quimioterapia convencional indicada para el tratamiento del CCR. Nuestros resultados mostraron que la inactivación de *NOMO* no afecta a la proliferación celular ni modifica la respuesta al tratamiento con 5-FU, irinotecán y oxaliplatino en las líneas celulares incluidas en este trabajo. Por tanto, no se puede hacer una indicación específica de alguno de estos fármacos para aquellos pacientes con EOCRC que presentasen delección del gen *NOMO1* en el tumor.

Sin embargo, las células carentes de *NOMO* mostraron resistencia al tratamiento con cisplatino. El fenotipo de resistencia de las células del colon al tratamiento con este fármaco es causado, principalmente, por alteraciones en la señalización del daño génico a través de p53 y/o por defectos en las vías de reparación del mismo²²⁵. En particular, la inactivación de p53 parece estar asociada a la quimiorresistencia y a un mal pronóstico. De

forma interesante, el perfil transcriptómico de las líneas celulares NOMO-KO empleadas en nuestro trabajo mostraron la desregulación de la vía de p53, lo cual podría explicar la resistencia observada en los ensayos de viabilidad. Por otro lado, la resistencia al tratamiento con cisplatino y no al oxaliplatino podría deberse al mecanismo de acción que presenta cada fármaco. De hecho, a diferencia del cisplatino, el oxaliplatino genera daños en el DNA que no son reconocidos por las proteínas del sistema MMR, lo que hace que la citotoxicidad inducida por este sea independiente del sistema de reparación^{226,227}. Esto explica la observación preclínica de que el oxaliplatino tenga una actividad sustancial en líneas celulares de cáncer de colon, normalmente deficientes en MMR y, por tanto, resistentes al cisplatino²²⁵. Por lo tanto, sería conveniente considerar el estado mutacional de *NOMO1* como un posible mecanismo de resistencia en aquellos tipos tumorales que no responden al tratamiento con cisplatino.

5. La inactivación del gen *NOMO1* genera un perfil transcripcional y proteómico diferencial en las líneas celulares HCT-116 y HS-5

Con el fin de determinar el papel que desempeña *NOMO1* en EOCRC, y dado que su inactivación no afecta a la actividad de la vía Nodal, decidimos estudiar los perfiles de expresión transcripcional y proteica de las líneas celulares WT y NOMO-KO.

En primer lugar, realizamos un análisis de la expresión de mRNA a través de dos aproximaciones: los microarrays de expresión y la secuenciación de RNA. En ambos ensayos, se unificaron los clones NOMO-KO y NOMO-WT de las dos líneas celulares para realizar la comparación del perfil de expresión, ya que de esta forma igualamos el fondo genético de cada línea celular y así los cambios observados podrían atribuirse específicamente a la pérdida de *NOMO1*. En el microarray de expresión, se compararon 4 clones NOMO-KO con 4 clones NOMO-WT. En el RNA-seq, se compararon 6 clones NOMO-KO con 6 clones NOMO-WT de las dos líneas celulares.

Nuestros resultados revelaron un perfil de expresión diferencial de los genes comúnmente desregulados en las líneas celulares HCT-116 y HS-5 debido a la inactivación de *NOMO1*, independientemente del ensayo utilizado. Sin embargo, los genes que aparecieron sobreexpresados o infraexpresados en las líneas carentes de *NOMO1* no mostraron una coincidencia total en los dos ensayos empleados. Este hecho podría explicarse por diferentes razones: i) el perfil transcriptómico varía a lo largo del tiempo, ya que las líneas celulares no solo adquieren mutaciones a medida que se dividen, sino que también la expresión de determinados genes responde a diversos estímulos que se dan en un momento y contexto celular concreto; ii) existe una diferencia cronológica en el momento

en el que se ha realizado cada ensayo; iii) los microarrays de expresión constituyen una técnica de hibridación de una mínima región de la secuencia codificante de cada gen, por lo que las alteraciones en el resto de la secuencia no son detectadas; iv) el RNA-seq se basa en la secuenciación completa del gen y presenta un mayor potencial para detectar pequeños cambios en los niveles de expresión del mismo.

Además, se realizó un análisis de enriquecimiento de rutas de señalización (GSEA) introduciendo el set de genes desregulados en cada uno de los ensayos. A pesar de la no coincidencia completa de los genes desregulados en cada aproximación, las rutas de señalización afectadas por los genes sobre e infra-expresados en el microarray de expresión y en el RNA-seq mostraron una estrecha similitud. Por ejemplo, las líneas celulares con inactivación de *NOMO1* presentan la desregulación de las rutas de inflamación, formación de vasos sanguíneos y de la vía de señalización de TNF α . Estos procesos se encuentran principalmente desregulados en la EII, la cual es un factor de riesgo para el desarrollo del CCR. También, de manera interesante, los cambios a nivel transcriptómico observados en las líneas carentes de *NOMO1* se asocian fundamentalmente con la alteración de los procesos de transición epitelio-mesénquima y migración celular. En ellos, participan los genes *QSOX1*, *LUM*, *VCAN*, *CXCL8*, *PTX3*, *EDIL3*, *LOXL1*, *BDNF*, *PCOLCE*, *CAPG*, *POSTN*, *CCN1*, *CCN2*, *CD44*, *SERPINE2*, *THBS1* y *VEGFA*, los cuales se encuentran desregulados en nuestras líneas celulares tras la inactivación de *NOMO1*. Estos procesos biológicos afectados son de especial relevancia en el contexto del EOCRC, ya que estos pacientes se caracterizan por desarrollar metástasis tempranas y presentar fenotipos tumorales más agresivos. Por lo tanto, la pérdida de *NOMO1* podría contribuir a un aumento de la capacidad metastásica en EOCRC.

Por este motivo, en primer lugar, decidimos estudiar el estado de E-Cadherina, y Vimentina, que son marcadores típicos de la EMT. En segundo lugar, también analizamos la expresión proteica de β -Catenina, dada su estrecha relación con los procesos de migración celular, invasión y metástasis. Acorde a nuestros resultados, estos marcadores no mostraron cambios significativos en su expresión tras la inactivación de *NOMO*, lo que sugiere que son otras proteínas las que podrían estar implicadas en la desregulación de la EMT y la migración celular observada en el perfil transcripcional de nuestras líneas celulares.

En este sentido, se realizó un análisis de LC-IMS/MS con el fin de dilucidar los cambios a nivel proteico que se estaban produciendo en las líneas celulares carentes de *NOMO*. Con este análisis se pudo identificar la desregulación de las proteínas CTND1 (delta-catenina), LMNB1 y HMGA1 en las células *NOMO-KO*, las cuales se asocian principalmente a los procesos de EMT, migración celular, invasión y metástasis. Por tanto, estos resultados

indicarían, nuevamente, que la pérdida de *NOMO1* podría contribuir a un aumento de la capacidad migratoria.

6. La inactivación de *NOMO* promueve la migración celular en EOCRC

Con todo esto, decidimos analizar el efecto de la pérdida de *NOMO1* en la capacidad de migración celular, debido a la alteración de las proteínas asociadas a la migración e invasión (CTND1, LMNB1, HMGA1) comentadas anteriormente. Interesantemente, encontramos que la inactivación de *NOMO1* aumentó significativamente ($p < 0,001$) la capacidad migratoria de las líneas celulares HCT-116 y HS-5 *knockout* para este gen. Por lo tanto, los datos proporcionados anteriormente por el transcriptoma y el proteoma podrían explicar la mayor capacidad migratoria observada en los clones *NOMO-KO*, y justifican la necesidad de estudiar en profundidad el papel que podrían desempeñar CTND1, LMNB1 y HMGA1 en el desarrollo del EOCRC.

CTND1 (infraexpresada en las dos líneas celulares carentes de *NOMO*) se ha asociado con la estabilización de E-Cadherina y con el mantenimiento de las uniones de adherencia celular²²⁸. Concretamente, el análisis del proteoma mostró la infra expresión de CTND1, entre otras proteínas, la cual está asociada con los procesos de uniones apicales. CTND1 participa en la vía de señalización Wnt/beta-catenina/CTNNB1 regulando la proliferación, migración y diferenciación celular de las células endoteliales en el crecimiento tumoral²²⁹. Otros estudios han descrito que la pérdida de CTND1 se asocia con un mal pronóstico y metástasis en el cáncer de mama ductal y se considera como un supresor tumoral²³⁰. En un estudio realizado en líneas celulares de carcinoma escamoso de cabeza y cuello (HNSCC), y de carcinoma oral de células escamosas (OSCC), la pérdida de CTND1 provoca un aumento de la migración celular, de la invasión y favorece la EMT^{231,232}.

La lamina B1 (LMNB1), infraexpresada en las dos líneas celulares *NOMO-KO* de nuestro estudio, se ha descrito como supresor tumoral en cáncer de pulmón. La supresión de LMNB1 en las células epiteliales de pulmón altera los procesos de EMT y promueve la migración celular, el crecimiento tumoral y la metástasis mediante la activación de la señalización RET/p38²³³.

HMGA1 se encuentra sobreexpresada en las células *NOMO-KO* de las líneas HCT-116 y HS-5. Este aumento de la expresión concuerda con lo observado por otros autores en tejidos de cáncer de cuello uterino, donde su expresión se correlaciona positivamente con la metástasis en los ganglios linfáticos y el estadio clínico avanzado. Además, se ha descrito

que HMGA1 aumenta el crecimiento tumoral y acelera la migración y la invasión en líneas celulares de cáncer de cuello uterino²³⁴.

Sin embargo, aunque un aumento de la migración inducido por la pérdida de *NOMO1* podría contribuir a la carcinogénesis colorrectal a través de la desregulación de CTND1, LMNB1 o HMGA1, no podemos excluir la posibilidad de que otras vías de señalización desreguladas por la pérdida de este gen puedan desempeñar un papel relevante en la patogénesis de la enfermedad.

7. La pérdida del gen *NOMO1* es una mutación *passenger* en EOCRC

Además, el sistema CRISPR/cas9 también se utilizó para crear un modelo animal condicional para *Nomo1*²⁰² que permitiera generar su inactivación de forma específica en el intestino, y así estudiar el desarrollo tumoral a lo largo del tiempo. La elección de un modelo condicional se debe a que la pérdida de *Nomo1* en estado embrionario es letal. Así, para estudiar el papel de *NOMO1* en la carcinogénesis colorrectal se utilizó el sistema Cre-loxp, con el cual se indujo la delección del gen en un momento y contexto celular específico. El funcionamiento del sistema Cre-loxp *in vivo* ha revelado resultados satisfactorios en cuanto a la edición génica en otros tejidos, por lo que es considerado como un modelo óptimo para establecer modelos KO de diversos genes^{202,235,236}.

Por ejemplo, en un estudio llevado a cabo por Huang *et al.* se empleó este modelo animal para observar la formación tumoral en el colon de ratones condicionales para los genes *APC* y *SMAD4* que presentaban mutación de *KRAS*²³⁶. A diferencia de lo observado por el grupo de Huang, los resultados de nuestro modelo *in vivo* mostraron que la pérdida de *Nomo1* en el intestino de los ratones analizados no contribuyó al desarrollo tumoral después de 20 meses de seguimiento. Este resultado indica de manera firme que la pérdida de *NOMO1* podría no ser una mutación *driver* del cáncer colorrectal. En este caso, cabe destacar que la influencia del contexto celular y medioambiental que presentan los modelos animales difiere, en gran medida, del papel que tiene el ambiente sobre las células del colon humano.

Bajo la premisa de que la pérdida de *NOMO1* podría tener un papel *passenger* en la génesis del EOCRC, decidimos desarrollar un modelo de carcinogénesis colorrectal mediante la administración semanal de DMH. La DMH es un potente carcinógeno que induce la formación de adenomas y adenocarcinomas en el colon de roedores^{205–208}, y que presenta gran similitud con el CCR humano en cuanto a su morfología, patrones de crecimiento y manifestaciones clínicas²⁰⁹. Desde el punto de vista molecular, este agente provoca la

metilación del DNA, generando mutaciones puntuales en genes que afectan a la proliferación, diferenciación y apoptosis de las células epiteliales del colon^{210–212}.

Tras 20 semanas de tratamiento con DMH, no se observaron cambios significativos a nivel colorrectal en aquellos animales que presentaban delección heterocigota de *Nomo1* (n=4) en el colon y en los animales del grupo control (n=5). Sin embargo, fue de especial interés que los únicos dos animales que progresaron a adenocarcinoma presentaban la delección heterocigota de *Nomo1*: uno de ellos presentó un adenocarcinoma *in situ* intramucoso en el colon, y el otro un adenocarcinoma de pulmón con patrón papilar. Interesantemente, estas dos lesiones se localizaron en tejidos donde la Cre se expresa de manera constitutiva, por lo que la delección heterocigota de *Nomo1* se habría producido. Estos datos sugieren que la pérdida de *Nomo1* podría estar desempeñando un papel secundario durante la carcinogénesis colorrectal, facilitando la malignización del tumor. El resto de los animales del grupo control no presentaron signos de neoplasia maligna, si bien se puede afirmar que el modelo de carcinogénesis había funcionado correctamente ya que la mayoría de los animales presentaban adenomas. Cabe destacar que el número de ratones incluidos en este modelo no fue lo suficientemente elevado como para concluir sobre el papel que desempeña *NOMO1* en la génesis del CCR.

Por tanto, la pérdida en heterocigosis de *NOMO1* podría contribuir a la aparición de una lesión pre-cancerosa en las células del colon, como ocurre con el gen *APC*. La inactivación de *APC* bloquea la diferenciación de las células intestinales y promueve la proliferación, conduciendo a la formación de adenomas en el colon que predisponen al desarrollo del CCR. En nuestro caso, la pérdida de *NOMO1* podría favorecer la proliferación celular permitiendo la aparición de adenomas, y una segunda mutación en otro gen sería el responsable de la transformación maligna. Además, un estudio preliminar de nuestro grupo, realizado en una serie de pólipos de pacientes menores de 50 años, demostró que más del 40% presentaban delección heterocigota de *NOMO1*. Este hecho situaría la pérdida de este gen como un evento precoz en el desarrollo del CCR. Por este motivo, sería de especial interés caracterizar la pérdida de *NOMO1* en un modelo animal con inactivación del gen *APC* (ratones *APC*^{min/+}; *Nomo1*^{Flox}). En resumen, los resultados proporcionados por nuestros modelos *in vivo* indican que la pérdida de *NOMO1* no contribuye directamente al desarrollo de tumores colorrectales en ratones.

En conjunto, nuestros datos sugieren que *NOMO1*, a pesar de estar delecionado en más del 80% de los tumores EOCRC¹⁵⁶, desempeña un papel *passenger* en el desarrollo de esta enfermedad. También es importante considerar que otros genes codificantes o no codificantes, localizados en la misma región cromosómica (16p13.11-13.12), podrían ser

relevantes en la carcinogénesis colorrectal de inicio temprano. En consecuencia, se necesitan más estudios mecanísticos para comprender el papel que desempeña *NOMO1* en el desarrollo del EOCRC, así como los mecanismos moleculares implicados en la carcinogénesis colorrectal de inicio temprano.

8. La longitud telomérica como un marcador predictivo del CCR

Por último, en este trabajo quisimos explorar nuevos marcadores predictivos en línea germinal que pudieran ser incluidos en los programas de cribado del CCR, de manera que facilitasen la detección de la población de riesgo de una manera precoz. Especialmente, el EOCRC presenta la limitación de que estos pacientes no son incluidos en las estrategias de prevención del CCR, ya que no cumplen criterios de edad, con la excepción de aquellos sujetos que tienen antecedentes familiares de CCR. La *European Society of Gastrointestinal Endoscopy* (ESGE) solo recomienda el cribado antes de los 50 años cuando el sujeto presenta al menos dos familiares de primer grado con CCR, o al menos un familiar de primer grado menor de 50 años con CCR²³⁷. Además, el EOCRC de tipo esporádico abarca casi el 80% de todos los casos de EOCRC, por lo que este grupo de sujetos no disponen de herramientas útiles que faciliten la prevención de la enfermedad. Por esta razón, quisimos analizar el estado de los telómeros en este grupo de pacientes, e investigar si el acortamiento telomérico podría predisponer al desarrollo del CCR, y así utilizar la LT como una herramienta útil en los programas de cribado.

En primer lugar, decidimos analizar mediante RT-qPCR la LT de leucocitos procedentes de SP de 87 casos esporádicos de EOCRC y 109 individuos sanos. Cabe destacar que, en el momento de la toma de muestra, ninguno de los pacientes había recibido tratamiento quirúrgico, neoadyuvante o adyuvante. Nuestros resultados mostraron que la LT de los pacientes con EOCRC fue significativamente menor ($p < 0,001$) que la LT de la población control. Así, la longitud media del telómero por cada célula diploide fue de 122 kb en EOCRC y de 296 kb en la población control. En un estudio sobre el acortamiento telomérico realizado por Lisa Boardman *et al.*, se observó que los telómeros más largos aumentaban el riesgo de desarrollar EOCRC, mientras que los telómeros más cortos incrementaban el riesgo de LOCRC¹⁴³. Aunque estos hallazgos contrastan con nuestros resultados, debemos tener en cuenta que nuestro estudio de casos y controles sólo se centra en la población menor de 50 años, ya que el objetivo principal del estudio es el de establecer la LT como una herramienta útil para detectar precozmente a los sujetos con riesgo de desarrollar EOCRC. Además, los telómeros de los leucocitos se acortan sistemáticamente con el paso de los años, posicionándose como un claro factor de envejecimiento celular, lo que podría explicar el mayor acortamiento observado por el grupo

de Lisa Boardman en los pacientes con LOCRC. Estos pacientes podrían tener un importante acortamiento telomérico debido a la edad, que conduce a una situación de crisis en los telómeros y a una inestabilidad cromosómica, permitiendo la malignización de las células. Por lo tanto, si consideramos el cáncer como una enfermedad del adulto mayor, la existencia de un acortamiento telomérico en línea germinal, en sujetos menores de 50 años, podría generar una situación de envejecimiento prematuro que aumenta el riesgo de desarrollar EOCRC.

En segundo lugar, estudiamos la distribución genotípica del SNP rs2853669 (A>G) del gen *hTERT*, el cual ha sido asociado con el desarrollo de diferentes tipos de tumores como gliomas, melanomas, cáncer de mama, carcinoma hepatocelular o cáncer de pulmón²³⁸. En estos casos, se ha evidenciado una asociación entre el genotipo GG y el desarrollo tumoral a través de una ganancia de función de la telomerasa, la cual promueve la inmortalización celular por elongación de los telómeros. En nuestro estudio, identificamos una asociación entre el alelo G y el riesgo de desarrollar EOCRC. La menor frecuencia del alelo G estaría relacionada con una pérdida de la actividad de la telomerasa, por lo que existiría un acortamiento telomérico que genera un envejecimiento celular prematuro, el cual predispone al desarrollo del EOCRC.

Conclusiones



Conclusiones

Primera. En relación con la caracterización funcional de la pérdida del gen *NOMO1*:

- La inactivación de *NOMO* genera la desestabilización del complejo que forma con Nicalina, pero no afecta a la expresión de TMEM147 en EOCRC. La regulación que ejerce *NOMO* sobre Nicalina se lleva a cabo mediante un mecanismo post-transcripcional, sin afectar a los niveles proteicos de TMEM147.
- La inactivación de *NOMO* no afecta a la actividad de la vía Nodal. La expresión de las proteínas que conforman la vía de señalización Nodal no se ve alterada en los clones *knockout* para *NOMO*. Esto sugiere que el papel de *NOMO* en la carcinogénesis colorrectal debe estar mediado por otra ruta de señalización no conocida hasta el momento.
- La pérdida del gen *NOMO1* en las líneas celulares empleadas en este trabajo no modifica la viabilidad celular ni el estado de diferenciación celular.
- Las líneas celulares carentes de *NOMO1* muestran un perfil transcripcional diferencial, donde se incluyen genes involucrados en la transición epitelio-mesénquima y en la migración celular.
- La inactivación de *NOMO1* promueve la migración celular. El aumento de la capacidad migratoria en las células carentes de este gen podría asociarse con el incremento de la capacidad invasiva y metastásica que presentan los tumores colorrectales de inicio temprano.
- Las líneas celulares con inactivación de *NOMO* muestran un perfil de expresión proteico diferencial. La desregulación de CTND1, LMNB1 y HMGA1 en las células *knockout* para *NOMO* podría asociarse con el incremento de la capacidad migratoria observada.

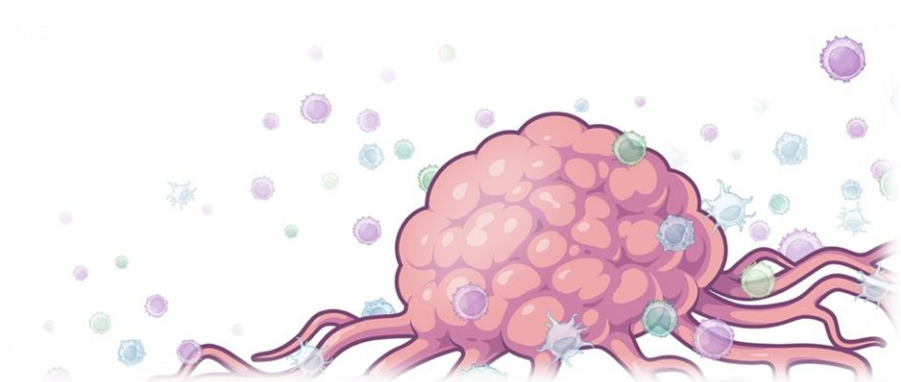
Segunda. En relación con el efecto tumorigénico de la pérdida del gen *NOMO1*:

- El sistema Cre-loxp funciona eficientemente para inducir la delección de *Nomo1* en un modelo animal condicional, en un momento y contexto celular específico.
- La pérdida de *Nomo1* no induce la formación de cáncer en modelos murinos, lo que sugiere que este gen actúa como *passenger* del cáncer colorrectal.

Tercera. La inactivación de *NOMO* no modifica la respuesta al tratamiento con 5-fluorouracilo, irinotecán y oxaliplatino en líneas celulares. No obstante, existe una resistencia de las células carentes de *NOMO* al tratamiento con cisplatino. Sería conveniente considerar el estado del gen *NOMO1* como un posible mecanismo de resistencia en aquellos tumores que no responden al tratamiento con cisplatino.

Cuarta. La medición de la longitud telomérica se posiciona como un método no invasivo que podría facilitar la identificación precoz de individuos con riesgo de desarrollar cáncer colorrectal, especialmente en aquellos individuos sin antecedentes familiares menores de 50 años.

Referencias



Referencias

1. Rubin, H. Fields and field cancerization: The preneoplastic origins of cancer. *BioEssays* **33**, 224–231 (2011).
2. Seyfried, T. N. & Huysentruyt, L. C. On the Origin of Cancer Metastasis. *Crit. Rev. Oncog.* **18**, 43 (2013).
3. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
5. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
6. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
7. Home | Redecan. <https://redecan.org/es>.
8. Definición de colon - Diccionario de cáncer del NCI - Instituto Nacional del Cáncer. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/colon>.
9. Arvelo, F., Sojo, F. & Cotte, C. Biology of colorectal cancer. *Ecancermedicalscience* **9**, (2015).
10. Schneider, N. I. & Langner, C. Prognostic stratification of colorectal cancer patients: current perspectives. *Cancer Manag. Res.* **6**, 291–300 (2014).
11. Conteduca, V., Sansonno, D., Russi, S. & Dammacco, F. Precancerous colorectal lesions (Review). *Int. J. Oncol.* **43**, 973–984 (2013).
12. Carethers, J. M. & Jung, B. H. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* **149**, 1177 (2015).
13. Jass, J. R. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* **50**, 113–130 (2007).
14. Strum, W. B. Colorectal Adenomas. <https://doi.org/10.1056/NEJMra1513581> **374**, 1065–1075 (2016).
15. East, J. E. et al. British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. *Gut* **66**, 1181 (2017).
16. Bae, S. I. & Kim, Y. S. Colon Cancer Screening and Surveillance in Inflammatory Bowel Disease. *Clin. Endosc.* **47**, 509 (2014).
17. Silva, S. M. arque. e. et al. INFLUENCE OF PATIENT AGE AND COLORECTAL POLYP SIZE ON HISTOPATHOLOGYFINDINGS. *Arq. Bras. Cir. Dig.* **27**, 109 (2014).
18. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J. Gastrointest. Oncol.* **3**, 153–173 (2012).
19. Marley, A. R. & Nan, H. Epidemiology of colorectal cancer. *Int. J. Mol. Epidemiol. Genet.* **7**, 105 (2016).
20. Li, J. et al. TNM staging of colorectal cancer should be reconsidered by T stage weighting. *World J. Gastroenterol.* **20**, 5104 (2014).

21. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
22. Age standardized (World) incidence rates, colorectal cancer, males, all ages. doi:10.2-15.2.
23. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
24. Aicr & WCRF. Diet, nutrition, physical activity and colorectal cancer.
25. Center, M. M., Jemal, A. & Ward, E. International trends in colorectal cancer incidence rates. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1688–1694 (2009).
26. Statistical Update 2018 | Human Development Reports. <https://hdr.undp.org/en/content/human-development-indices-indicators-2018-statistical-update>.
27. Ouakrim, D. A. *et al.* Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. *BMJ* **351**, (2015).
28. Schreuders, E. H. *et al.* Colorectal cancer screening: a global overview of existing programmes. *Gut* **64**, 1637–1649 (2015).
29. Siegel, R. L. *et al.* Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J. Natl. Cancer Inst.* **109**, (2017).
30. Bailey, C. E. *et al.* Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg.* **150**, 17–22 (2015).
31. Wolf, A. M. D. *et al.* Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA. Cancer J. Clin.* **68**, 250–281 (2018).
32. Kasi, P. M. *et al.* Rising Proportion of Young Individuals With Rectal and Colon Cancer. *Clin. Colorectal Cancer* **18**, e87–e95 (2019).
33. *Las cifras del cáncer en España 2022.* (2022).
34. Rodriguez-Salas, N. *et al.* Clinical relevance of colorectal cancer molecular subtypes. *Crit. Rev. Oncol. Hematol.* **109**, 9–19 (2017).
35. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. *CA. Cancer J. Clin.* **67**, 93–99 (2017).
36. Labianca, R. *et al.* Colon cancer. *Crit. Rev. Oncol. Hematol.* **74**, 106–133 (2010).
37. Seifert, G., Brocheriou, C., Cardesa, A. & Eveson, J. W. WHO International Histological Classification of Tumours Tentative Histological Classification of Salivary Gland Tumours. *Pathol. - Res. Pract.* **186**, 555–581 (1990).
38. Kuipers, E. J. *et al.* Colorectal cancer. *Nat. Rev. Dis. Prim.* **1**, (2015).
39. Keum, N. N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 713–732 (2019).
40. Sánchez-Gundín, J., Fernández-Carballido, A. M., Martínez-Valdivieso, L., Barreda-Hernández, D. & Torres-Suárez, A. I. New Trends in the Therapeutic Approach to

- Metastatic Colorectal Cancer. *Int. J. Med. Sci.* **15**, 659–665 (2018).
41. Messersmith, W. A. NCCN Guidelines Updates: Management of Metastatic Colorectal Cancer. *J. Natl. Compr. Canc. Netw.* **17**, 599–601 (2019).
 42. Brown, K. G. M., Solomon, M. J., Mahon, K. & O'Shannassy, S. Management of colorectal cancer. *BMJ* **366**, (2019).
 43. Labianca, R. *et al.* Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **24 Suppl 6**, (2013).
 44. Van Cutsem, E., Cervantes, A., Nordlinger, B., Arnold, D. & The ESMO Guidelines Working Group. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **25 Suppl 3**, iii1–iii9 (2014).
 45. G, C. *et al.* Phase III randomized trial of FOLFIRI versus FOLFOX4 in the treatment of advanced colorectal cancer: a multicenter study of the Gruppo Oncologico Dell'Italia Meridionale. *J. Clin. Oncol.* **23**, 4866–4875 (2005).
 46. Vera, R. *et al.* Current controversies in the management of metastatic colorectal cancer. *Cancer Chemother. Pharmacol.* **76**, 659–677 (2015).
 47. Cassidy, J. *et al.* XELOX (capecitabine plus oxaliplatin): active first-line therapy for patients with metastatic colorectal cancer. *J. Clin. Oncol.* **22**, 2084–2091 (2004).
 48. Goldberg, R. M. *et al.* A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J. Clin. Oncol.* **22**, 23–30 (2004).
 49. Tiwari, A. *et al.* Novel targeting approaches and signaling pathways of colorectal cancer: An insight. *World J. Gastroenterol.* **24**, 4428–4435 (2018).
 50. Krishnamurthy, N. & Kurzrock, R. Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. *Cancer Treat. Rev.* **62**, 50–60 (2018).
 51. Xie, Y. H., Chen, Y. X. & Fang, J. Y. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct. Target. Ther.* **5**, (2020).
 52. Amersi, F., Agustin, M. & Ko, C. Y. Colorectal cancer: epidemiology, risk factors, and health services. *Clin. Colon Rectal Surg.* **18**, 133–140 (2005).
 53. Keller, D. S., Windsor, A., Cohen, R. & Chand, M. Colorectal cancer in inflammatory bowel disease: review of the evidence. *Tech. Coloproctol.* **23**, 3–13 (2019).
 54. Shussman, N. & Wexner, S. D. Colorectal polyps and polyposis syndromes. *Gastroenterol. Rep.* **2**, 1–15 (2014).
 55. Yao, C., Nash, G. F. & Hickish, T. Management of colorectal cancer and diabetes. *J. R. Soc. Med.* **107**, 103–109 (2014).
 56. Peeters, P. J. H. L., Bazelier, M. T., Leufkens, H. G. M., De Vries, F. & De Bruin, M. L. The risk of colorectal cancer in patients with type 2 diabetes: associations with treatment stage and obesity. *Diabetes Care* **38**, 495–502 (2015).
 57. Valle, L. Genetic predisposition to colorectal cancer: Where we stand and future perspectives. *World J. Gastroenterol.* **20**, 9828 (2014).

58. Wong, S. H. & Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 690–704 (2019).
59. Saus, E., Iraola-Guzmán, S., Willis, J. R., Brunet-Vega, A. & Gabaldón, T. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. *Mol. Aspects Med.* **69**, 93–106 (2019).
60. Sánchez-Alcoholado, L. *et al.* The Role of the Gut Microbiome in Colorectal Cancer Development and Therapy Response. *Cancers (Basel)*. **12**, (2020).
61. Wieczorska, K., Stolarek, M. & Stec, R. The Role of the Gut Microbiome in Colorectal Cancer: Where Are We? Where Are We Going? *Clin. Colorectal Cancer* **19**, 5–12 (2020).
62. Cheng, Y., Ling, Z. & Li, L. The Intestinal Microbiota and Colorectal Cancer. *Front. Immunol.* **11**, 3100 (2020).
63. Article, R. Increasing Incidence of Early-Onset Colorectal Cancer. 13–15 (2022) doi:10.1056/NEJMra2200869.
64. Worldwide incidence and mortality of colorectal cancer and human development index (HDI): an ecological study - WCRJ. <https://www.wcrj.net/article/1433>.
65. Peterse, E. F. P. *et al.* The impact of the rising colorectal cancer incidence in young adults on the optimal age to start screening: Microsimulation analysis I to inform the American Cancer Society colorectal cancer screening guideline. *Cancer* **124**, 2964–2973 (2018).
66. Vuik, F. E. R. *et al.* Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut* **68**, (2019).
67. Schmuck, R. *et al.* Gender comparison of clinical, histopathological, therapeutic and outcome factors in 185,967 colon cancer patients. *Langenbeck's Arch. Surg.* **405**, 71–80 (2020).
68. Ellis, L. *et al.* Colorectal Cancer Incidence Trends by Age, Stage, and Racial/Ethnic Group in California, 1990-2014. *Cancer Epidemiol. Biomarkers Prev.* **27**, 1011–1018 (2018).
69. Aykan, N. F. Red Meat and Colorectal Cancer. *Oncol. Rev.* **9**, 38–44 (2015).
70. Aran, V., Victorino, A. P., Thuler, L. C. & Ferreira, C. G. Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. *Clin. Colorectal Cancer* **15**, 195–203 (2016).
71. Aicr & WCRF. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective A summary of the Third Expert Report.
72. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz. Gastroenterol.* **14**, 89–103 (2019).
73. Murphy, N. *et al.* Lifestyle and dietary environmental factors in colorectal cancer susceptibility. *Mol. Aspects Med.* **69**, 2–9 (2019).
74. Carr, P. R. *et al.* Lifestyle factors and risk of sporadic colorectal cancer by microsatellite instability status: a systematic review and meta-analyses. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **29**, 825–834 (2018).
75. Biomarkers, E. G.-C. E. and P. & 2001, undefined. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. *AACR*

- (2001).
76. Yurgelun, M. B. *et al.* Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J. Clin. Oncol.* **35**, 1086–1095 (2017).
 77. DeRycke, M. S. *et al.* Targeted sequencing of 36 known or putative colorectal cancer susceptibility genes. *Mol. Genet. genomic Med.* **5**, 553–569 (2017).
 78. AlDubayan, S. H. *et al.* Inherited DNA-Repair Defects in Colorectal Cancer. *Am. J. Hum. Genet.* **102**, 401–414 (2018).
 79. Win, A. K. *et al.* Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. *Cancer Epidemiol. Biomarkers Prev.* **26**, 404–412 (2017).
 80. Møller, P. *et al.* Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut* **66**, 464–472 (2017).
 81. Watson, P. *et al.* The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int. J. cancer* **123**, 444–449 (2008).
 82. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895-2015. *Nat. Rev. Cancer* **15**, 181–194 (2015).
 83. Kalady, M. F. *et al.* Risk of colorectal adenoma and carcinoma after colectomy for colorectal cancer in patients meeting Amsterdam criteria. *Ann. Surg.* **252**, 507–511 (2010).
 84. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nat.* 2012 4877407 **487**, 330–337 (2012).
 85. Schwitalle, Y. *et al.* Immune response against frameshift-induced neopeptides in HNPCC patients and healthy HNPCC mutation carriers. *Gastroenterology* **134**, 988–997 (2008).
 86. Ligtenberg, M. J. L. *et al.* Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat. Genet.* **41**, 112–117 (2009).
 87. Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819 (1993).
 88. Giang, H. *et al.* 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.* **19**, 4–9 (2018).
 89. Weren, R. D. A. *et al.* A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat. Genet.* **47**, 668–671 (2015).
 90. Adam, R. *et al.* Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *Am. J. Hum. Genet.* **99**, 337–351 (2016).
 91. Win, A. K. *et al.* Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. *Gastroenterology* **146**, (2014).
 92. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45**, 136–143

- (2013).
93. Valle, L. *et al.* HHS Public Access. **247**, 574–588 (2019).
 94. Gammon, A., Jasperson, K., Kohlmann, W. & Burt, R. W. Hamartomatous polyposis syndromes. *Best Pract. Res. Clin. Gastroenterol.* **23**, 219–231 (2009).
 95. Jaeger, E. *et al.* Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat. Genet.* **44**, 699–703 (2012).
 96. Rohlin, A. *et al.* GREM1 and POLE variants in hereditary colorectal cancer syndromes. *Genes. Chromosomes Cancer* **55**, 95–106 (2016).
 97. Paul, J. & Medema, J. P. Cancer stem cells: The challenges ahead. *Nat. Cell Biol.* **2013** **15**, 338–344 (2013).
 98. Nassar, D. & Blanpain, C. Cancer Stem Cells: Basic Concepts and Therapeutic Implications. *Annu. Rev. Pathol.* **11**, 47–76 (2016).
 99. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
 100. Pino, M. S. & Chung, D. C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059–2072 (2010).
 101. Markowitz, S. D. & Bertagnolli, M. M. Molecular Basis of Colorectal Cancer. *N. Engl. J. Med.* **361**, 2449–2460 (2009).
 102. Tsang, A. H. F. *et al.* Current and future molecular diagnostics in colorectal cancer and colorectal adenoma. *World J. Gastroenterol.* **20**, 3847–3857 (2014).
 103. Baker, S. J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217–221 (1989).
 104. Thiagalingam, S. *et al.* Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat. Genet.* **13**, 343–346 (1996).
 105. Markowitz, S. *et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336–1338 (1995).
 106. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
 107. Diep, C. B. *et al.* The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes. Chromosomes Cancer* **45**, 31–41 (2006).
 108. Jasmine, F. *et al.* A Genome-Wide Study of Cytogenetic Changes in Colorectal Cancer Using SNP Microarrays: Opportunities for Future Personalized Treatment. *PLoS One* **7**, 31968 (2012).
 109. Baudis, M. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* **7**, 1–15 (2007).
 110. Geiersbach, K. B. & Samowitz, W. S. Microsatellite instability and colorectal cancer. *Arch. Pathol. Lab. Med.* **135**, 1269–1277 (2011).
 111. Meyer, L. A., Broaddus, R. R. & Lu, K. H. Endometrial cancer and Lynch syndrome: clinical

- and pathologic considerations. *Cancer Control* **16**, 14–22 (2009).
112. Nayak, S. S. *et al.* Prevalence estimation of microsatellite instability in colorectal cancers using tissue microarray based methods – A tertiary care center experience. *Indian J. Pathol. Microbiol.* **61**, 520 (2018).
 113. Kaiser, J. C., Meckbach, R. & Jacob, P. Genomic instability and radiation risk in molecular pathways to colon cancer. *PLoS One* **9**, (2014).
 114. Domingo, E. *et al.* BRAF-V600E is not involved in the colorectal tumorigenesis of HNPCC in patients with functional MLH1 and MSH2 genes. *Oncogene* **24**, 3995–3998 (2005).
 115. Frouws, M. A. *et al.* The influence of BRAF and KRAS mutation status on the association between aspirin use and survival after colon cancer diagnosis. *PLoS One* **12**, (2017).
 116. Venderbosch, S. *et al.* Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: a pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. *Clin. Cancer Res.* **20**, 5322–5330 (2014).
 117. Wang, Y. *et al.* Distinct impacts of KRAS, NRAS and BRAF mutations on survival of patients with metastatic colorectal cancer. https://doi.org/10.1200/JCO.2018.36.15_suppl.3513 **36**, 3513–3513 (2018).
 118. Wang, Y., Liu, D., Jin, X., Song, H. & Lou, G. Genome-wide characterization of aberrant DNA methylation patterns and the potential clinical implications in patients with endometrial cancer. *Pathol. Res. Pract.* **215**, 137–143 (2019).
 119. Magzoub, M. M., Prunello, M., Brennan, K. & Gevaert, O. The impact of DNA methylation on the cancer proteome. *PLoS Comput. Biol.* **15**, (2019).
 120. Hesson, L. B. *et al.* CpG island promoter hypermethylation of a novel Ras-effector gene RASSF2A is an early event in colon carcinogenesis and correlates inversely with K-ras mutations. *Oncogene* **24**, 3987–3994 (2005).
 121. Cohen, Y. *et al.* Hypermethylation of CpG island loci of multiple tumor suppressor genes in retinoblastoma. *Exp. Eye Res.* **86**, 201–206 (2008).
 122. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* **38**, 787–793 (2006).
 123. Ogino, S. *et al.* Molecular correlates with MGMT promoter methylation and silencing support CpG island methylator phenotype-low (CIMP-low) in colorectal cancer. *Gut* **56**, 1564–1571 (2007).
 124. Shen, L. *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18654–18659 (2007).
 125. Ogino, S. *et al.* Combined Analysis of COX-2 and p53 Expressions Reveals Synergistic Inverse Correlations with Microsatellite Instability and CpG Island Methylator Phenotype in Colorectal Cancer. *Neoplasia* **8**, 458 (2006).
 126. Ogino, S. *et al.* CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut* **55**, 1000–1006 (2006).

127. Suzuki, H., Yamamoto, E., Maruyama, R., Niinuma, T. & Kai, M. Biological significance of the CpG island methylator phenotype. *Biochem. Biophys. Res. Commun.* **455**, 35–42 (2014).
128. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
129. Jacobs, D. *et al.* Defining Early-Onset Colon and Rectal Cancers. *Front. Oncol.* **8**, (2018).
130. Lui, R. N. *et al.* Global Increasing Incidence of Young-Onset Colorectal Cancer Across 5 Continents: A Joinpoint Regression Analysis of 1,922,167 Cases. *Cancer Epidemiol. Biomarkers Prev.* **28**, 1275–1282 (2019).
131. Bhandari, A., Woodhouse, M. & Gupta, S. Colorectal cancer is a leading cause of cancer incidence and mortality among adults younger than 50 years in the USA: a SEER-based analysis with comparison to other young-onset cancers. *J. Investig. Med.* **65**, 311–315 (2017).
132. Saraste, D., Järås, J. & Martling, A. Population-based analysis of outcomes with early-age colorectal cancer. *Br. J. Surg.* **107**, 301–309 (2020).
133. Kneuert, P. J. *et al.* Overtreatment of young adults with colon cancer: more intense treatments with unmatched survival gains. *JAMA Surg.* **150**, 402–409 (2015).
134. O'Connell, J. B., Maggard, M. A., Livingston, E. H. & Yo, C. K. Colorectal cancer in the young. *Am. J. Surg.* **187**, 343–348 (2004).
135. Berger, N. A. *et al.* Cancer in the Elderly. *Trans. Am. Clin. Climatol. Assoc.* **117**, 147 (2006).
136. Willauer, A. N. *et al.* Clinical and molecular characterization of early-onset colorectal cancer. *Cancer* **125**, 2002–2010 (2019).
137. Meyer, J. E. *et al.* Increasing incidence of rectal cancer in patients aged younger than 40 years: An analysis of the surveillance, epidemiology, and end results database. *Cancer* **116**, 4354–4359 (2010).
138. McCawley, N. *et al.* Mucinous Rectal Adenocarcinoma Is Associated with a Poor Response to Neoadjuvant Chemoradiotherapy: A Systematic Review and Meta-analysis. *Dis. Colon Rectum* **59**, 1200–1208 (2016).
139. Perea, J. *et al.* Early-onset colorectal cancer is an easy and effective tool to identify retrospectively Lynch syndrome. *Ann. Surg. Oncol.* **18**, 3285–3291 (2011).
140. Mauri, G. *et al.* Early-onset colorectal cancer in young individuals. *Mol. Oncol.* **13**, 109–131 (2019).
141. Ballester, V., Rashtak, S. & Boardman, L. Clinical and molecular features of young-onset colorectal cancer. *World J. Gastroenterol.* **22**, 1736 (2016).
142. Stigliano, V., Sanchez-Mete, L., Martayan, A. & Anti, M. Early-onset colorectal cancer: a sporadic or inherited disease? *World J. Gastroenterol.* **20**, 12420–12430 (2014).
143. Boardman, L. A. *et al.* The association of telomere length with colorectal cancer differs by the age of cancer onset. *Clin. Transl. Gastroenterol.* **5**, (2014).
144. Mione, C., Vertecchi, E., Rizzo, A. & Salvati, E. Telomere Targeting Approaches in

- Cancer: Beyond Length Maintenance. *Int. J. Mol. Sci.* 2022, Vol. 23, Page 3784 **23**, 3784 (2022).
145. Serebriiskii, I. G. *et al.* Comprehensive characterization of RAS mutations in colon and rectal cancers in old and young patients. *Nat. Commun.* **10**, (2019).
 146. Yiu, R., Qiu, H., Lee, S. H. & García-Aguilar, J. Mechanisms of microsatellite instability in colorectal cancer patients in different age groups. *Dis. Colon Rectum* **48**, 2061–2069 (2005).
 147. Antelo, M. *et al.* A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PLoS One* **7**, (2012).
 148. Kirzin, S. *et al.* Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. *PLoS One* **9**, (2014).
 149. Lieu, C. H. *et al.* Comprehensive genomic landscapes in early and later onset colorectal cancer. *Clin. Cancer Res.* **25**, 5852–5858 (2019).
 150. Perea, J. *et al.* Age at onset should be a major criterion for subclassification of colorectal cancer. *J. Mol. Diagn.* **16**, 116–126 (2014).
 151. Silla, I. O. *et al.* Early-onset colorectal cancer: A separate subset of colorectal cancer. *World J. Gastroenterol.* **20**, 17288 (2014).
 152. Banerjea, A., Hands, R. E., Powar, M. P., Bustin, S. A. & Dorudi, S. Microsatellite and chromosomal stable colorectal cancers demonstrate poor immunogenicity and early disease recurrence. *Colorectal Dis.* **11**, 601–608 (2009).
 153. Arriba, M. *et al.* DNA copy number profiling reveals different patterns of chromosomal instability within colorectal cancer according to the age of onset. *Mol. Carcinog.* **55**, 705–716 (2016).
 154. Arriba, M. *et al.* Unsupervised Analysis of Array Comparative Genomic Hybridization Data from Early-Onset Colorectal Cancer Reveals Equivalence with Molecular Classification and Phenotypes. *Neoplasia* **19**, 28 (2017).
 155. Berg, M. *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* **9**, 100 (2010).
 156. Perea, J. *et al.* NOMO-1 gene is deleted in early-onset colorectal cancer. *Oncotarget* **8**, 24429–24436 (2017).
 157. Ramalingam, A. *et al.* 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders. *J. Hum. Genet.* **56**, 541–544 (2011).
 158. Ingason, A. *et al.* Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol. Psychiatry* **16**, 17–25 (2011).
 159. Zhang, H. *et al.* Silencing of nodal modulator 1 inhibits the differentiation of P19 cells into cardiomyocytes. *Exp. Cell Res.* **331**, 369–376 (2015).
 160. Sandomenico, A. & Ruvo, M. Targeting Nodal and Cripto-1: Perspectives Inside Dual Potential Theranostic Cancer Biomarkers. *Curr. Med. Chem.* **26**, 1994–2050 (2019).
 161. Haffner, C. *et al.* Nicalin and its binding partner Nomo are novel Nodal signaling

- antagonists. *EMBO J.* **23**, 3041 (2004).
162. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
 163. 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. *J. Biol. Chem.* **282**, 10632–10638 (2007).
 164. Dettmer, U. *et al.* Transmembrane protein 147 (TMEM147) is a novel component of the Nicalin-NOMO protein complex. *J. Biol. Chem.* **285**, 26174–26181 (2010).
 165. Salnikov, E. *et al.* Lipid bilayer position and orientation of novel carprofens, modulators of γ -secretase in Alzheimer's disease. *Biochim. Biophys. acta. Biomembr.* **1860**, 2224–2233 (2018).
 166. Zhu, B. *et al.* ER-associated degradation regulates Alzheimer's amyloid pathology and memory function by modulating γ -secretase activity. *Nat. Commun.* **2017** *8*, 1–15 (2017).
 167. Shen, M. M. Nodal signaling: developmental roles and regulation. *Development* **134**, 1023–1034 (2007).
 168. Zinski, J., Tajer, B. & Mullins, M. C. TGF- β Family Signaling in Early Vertebrate Development. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).
 169. Wei, S. & Wang, Q. Molecular regulation of Nodal signaling during mesendoderm formation. *Acta Biochim. Biophys. Sin. (Shanghai)*. **50**, 74–81 (2018).
 170. Ishii, H., Afify, S. M., Hassan, G., Salomon, D. S. & Seno, M. Cripto-1 as a Potential Target of Cancer Stem Cells for Immunotherapy. *Cancers* **2021**, Vol. 13, Page 2491 **13**, 2491 (2021).
 171. A Nodal- and ALK4-independent signaling pathway activated by Cripto-1 through Glypican-1 and c-Src - PubMed. <https://pubmed.ncbi.nlm.nih.gov/12649175/>.
 172. Harms, P. W. & Chang, C. Tomoregulin-1 (TMEFF1) inhibits nodal signaling through direct binding to the nodal coreceptor Cripto. *Genes Dev.* **17**, 2624–2629 (2003).
 173. Iratni, R. *et al.* Inhibition of excess nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. *Science* **298**, 1996–1999 (2002).
 174. Arboreto, P., Cillo, M. & Leonardi, A. New Insights into Cancer Targeted Therapy: Nodal and Cripto-1 as Attractive Candidates. *Int. J. Mol. Sci.* **22**, (2021).
 175. Topczewska, J. M. *et al.* Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat. Med.* **12**, 925–932 (2006).
 176. Quail, D. F. *et al.* Nodal signalling in embryogenesis and tumourigenesis. *Int. J. Biochem. Cell Biol.* **45**, 885–898 (2013).
 177. Li, P. *et al.* Elevated expression of Nodal and YAP1 is associated with poor prognosis of gastric adenocarcinoma. *J. Cancer Res. Clin. Oncol.* **142**, 1765–1773 (2016).
 178. Park, K. S. *et al.* High cripto-1 and low miR-205 expression levels as prognostic markers in early stage non-small cell lung cancer. *Lung Cancer* **116**, 38–45 (2018).
 179. Gong, Y. *et al.* Nodal promotes the self-renewal of human colon cancer stem cells via an autocrine manner through Smad2/3 signaling pathway. *Biomed Res. Int.* **2014**, (2014).

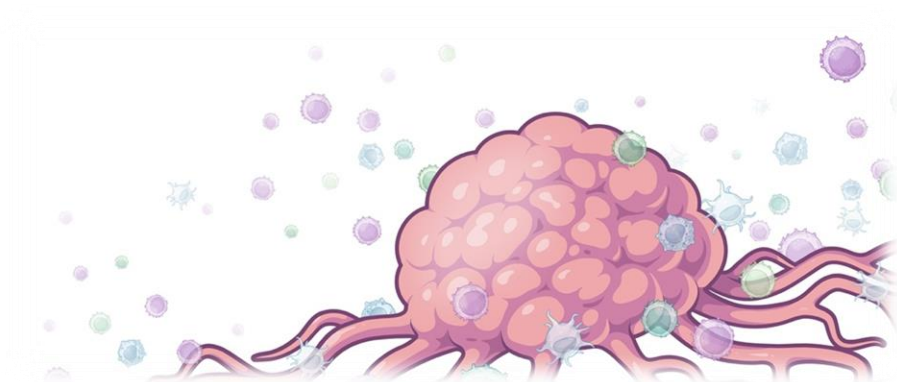
180. Strizzi, L. *et al.* Epithelial mesenchymal transition is a characteristic of hyperplasias and tumors in mammary gland from MMTV-Cripto-1 transgenic mice. *J. Cell. Physiol.* **201**, 266–276 (2004).
181. Shani, G. *et al.* GRP78 and Cripto form a complex at the cell surface and collaborate to inhibit transforming growth factor beta signaling and enhance cell growth. *Mol. Cell. Biol.* **28**, 666–677 (2008).
182. Perea, J. *et al.* Cohort profile: the Spanish Early-onset Colorectal Cancer (SECOC) cohort: a multicentre cohort study on the molecular basis of colorectal cancer among young individuals in Spain. *BMJ Open* **11**, e055409 (2021).
183. Arya, M. *et al.* Basic principles of real-time quantitative PCR. *Expert Rev. Mol. Diagn.* **5**, 209–219 (2005).
184. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
185. Díez, P. *et al.* Integration of Proteomics and Transcriptomics Data Sets for the Analysis of a Lymphoma B-Cell Line in the Context of the Chromosome-Centric Human Proteome Project. *J. Proteome Res.* **14**, 3530–3540 (2015).
186. Meier, F. *et al.* Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).
187. Lawrence, R. T., Searle, B. C., Llovet, A. & Villén, J. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat. Methods* **2016** 135 **13**, 431–434 (2016).
188. Stokes, M. P. *et al.* PTMScan Direct: Identification and Quantification of Peptides from Critical Signaling Proteins by Immunoaffinity Enrichment Coupled with LC-MS/MS. *Mol. Cell. Proteomics* **11**, 187–201 (2012).
189. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199 (2019).
190. T, M. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **65**, 55–63 (1983).
191. Yeung, T. M., Gandhi, S. C., Wilding, J. L., Muschel, R. & Bodmer, W. F. Cancer stem cells from colorectal cancer-derived cell lines. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3722 (2010).
192. Jaggupilli, A. & Elkord, E. Significance of CD44 and CD24 as cancer stem cell markers: An enduring ambiguity. *Clin. Dev. Immunol.* **2012**, (2012).
193. Liang, C. C., Park, A. Y. & Guan, J. L. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.* **2007** 22 **2**, 329–333 (2007).
194. Adli, M. The CRISPR tool kit for genome editing and beyond. doi:10.1038/s41467-018-04252-2.
195. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and

- hypothetical mechanisms of action. *Biol. Direct* **1**, (2006).
196. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
 197. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
 198. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
 199. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
 200. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013** *8*, 2281–2308 (2013).
 201. Harms, D. W. *et al.* Mouse Genome Editing using CRISPR/Cas System. *Curr. Protoc. Hum. Genet.* **83**, 15.7.1 (2014).
 202. García-Tuñón, I. *et al.* Establishment of a conditional Nomo1 mouse model by CRISPR/Cas9 technology. *Mol. Biol. Rep.* **47**, 1381–1391 (2020).
 203. Durkin, M. E., Qian, X., Popescu, N. C. & Lowy, D. R. Isolation of Mouse Embryo Fibroblasts. *Bio-protocol* **3**, (2013).
 204. Perez, M., Mendez-Sanchez, L. & Garcia, J. Msi1-CRE transgenic mice: driving transgene expression to the putative gut stem cell Multiple primary colorectal cancer: Synchronous and metachronous colorectal cancer View project Creating mouse models to study human biology and human diseases View proje. *Conf. Pap. Transgenic Res.* (2011).
 205. Kohno, H., Suzuki, R., Sugie, S. & Tanaka, T. β -Catenin mutations in a mouse model of inflammation-related colon carcinogenesis induced by 1,2-dimethylhydrazine and dextran sodium sulfate. *Cancer Sci.* **96**, 69–76 (2005).
 206. Newell, L. E. & Heddle, J. A. The potent colon carcinogen, 1,2-dimethylhydrazine induces mutations primarily in the colon. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **564**, 1–7 (2004).
 207. Moulahoum, H. *et al.* Aberrant crypt foci are regionally affected by zinc treatment in a 1,2-dimethylhydrazine induced colon carcinogenesis model. *J. Trace Elem. Med. Biol.* **47**, 21–30 (2018).
 208. Tawfik, M. K. & Mohamed, M. I. Exenatide suppresses 1,2-dimethylhydrazine-induced colon cancer in diabetic mice: Effect on tumor angiogenesis and cell proliferation. *Biomed. Pharmacother.* **82**, 106–116 (2016).
 209. Gilbert, J. M. Experimental colorectal cancer as a model of human disease. *Ann. R. Coll. Surg. Engl.* **69**, 48–53 (1987).
 210. Blakey, D. H. *et al.* Detection of Nuclear Anomalies in the Colonic Epithelium of the Mouse. *Cancer Res.* **45**, 242–249 (1985).
 211. Choudhary, G. & Hansen, H. Human health perspective on environmental exposure to hydrazines: A review. *Chemosphere* **37**, 801–843 (1998).
 212. Ma, Q. Y., Williamson, K. E. & Rowlands, B. J. Variability of cell proliferation in the proximal

- and distal colon of normal rats and rats with dimethylhydrazine induced carcinogenesis. *World J. Gastroenterol.* **8**, 847–852 (2002).
213. Ali, M. S., Hussein, R. M. & Kandeil, M. A. The pro-oxidant, apoptotic and anti-angiogenic effects of selenium supplementation on colorectal tumors induced by 1,2-dimethylhydrazine in BALB/C mice. *Reports Biochem. Mol. Biol.* **8**, 216–226 (2019).
 214. Gurley, K. E., Moser, R. D. & Kemp, C. J. Induction of colon cancer in mice with 1,2-dimethylhydrazine. *Cold Spring Harb. Protoc.* **2015**, 814–815 (2015).
 215. Goswami, A. *et al.* Association study of leukocyte telomere length and genetic polymorphism within hTERT promoter with type 2 diabetes in Bangladeshi population. *Mol. Biol. Rep.* **48**, 285–295 (2021).
 216. M, A. *et al.* Intermediate-onset colorectal cancer: A clinical and familial boundary between both early and late-onset colorectal cancer. *PLoS One* **14**, (2019).
 217. N, A. *et al.* Rising incidence of early-onset colorectal cancer - a call to action. *Nat. Rev. Clin. Oncol.* **18**, 230–243 (2021).
 218. AN, B.-H., JK, L., J, D. & S, G. An Update on the Epidemiology, Molecular Characterization, Diagnosis, and Screening Strategies for Early-Onset Colorectal Cancer. *Gastroenterology* **160**, 1041–1049 (2021).
 219. Selvakumar, S. C. *et al.* CRISPR/Cas9 and next generation sequencing in the personalized treatment of Cancer. *Mol. Cancer* **21**, (2022).
 220. Kalyan, A. *et al.* Nodal Signaling as a Developmental Therapeutics Target in Oncology. *Mol. Cancer Ther.* **16**, 787–792 (2017).
 221. RL, E. & GC, B. Role of transforming growth factor Beta in human cancer. *J. Clin. Oncol.* **23**, 2078–2093 (2005).
 222. Gong, W. *et al.* Nodal signaling promotes vasculogenic mimicry formation in breast cancer via the Smad2/3 pathway. *Oncotarget* **7**, 70152–70167 (2016).
 223. Kirsammer, G. *et al.* Nodal signaling promotes a tumorigenic phenotype in human breast cancer. *Semin. Cancer Biol.* **29**, 40–50 (2014).
 224. Lyu, J. & Cheng, C. Regulation of Alternative Splicing during Epithelial-Mesenchymal Transition. *Cells Tissues Organs* **211**, 238–251 (2022).
 225. Köberle, B. & Schoch, S. Platinum Complexes in Colorectal Cancer and Other Solid Tumors. *Cancers (Basel)*. **13**, (2021).
 226. Kweekel, D. M., Gelderblom, H. & Guchelaar, H. J. Pharmacology of oxaliplatin and the use of pharmacogenomics to individualize therapy. *Cancer Treat. Rev.* **31**, 90–105 (2005).
 227. Zdraveski, Z. Z., Mello, J. A., Farinelli, C. K., Essigmann, J. M. & Marinus, M. G. MutS preferentially recognizes cisplatin- over oxaliplatin-modified DNA. *J. Biol. Chem.* **277**, 1255–1260 (2002).
 228. Venhuizen, J. H., Jacobs, F. J. C., Span, P. N. & Zegers, M. M. P120 and E-cadherin: Double-edged swords in tumor metastasis. *Semin. Cancer Biol.* **60**, 107–120 (2020).
 229. Duñach, M., Del Valle-Pérez, B. & García de Herreros, A. p120-catenin in canonical Wnt signaling. *Crit. Rev. Biochem. Mol. Biol.* **52**, 327–339 (2017).

230. Kurley, S. J. *et al.* A requirement for p120-catenin in the metastasis of invasive ductal breast cancer. *J. Cell Sci.* **134**, (2021).
231. Kidacki, M., Lehman, H. L., Green, M. V., Warrick, J. I. & Stairs, D. B. p120-catenin downregulation and PIK3CA mutations cooperate to induce invasion through MMP1 in HNSCC. *Mol. Cancer Res.* **15**, 1398–1409 (2017).
232. Zhong, C., Zuo, Z., Ji, Q. & Feng, D. P120ctn may participate in epithelial-mesenchymal transition in OSCC. *Indian J. Cancer* **53**, 20–24 (2016).
233. Jia, Y. *et al.* Lamin B1 loss promotes lung cancer development and metastasis by epigenetic derepression of RET. *J. Exp. Med.* **216**, 1377–1395 (2019).
234. Fu, F. *et al.* HMGA1 exacerbates tumor growth through regulating the cell cycle and accelerates migration/invasion via targeting miR-221/222 in cervical cancer article. *Cell Death Dis.* **9**, (2018).
235. Loesch, R. *et al.* Deleting the β -catenin degradation domain in mouse hepatocytes drives hepatocellular carcinoma or hepatoblastoma-like tumor growth. *J. Hepatol.* (2022) doi:10.1016/J.JHEP.2022.02.023.
236. Huang, X. D. *et al.* [Mouse models for human colorectal cancer with liver metastasis]. *Zhonghua Yi Xue Za Zhi* **99**, 2701–2705 (2019).
237. van Leerdam, M. E. *et al.* Endoscopic management of Lynch syndrome and of familial risk of colorectal cancer: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* **51**, 1082–1093 (2019).
238. Shen, N. *et al.* Association between rs2853669 in TERT gene and the risk and prognosis of human cancer: a systematic review and meta-analysis. *Oncotarget* **8**, 50864–50872 (2017).

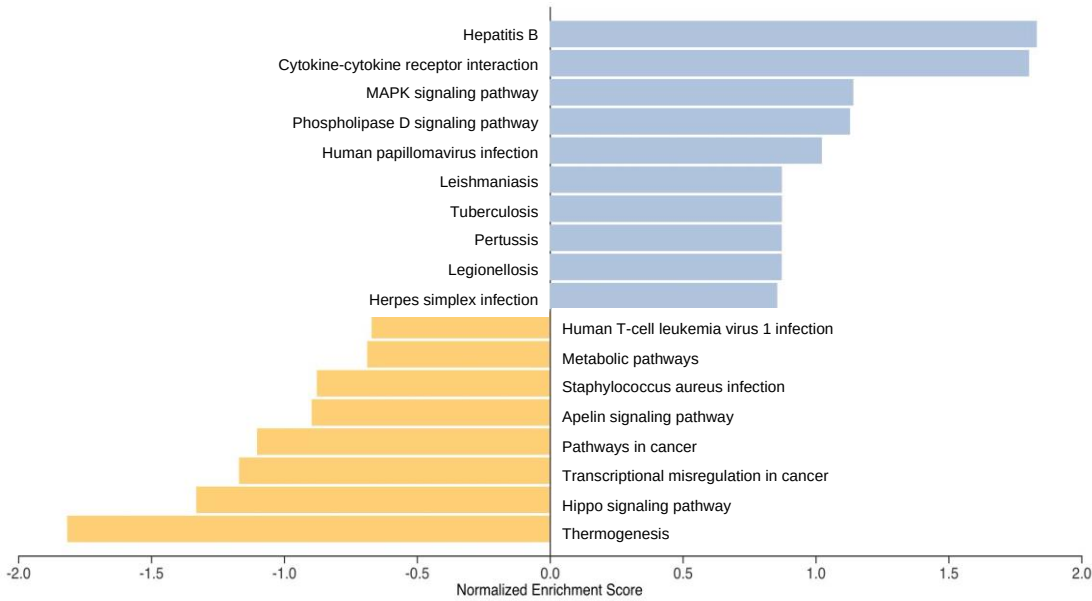
Anexos



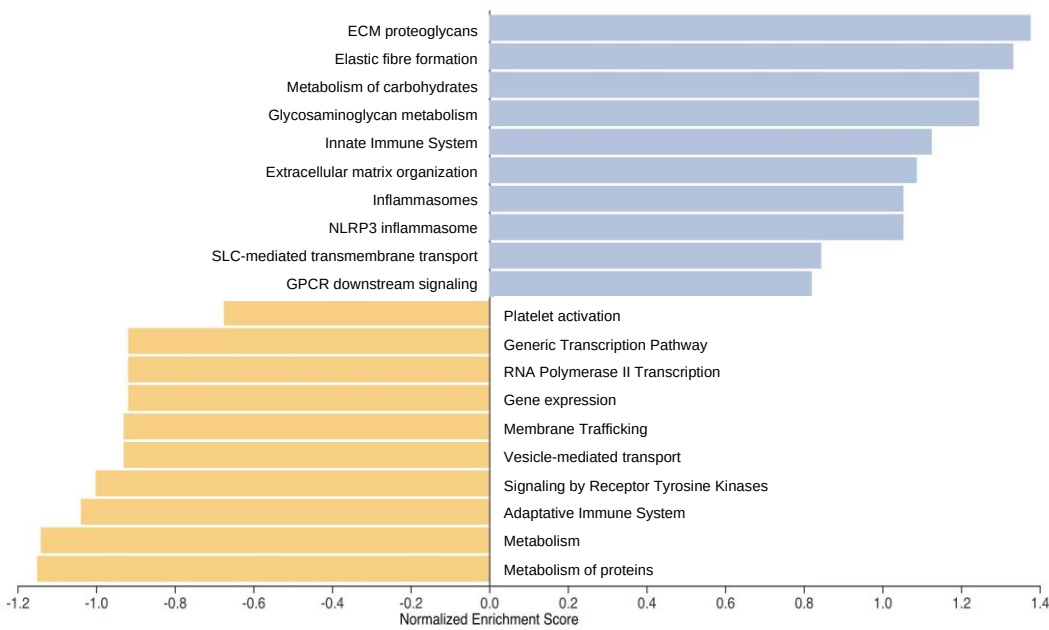
Anexos

Anexo 1. Rutas afectadas en las líneas celulares HCT-116 y HS-5 tras la inactivación de *NOMO1*. Set de genes identificados en el microarray de expresión comparados con las bases de datos “KEGG” y “Reactome”.

KEGG

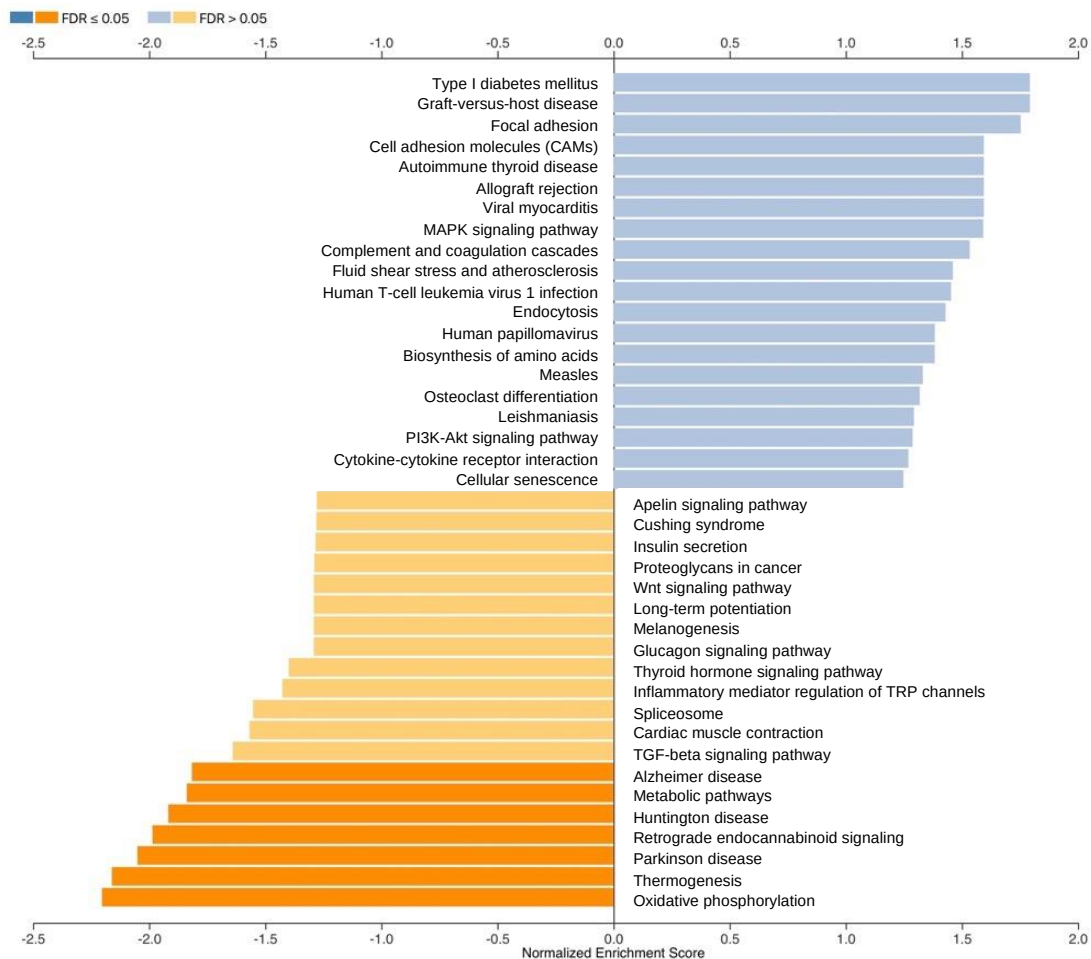


Reactome

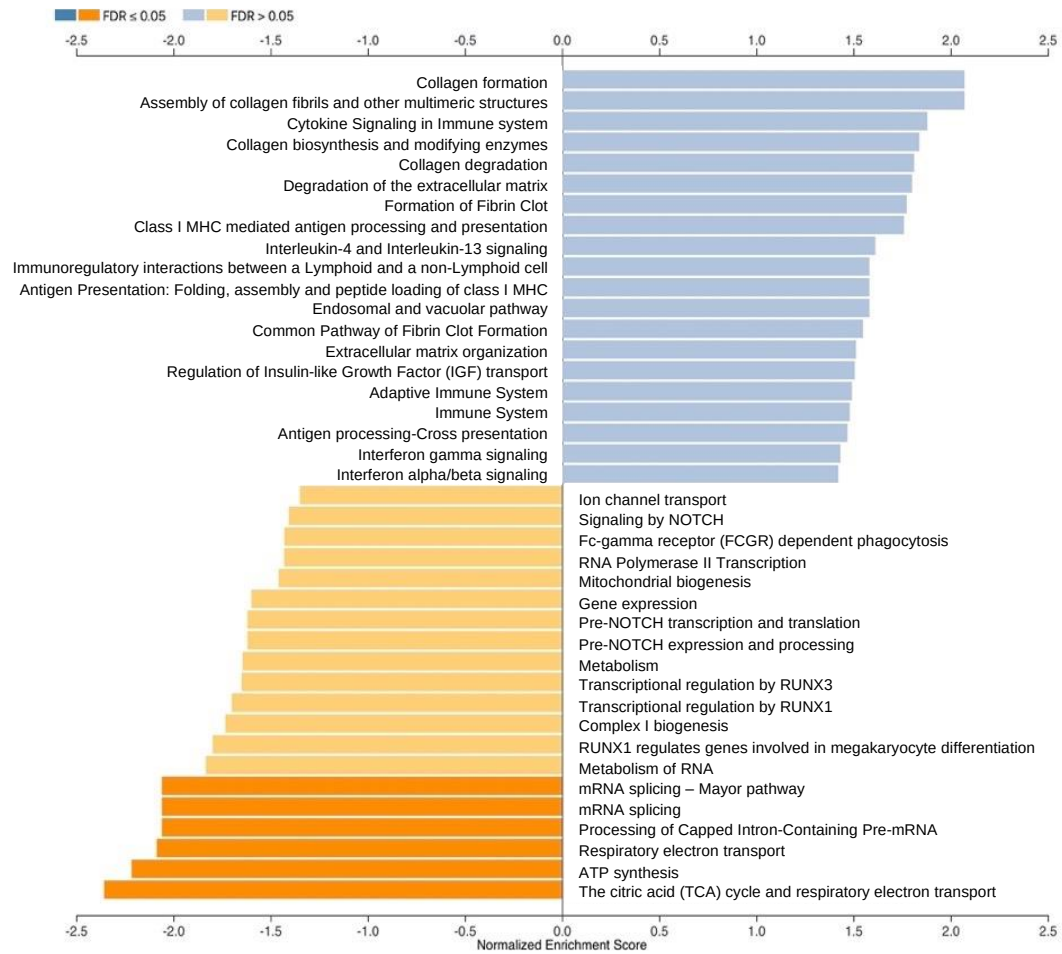


Anexo 2. Rutas afectadas en las líneas celulares HCT-116 y HS-5 tras la inactivación de *NOMO1*. Set de genes identificados en la secuenciación de RNA comparados con las bases de datos “KEGG” y “Reactome”.

KEGG

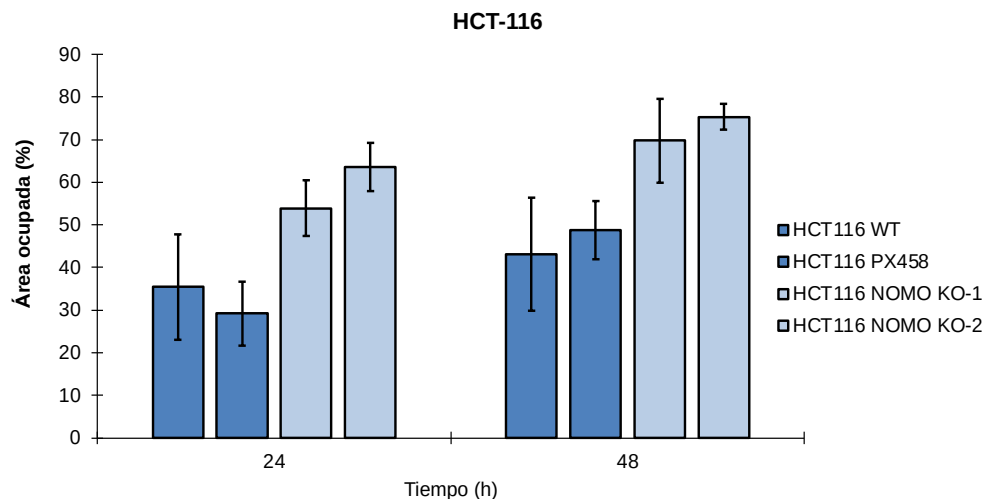


Reactome

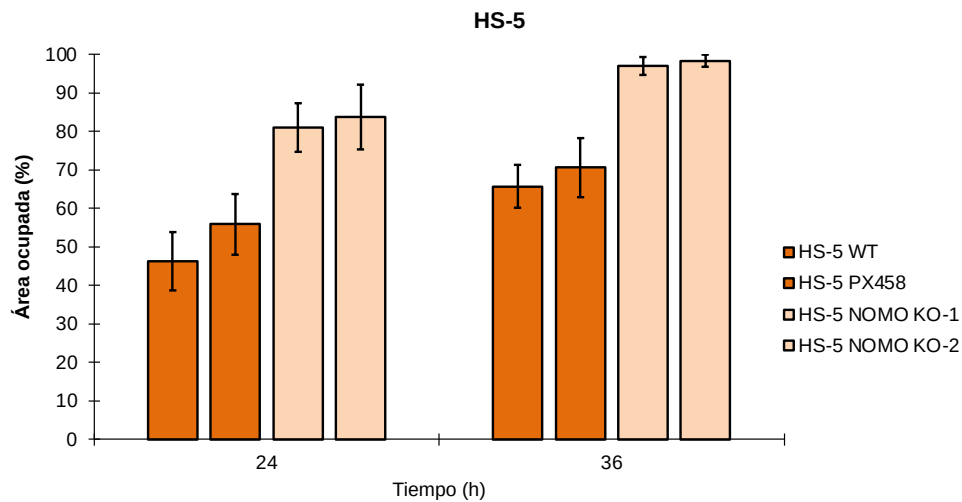


Anexo 3. Cuantificación del porcentaje de área ocupada de los clones NOMO-WT y NOMO-KO de cada línea celular. Se representan todas las réplicas a todos los tiempos estudiados.

HCT-116 WT					
Tiempo	0 horas	12 horas	24 horas	36 horas	48 horas
% ocupado WT.1	0,00	40,84	46,69	51,29	55,27
% ocupado WT.2	0,00	32,31	37,60	39,92	44,91
% ocupado WT.3	0,00	15,75	22,16	24,97	29,04
Promedio % ocupado	0,00	29,64	35,48	38,73	43,07
Desviación Estándar	0,00	12,76	12,40	13,20	13,21
HCT-116 PX458					
Tiempo	0 horas	12 horas	24 horas	36 horas	48 horas
% ocupado PX458.1	0,00	16,30	22,53	34,44	45,97
% ocupado PX458.2	0,00	28,68	37,55	47,94	56,60
% ocupado PX458.3	0,00	19,59	27,58	35,58	43,72
Promedio % ocupado	0,00	21,52	29,22	39,32	48,76
Desviación Estándar	0,00	6,34	7,50	7,47	6,85
HCT-116 NOMO KO-1					
Tiempo	0 horas	12 horas	24 horas	36 horas	48 horas
% ocupado NOMO KO-1.1	0,00	34,61	47,56	56,80	61,03
% ocupado NOMO KO-1.2	0,00	39,84	53,48	62,21	67,92
% ocupado NOMO KO-1.3	0,00	42,83	60,70	71,77	80,32
Promedio % ocupado	0,00	39,09	53,91	63,59	69,76
Desviación Estándar	0,00	4,16	6,58	7,58	9,78
HCT-116 NOMO KO-2					
Tiempo	0 horas	12 horas	24 horas	36 horas	48 horas
% ocupado NOMO KO-2.1	0,00	51,99	64,86	70,73	77,68
% ocupado NOMO KO-2.2	0,00	42,65	57,46	61,97	71,89
% ocupado NOMO KO-2.3	0,00	57,39	68,44	72,97	76,57
Promedio % ocupado	0,00	50,68	63,59	68,56	75,38
Desviación Estándar	0,00	7,46	5,60	5,81	3,07

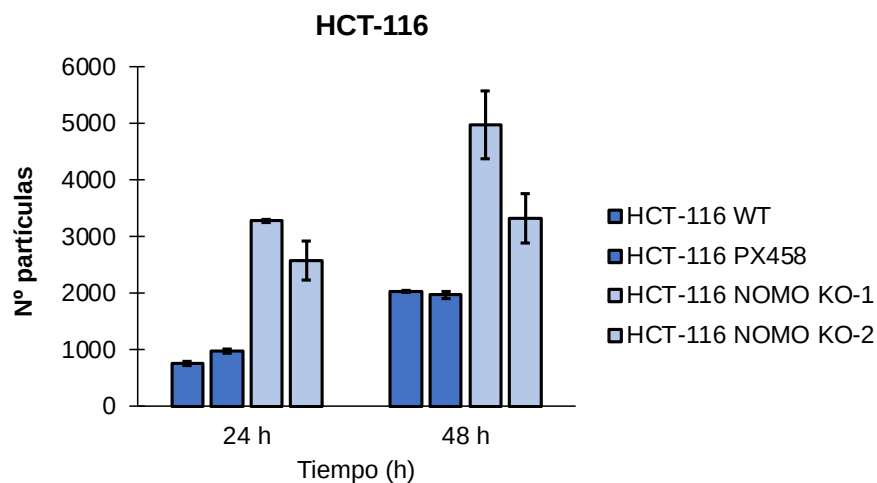


HS-5 WT				
Tiempo	0 horas	12 horas	24 horas	36 horas
% ocupado WT.1	0,00	24,00	43,55	63,85
% ocupado WT.2	0,00	21,50	57,46	67,26
% ocupado WT.3	0,00	13,37	40,28	59,29
Promedio % ocupado	0,00	19,62	47,10	63,47
Desviación Estándar	0,00	5,56	9,12	4,00
HS-5 PX458				
Tiempo	0 horas	12 horas	24 horas	36 horas
% ocupado PX458.1	0,00	28,82	56,88	76,58
% ocupado PX458.2	0,00	23,52	47,57	61,97
% ocupado PX458.3	0,00	27,96	63,27	73,30
Promedio % ocupado	0,00	26,77	55,91	70,62
Desviación Estándar	0,00	2,85	7,89	7,67
HS-5 NOMO KO-1				
Tiempo	0 horas	12 horas	24 horas	36 horas
% ocupado NOMO KO-2.1	0,00	39,96	86,16	98,29
% ocupado NOMO KO-2.2	0,00	34,92	85,39	98,00
% ocupado NOMO KO-2.3	0,00	36,17	72,23	98,12
Promedio % ocupado	0,00	37,02	81,26	98,14
Desviación Estándar	0,00	2,63	7,83	0,15
HS-5 NOMO KO-2				
Tiempo	0 horas	12 horas	24 horas	36 horas
% ocupado NOMO KO-1.1	0,00	37,44	82,56	98,71
% ocupado NOMO KO-1.2	0,00	41,00	82,78	96,10
% ocupado NOMO KO-1.3	0,00	42,70	95,04	100,00
Promedio % ocupado	0,00	40,38	86,79	98,27
Desviación Estándar	0,00	2,69	7,14	1,98

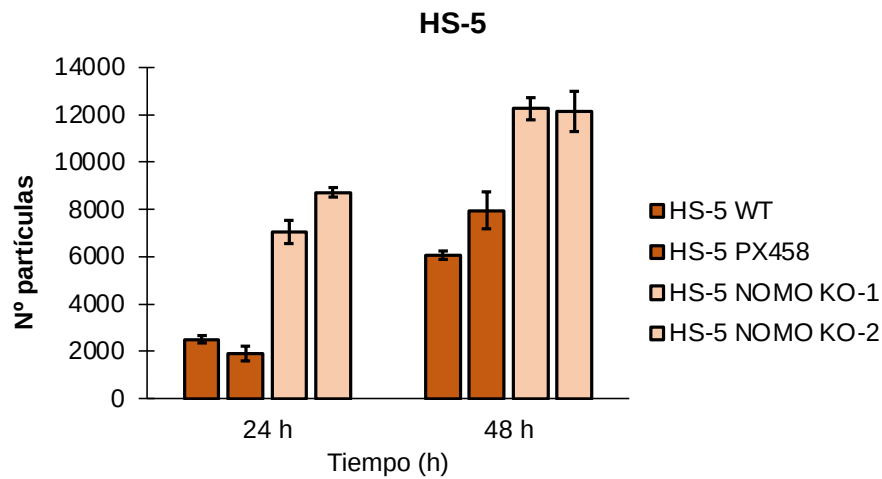


Anexo 4. Cuantificación del número de partículas positivas (células que han migrado) de los clones NOMO-WT y NOMO-KO de cada línea celular. Se representan todas las réplicas a todos los tiempos estudiados.

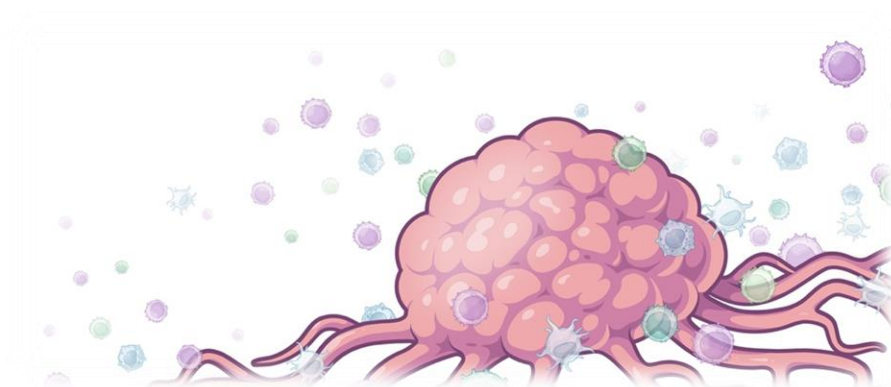
HCT-116 WT			
Tiempo	0 horas	24 horas	48 horas
Partículas + WT.1	0,00	729,00	2052,00
Partículas + WT.2	0,00	798,00	2033,00
Partículas + WT.3	0,00	777,00	2015,00
Promedio Partículas +	0,00	768,00	2033,33
Desviación Estándar	0,00	35,37	18,50
HCT-116 PX458			
Tiempo	0 horas	24 horas	48 horas
Partículas + PX458.1	0,00	947,00	2044,00
Partículas + PX458.2	0,00	1009,00	1921,00
Partículas + PX458.3	0,00	985,00	1951,00
Promedio Partículas +	0,00	980,33	1972,00
Desviación Estándar	0,00	31,26	64,13
HCT-116 NOMO KO-1			
Tiempo	0 horas	24 horas	48 horas
Partículas + NOMO KO-2.1	0,00	3281,00	4287,00
Partículas + NOMO KO-2.2	0,00	3305,00	5330,00
Partículas + NOMO KO-2.3	0,00	3243,00	5299,00
Promedio Partículas +	0,00	3276,33	4972,00
Desviación Estándar	0,00	31,26	593,43
HCT-116 NOMO KO-2			
Tiempo	0 horas	24 horas	48 horas
Partículas + NOMO KO-1.1	0,00	2620,00	3301,00
Partículas + NOMO KO-1.2	0,00	2223,00	2904,00
Partículas + NOMO KO-1.3	0,00	2910,00	3761,00
Promedio Partículas +	0,00	2584,33	3322,00
Desviación Estándar	0,00	344,89	428,89



HS-5 WT			
Tiempo	0 horas	24 horas	48 horas
Partículas + WT.1	0,00	2319,00	6060,00
Partículas + WT.2	0,00	2645,00	6221,00
Partículas + WT.3	0,00	2512,00	5895,00
Promedio Partículas +	0,00	2492,00	6058,67
Desviación Estándar	0,00	163,92	163,00
HS-5 PX458			
Tiempo	0 horas	24 horas	48 horas
Partículas + PX458.1	0,00	2219,00	8383,00
Partículas + PX458.2	0,00	1615,00	7057,00
Partículas + PX458.3	0,00	1879,00	8412,00
Promedio Partículas +	0,00	1904,33	7950,67
Desviación Estándar	0,00	302,80	774,07
HS-5 NOMO KO-1			
Tiempo	0 horas	24 horas	48 horas
Partículas + NOMO KO-2.1	0,00	6553,00	11759,00
Partículas + NOMO KO-2.2	0,00	7555,00	12326,00
Partículas + NOMO KO-2.3	0,00	6991,00	12671,00
Promedio Partículas +	0,00	7033,00	12252,00
Desviación Estándar	0,00	502,32	460,48
HS-5 NOMO KO-2			
Tiempo	0 horas	24 horas	48 horas
Partículas + NOMO KO-1.1	0,00	8720,00	11276,00
Partículas + NOMO KO-1.2	0,00	8904,00	12216,00
Partículas + NOMO KO-1.3	0,00	8501,00	12951,00
Promedio Partículas +	0,00	8708,33	12147,67
Desviación Estándar	0,00	201,75	839,59



Artículos



Artículo 1

Title: RECURRENT *NOMO1* GENE DELETION IS A POTENTIAL CLINICAL MARKER IN EARLY-ONSET COLORECTAL CANCER

Authors: Jérica Pérez García^{1,2,3, †}, Abel Martel-Martel^{1,2,4, †}, Paula García-Vallés^{1,2}, Luis Corchete^{3,5,6}, Juan L. García^{1,3}, Nerea Gestoso-Uzal^{1,2}, Rosario Vidal Tocino^{1,4}, Oscar Blanco^{1,7}, Lucía Méndez^{1,8}, Manuel Sánchez-Martín^{1,8}, Manuel Fuentes^{9,10}, Ana B. Herrero^{1,2,3}, Andreana N. Holowatyj¹¹, José Perea¹, Rogelio González-Sarmiento^{1,2,3}.

Affiliations: ¹Institute of Biomedical Research of Salamanca (IBSAL), SACYL-University of Salamanca-CSIC, Salamanca, Spain. ²Molecular Medicine Unit, Department of Medicine, University of Salamanca, Salamanca, Spain. ³Institute of Molecular and Cellular Biology of Cancer (IBMCC), University of Salamanca-CSIC, Salamanca, Spain. ⁴Medical Oncology Department, Complejo Asistencial Universitario de Salamanca-IBSAL, Salamanca, Spain. ⁵Hematology Department, University Hospital of Salamanca, Institute of Biomedical Research of Salamanca (IBSAL), Salamanca, Spain. ⁶Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. ⁷Anatomy Pathology Service, University Hospital of Salamanca, Salamanca, Spain. ⁸Transgenic Service, Nucleus, University of Salamanca, Salamanca, Spain. ⁹Department of Medicine and Cytometry General Service-Nucleus, CIBERONC, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), Salamanca, Spain. ¹⁰Proteomics Unit, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), Salamanca, Spain. ¹¹Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA.

† Jérica Pérez García and Abel Martel-Martel contributed equally to this work.

Estado: sometido a publicación.

ABSTRACT

Early-onset colorectal cancer (EOCRC; age younger than 50 years) incidence has been steadily increasing over the last decades worldwide with causes unexplained. A unique molecular feature of EOCRC is that these cases harbor a greater incidence of *Nodal Modulator 1 (NOMO1)* somatic deletions compared with late-onset colorectal cancer. Yet the mechanisms of *NOMO1* in early-onset colorectal carcinogenesis are currently unknown. Here we show that 50% of heterozygous *NOMO1* deleted–EOCRCs presented pathogenic mutations in this gene, suggesting that *NOMO1* can be inactivated by deletion or mutation in EOCRC. To study its role in EOCRC, CRISPR/cas9 technology was used to generate *NOMO1* knockout HCT-116 (EOCRC) and HS-5 (bone marrow) cell lines. Loss of *NOMO1* in these cell lines did not perturb Nodal pathway signaling. Results from expression microarray, RNA sequencing and protein expression analysis by LC-IMS/MS revealed that *NOMO1* inactivation deregulates other signaling pathways independent of the Nodal pathway such as epithelial-mesenchymal transition and cell migration. Importantly, *NOMO1* inactivation increased the migration capacity of CRC cells. Additionally, a gut-specific conditional KO mouse model of *NOMO1* revealed no subsequent tumor development in mice. Overall, these findings suggest that *NOMO1* could not be a driver of early-onset colorectal carcinogenesis, but its loss promotes an increased migration capacity of CRC cells. Further study is warranted to explore other signaling pathways deregulated by the loss of *NOMO1* that may play a relevant role in the pathogenesis of the disease.

INTRODUCTION

Colorectal cancer (CRC) is the third most commonly diagnosed cancer among men and women worldwide, estimated to have accounted for one in every ten new cancer cases in 2020(1). Despite a reduction in the absolute numbers of patients diagnosed with CRC, there has been an increase in CRC incidence among individuals diagnosed before age 50 years (Early-Onset CRC; EOCRC) that remains not well understood(2). Indeed, the pathogenesis of EOCRC is well characterized among individuals with hereditary CRC. However, the majority (over 80%) of EOCRC cases do not carry a germline mutation associated with cancer predisposition (sporadic EOCRC)(3,4). Consequently, there is a timely need to elucidate the molecular etiologies of sporadic EOCRC to accelerate the translation of research findings into clinical application and reduce this disease burden.

Evidence is accumulating to suggest that EOCRC carries a distinct clinical, pathologic, and molecular presentation compared with CRCs from cases diagnosed at age 50+ years (Late-Onset CRC; LOCRC)(5–9). For example, from a clinical point of view, sporadic early-onset tumors have a worse prognosis than the late-onset ones: they are more aggressive, develop early metastasis and therefore they are associated with poorer survival. From a molecular point of view, EOCRCs show poor differentiation, signet ring and mucinous histology, typical features of tumors associated with Lynch syndrome. Moreover, these tumors show substantial dissimilarities regarding the CIN pattern in comparison with LOCRC(10,11).

Assessment of DNA copy number alterations (CNAs) between EOCRC and LOCRC cases has also yielded intriguing observations for further study. We previously identified a recurrent deletion in the short arm of chromosome 16 (16p13.12-p13.11) that presented alone or in combination with other gains or losses of genetic material to be more frequent in EOCRC than LOCRC cases (33% vs. 16.3%, respectively)(12). Importantly, in this chromosomal region was located the Nodal Modulator 1 (*NOMO1*) gene—which was subsequently found to be deleted in 70% of EOCRC cases. However, among late-onset CRCs, only 4.5% of cases carried a *NOMO1* deletion. Together, these findings suggest that *NOMO1* may be a promising molecular feature distinct to EOCRC cases for further molecular study(13).

The *NOMO1* gene encodes a 130kDa transmembrane protein located in the endoplasmic reticulum that forms a protein complex together with Nicalin (NCLN) and the Transmembrane Protein 147 (TMEM147) that inhibits Nodal signaling during embryonic development(14,15). Importantly, the Nodal pathway is a signal transduction pathway critical for differentiation during embryonic development and for the maintenance of pluripotency in human embryonic stem cells(16–19). The ligand, Nodal, propagates its signal by binding the type I Activin receptor ALK4/7 (ACVR1B/ACVR1C) and type II ACTRIIA/ACTRIIB (ACVR2A/ACVR2B) in cooperation with its co-receptor Cripto-1(16,17,20). This co-receptor binds ALK4 and promotes the recruitment of ACTRIIA/B, facilitating the formation of the receptor signaling complex. Once activated, the receptor phosphorylates Smad proteins (P-Smads) which bind to a common intermediate, SMAD4 (Co-Smad), allowing the translocation of the complex to the nucleus where it regulates the transcription of target genes(20). Although Nodal is not active in most adult tissues, its re-expression and signaling have been linked to multiple types of human cancers(21). Nodal signaling maintains the self-renewal capacity of cancer stem cells (CSC) and promotes the invasiveness of several solid tumor-including melanoma, breast, colon, ovarian, prostate, endometrial and pancreas tumours(18,21–23). In addition, other complex components, including co-receptor Cripto-1, have been found to be overexpressed in a variety of human tumors and human cancer cell lines, whereas low or undetectable levels of expression were detected in normal adult tissues and in non-transformed normal cell lines(16). Despite the growing importance of Nodal signaling in carcinogenesis, the role of *NOMO1* in the pathogenesis or progression of CRC, particularly EOCRC, remains unexplored.

In the study presented here, we used CRISPR-Cas9 technology to delete endogenous *NOMO1* in multiple independent cell lines bearing the wild type (WT) gene. We also used a gut-specific conditional knockout (KO) mouse model of *NOMO1* to study subsequent tumor development. Specifically, the characterization of the *NOMO1*-KO clones revealed that loss of *NOMO1* did not affect Nodal signaling pathway activity nor cell proliferation. Importantly, loss of *NOMO1* deregulate the epithelial-mesenchymal transition (EMT) process and increased CRC cell migration. Deletion of *NOMO1* using in vivo models also yielded no subsequent tumor development. Together, these findings suggest that *NOMO1* could not be a driver of early-onset colorectal carcinogenesis, and that other signaling pathways deregulated by its loss may play a relevant role in the pathogenesis of EOCRC.

RESULTS

Quantitative PCR reveals a single *NOMO* gene

The presence of three highly similar genes named *NOMO1*, *NOMO-2* and *NOMO-3* have been previously reported in a region of duplication located on the p arm of chromosome 16(35). These genes encode closely related proteins with a homology of 99.6% that may have identical function. To confirm the existence of three *NOMO* genes, we used the RT-qPCR method(36). A specific pair of oligonucleotides were designed to amplify the *NOMO* genes. Using these primers, a DNA fragment that contains part of exon 4 and part of intron 4 (whose sequence is 100% similar in the three reported *NOMO* genes) was amplified. Several oligonucleotides were also designed to amplify and quantify different controls (Table S1). To further quantify *NOMO* content, we amplified another fragment of the *NOMO* gene that contains part of exon 1-a sequence which is 100% identical across the three reported *NOMO* genes. As an internal control, we amplified part of exon 1 of the single copy gene *LEMD3* (two alleles), a DNA fragment between exon2-intron 2 of *RBFOX1*, a gene located in proximity to the *NOMO* genes on the same chromosomal region (two alleles), and a fragment of exon 30 of *PKD1*-that has no homology with any pseudogene (two alleles). We also amplified, as a control of gene duplication, exon 13 of *PKD1*, that has a pseudogene that contains a 100% homologous sequence (4 alleles), and exon 5 of the *STS* gene, located on chromosome X that has only one allele in males. Our results showed that the amount of *NOMO* DNA amplified in RT-qPCR was similar to single copy genes *LEMD3*, *RBFOX1* and exon 30 of *PKD1*, whereas this amount was half the amount of exon 13 of *PKD1* and double the amount of *STS* exon 5 (Fig. 1).

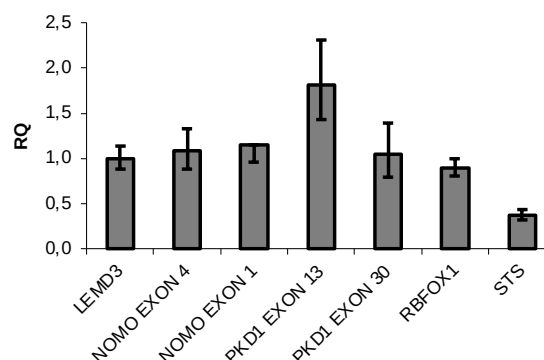


Fig. 1. DNA content of the indicated genes determined by q-PCR. Data are presented as the mean of three replicates $\pm$ SD. RQ represents the relative expression levels of each amplicon normalized with a control (*LEMD3*).

Next, we compared coding regions of the three reported *NOMO* genes and found that they differ in 20 nucleotides that correspond to 13 silent and 7 missense mutations (Table 1). Only one of these missense mutations (c.26C>T/p.L9P) has not been included in the dSNP database(37). The allele frequency of these mutations obtained from the ExAC aggregated population(38) showed that most of the *NOMO* WT alleles have a population frequency near 100%. As the presence of two genes would yield a 50% frequency of the mutated allele, these results accumulate to strongly suggest that there is a single *NOMO* gene.

Table 1. Allele frequency of nucleotides that differ between the reported *NOMO1*, *NOMO2* and *NOMO3*. Missense mutations are marked in bold.

GVSc	<i>NOMO1</i>	<i>NOMO2</i>	<i>NOMO3</i>	POBLATION ID	SUBJECTS	FREQ. ALE>	FREQ. ALE<
c.26A>T p.(L9P)	C	T	C				
c.156T>G p.(S52S)	T	G	T				
c.696C>G p.(N232N)	C	G	C	ExAc	7964	C=0.997	G=0.003
c.1185G>A p.(P395P)	G	G	A	ExAc	121402	G=0.999	A=0.001
c.1210 A>G p.(I404V)	A	G	G	ExAc	121400	A=0.999	G=0.001
c.1238 A>G p.(Q413R)	A	G	A	ExAc	11862	A=0.999	G=0.000
c.1260C>G p.(P420P)	C	G	C	ExAc	25306	C=0.999	G=0.001
c.1374 T>G p.(K458N)	T	T	G	ExAc	27102	T=0.777	G=0.223
c.1468A>G p.(N490D)	A	G	G	ExAc	23336	A=0.943	G=0.056
c.1477 A>G p.(M493V)	A	G	G	ExAc	121206	A=0.999	G=0.001
c.1738 A>G p.(M580V)	A	G	G	ExAc	119700	A=0.989	G=0.010
c.2187C>T p.(G729G)	C	T	T	ExAc	121396	C=0.999	T=0.001
c.2211C>T p.(P737P)	C	C	T	ExAc	814	T=0.814	C=0.185
c.2388T>C p.(H796H)	T	C	C				
c.2586G>A p.(A862A)	G	A	A				
c.2694C>T p.(S898S)	C	T	T	ExAc	121402	C=0.999	T=0.001
c.3216A>G p.(T1072T)	A	G	G	ExAc	121310	A=0.999	G=0.001
c.3318C>T p.(D1106D)	C	T	T	ExAc	121406	C=0.999	T=0.001
c.3583C>G p.(R1195R)	C	G	G				
c.3666T>G p.(T1222T)	T	G	G				

***NOMO1* is frequently inactivated by deletion or mutation in EOCRC**

We have previously reported that a high proportion of EOCRCs carry a homozygous deletion of the *NOMO1* gene(13). To expand upon these findings, we sequenced the *NOMO1* gene in 20 EOCRCs bearing a heterozygous deletion of *NOMO1*. Mutational profiling obtained through next-generation sequencing showed that half (8 of 16) carried a pathogenic mutation in the *NOMO1* gene (Table 2). The majority of these mutations were nonsense mutations (7/8; 87.5%) with a single frameshift mutation (1/8; 12.5%). In addition, 8 VUS were classified by prediction programs as likely pathogenic. Therefore, more than half of the tumors with heterozygous deletion of the *NOMO1* gene harbored a pathogenic mutation in the remaining allele or a VUS described as likely pathogenic (Table 2). Together, these results indicate that *NOMO1* is frequently inactivated in EOCRC either through homozygous deletion or a deletion of one allele and a pathogenic mutation in the second allele.

Table 2. Pathogenic mutations and variants of unknown clinical significance (VUS) identified in CRC tumors with heterozygous deletion of the *NOMO1* gene.

	VARIANT	ALLELIC FREQUENCY (%)	SIFT	POLYPHEN	CADD SCORE
Sample 1	c.3587C>T/p.(Ala1196Val)	40.8	deleterious (0.01)	probably damaging (0.931)	24.1
Sample 2	c.3651_3652delGA/p.(Lys1218AspfsTer31)	46.9	-	-	
Sample 4	c.1692G>A/p.(Trp564Ter)	12.5	-	-	
Sample 4	c.3107T>A/p.(Ile1036Asn)	26.7	deleterious (0)	possibly damaging (0.838)	23.2
Sample 4	c.3284C>G/p.(Ser1095Cys)	16.3	deleterious (0)	probably damaging (0.979)	17.2
Sample 5	c.2684T>A/p.(Leu895Ter)	22.5	-	-	
Sample 5	c.2810C>T/p.(Pro937Leu)	52.9	deleterious (0)	probably damaging (1)	29.5
Sample 6	c.3637C>T/p.(Gln1213Ter)	27.3	-	-	
Sample 7	c.2434G>T / p.(Glu812Ter)	18.7	-	-	
Sample 8	c.3649C>T / p.(Gln1217Ter)	20.4	-	-	
Sample 9	c.118G>T / p.(Gly40Cys)	15.2	deleterious (0)	probably damaging (1)	23.1
Sample 12	c.3002G>A / p.(Gly1001Asp)	38.3	deleterious (0.01)	probably damaging (0.998)	25.1
Sample 14	c.2278G>A / p.(Gly760Arg)	11.5	deleterious (0.01)	probably damaging (0.992)	29.3
Sample 14	c.2833C>T/p.(Gln945Ter)	10.5	-	-	
Sample 14	c.3019G>A / p.(Gly1007Arg)	25.0	deleterious (0)	probably damaging (1)	26.0
Sample 15	c.2428G>T/p.(Glu810Ter)	17.2	-	-	

CRISPR/Cas9 technology efficiently produces inactivation of *NOMO1*

To investigate the consequences of *NOMO1* inactivation, we constructed *NOMO1*-KO cell lines using CRISPR-Cas9 technology. We deleted endogenous *NOMO1* in one cell line derived from EOCRC (HCT-116) and in a bone marrow-derived non-cancerous mesenchymal cell line (HS-5). Each cell line was transfected with a combination of two plasmids (Fig. S1). In addition, cell lines were also transfected with the PX458 empty plasmid. After single-cell sorting of GFP+ cells, clones were expanded and screened by RT-qPCR (Fig. 2A), PCR-Sanger sequencing (Fig. 2B) and WB (Fig. 2C). At least two *NOMO1*-KO clones and two control WT clones for each cell line were efficiently generated and selected for further studies.

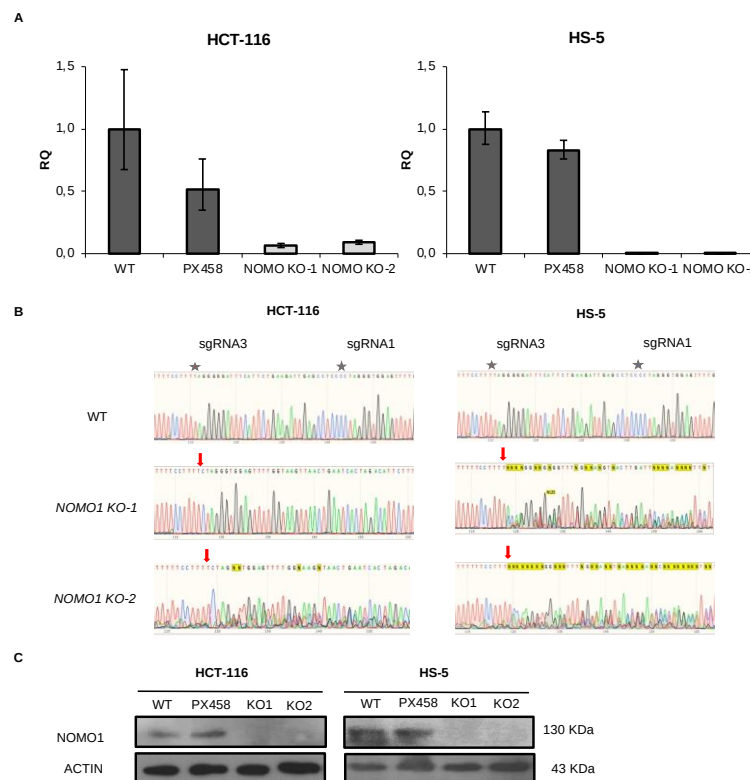


Fig. 2. Generation of *NOMO1*-KO cell lines using CRISPR/Cas9 technology. (A) qPCR amplification of the fragment corresponding to the sgRNAs used in *NOMO1*-KO and control clones. The RQ was calculated according to $2^{-\Delta\Delta CT}$ method, using the WT clones for normalization. Data are shown as the mean $\pm$ SD of

three replicates. (B) Confirmation by Sanger sequencing of the reading frame change in the nucleotide sequence generated by Cas9 in *NOMO1*-KO clones. The Cas9 breakpoint guided by sRNA1/sgRNA3 is marked with a red arrow. (C) *NOMO1* expression by WB in WT and KO clones in HCT-116 and HS-5 cell lines.

Inactivation of *NOMO1* significantly reduces the expression of *NCLN* in cell lines

It has been reported that *NOMO1*, together with *NCLN* and *TMEM147* form a protein complex located in the endoplasmic reticulum(14,39). Using an RNA interference approach, previous studies have demonstrated that *NCLN* and *NOMO1* become unstable and therefore decrease in the absence of the respective binding partner, suggesting that complex formation has a stabilizing effect(39). Similarly, it was later shown that *TMEM147* expression was strongly reduced in *NCLN* and *NOMO1* knockdown cells(14). Here, we found that the expression level of *TMEM147* was not affected by the loss of *NOMO1* in our study (Fig. 3A). In contrast, *NCLN* expression was strongly reduced in *NOMO1*-KO cells across all two cell lines tested. As the reduction in *NCLN* levels was not accompanied by a significant decline in the corresponding mRNA expression levels as determined by qRT-PCR, these results suggest that protein reduction is caused by a posttranscriptional mechanism (Fig. 3B).

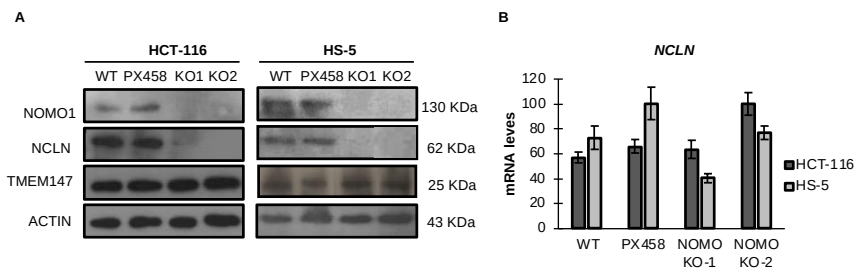


Fig. 3. Expression levels of Nicalin and TMEM147 in the presence or absence of *NOMO1* across cell lines. (A) *NOMO1*, *NCLN* and *TMEM147* protein levels determined by WB. (B) mRNA expression of Nicalin determined by qRT-PCR in *NOMO1*-KO and control clones. The expression of each clone was determined using the $2^{-\Delta\Delta CT}$ method and GAPDH was used for normalization. Data are shown as the mean $\pm$ SD of three replicates.

Inactivation of *NOMO1* does not affect Nodal signaling pathway activity nor cell proliferation

The *NCLN*-*NOMO1*-*TMEM147* complex has been shown to modulate Nodal signaling in developing zebrafish embryos by an unknown mechanism(14,15,39). Therefore, we next analyzed the abundance of several members of the Nodal pathway in *NOMO1*-KO and control cells. First, we quantified protein expression of receptor proteins ALK4, ACTRII, and co-receptor Cripto-1, by WB(18,23). We observed that no alterations in the abundance of any of these proteins were observed after *NOMO1*-KO (Fig. 4A). Once activated by the Nodal ligand, the receptors phosphorylate the Smad2 and Smad3 proteins that bind to Smad4 (Co-Smad). We found that Smad4 expression was not affected by *NOMO1* inactivation in any of the cell lines analyzed (Fig. 4B). To further examine whether the loss of *NOMO1* affects the activity of the Nodal signaling pathway, we analyzed Smad2/3 and p-Smad2/3 levels in the presence or absence of recombinant human Nodal (rhNodal) (R&D Systems). Smad2/3 and p-Smad2/3 proteins expression was found to be similar in WT and *NOMO1*-KO clones (Fig. 4C). This strongly indicates that Nodal pathway activity is not affected by *NOMO1* inactivation.

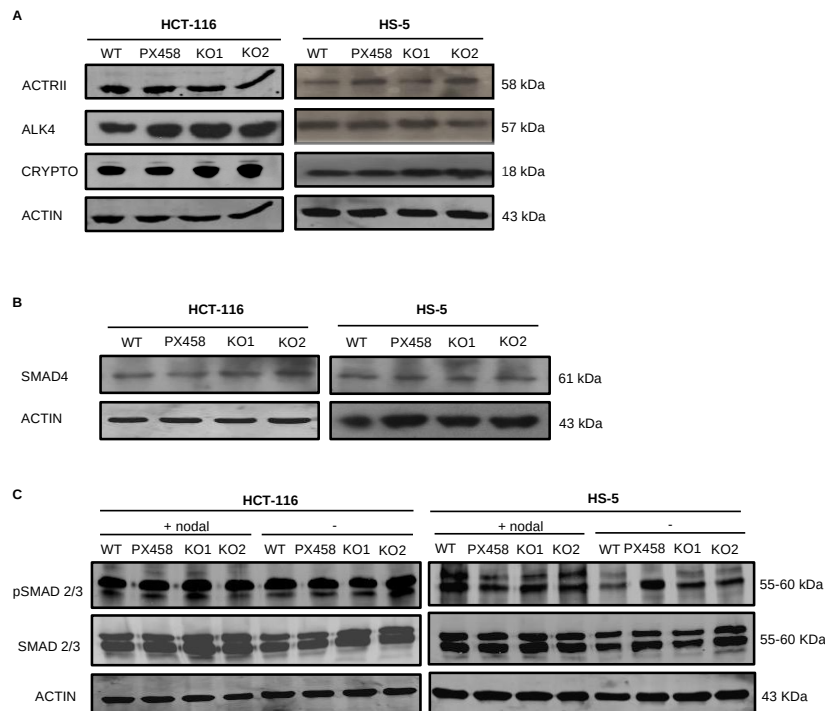


Fig. 4. Expression levels of proteins involved in the Nodal signaling pathway in *NOMO1*-KO and control clones. (A) Expression of ALK4 and ACTRII Nodal pathway receptor proteins and the co-receptor Cripto1 detected by WB. (B) Protein expression of SMAD4. (C) Levels of Smad2/3 and p-Smad2/3 protein expression in clones treated or untreated with rhNodal (300ng/mL) for 24h.

To determine whether loss of *NOMO1* affects cell growth rates, WT and *NOMO1*-KO clones were cultured and proliferation was quantified by MTT assays. No significant differences ($p > 0.05$) were found between the growth rates of the *NOMO1*-KO and control clones (Fig. S2).

Gene expression profiling in *NOMO1*-KO cell lines

Although our results indicate that *NOMO1* inactivation does not affect Nodal pathway activity, we aimed to test whether loss of *NOMO1* affected gene expression profiles. Total RNAs were extracted from *NOMO1*-KO and WT clones and subsequently processed for expression microarray and RNA sequencing (RNA-seq). For both analyses, a FC greater than 1.5 was considered for up-regulated genes and a FC lower than -1.5 for down-regulated genes.

According to the microarray analysis revealed a total of 63 genes were found to be de-regulated commonly to the two cell lines (HCT-116 and HS-5) when compared the 5 *NOMO1*-KO clones to 5 *NOMO1*-WT clones ($FC > 1.5$; $FC < -1.5$). A total of 37 genes were up-regulated and 26 down-regulated in *NOMO1*-KO clones compared to WT (Fig. S3A; Table S4). Results from the Gene Set Enrichment Analysis (GSEA) revealed the existence of 10 pathways to be enriched by the deregulated genes associated with *NOMO1* loss (Fig. S4A).

According to the RNA-seq data, a total of 592 genes were found to be de-regulated ($FDR < 0.05$) commonly to the two cell lines when compared the 6 *NOMO1*-KO clones to 6 control clones. A total of 108 genes were significantly up-regulated and 213 down-regulated in *NOMO1*-KO clones compared to WT (Fig. S3B; Table S5). The GSEA analysis showed that *NOMO1* inactivation leads to the enrichment of 20 signaling pathways (Fig. S4B).

Interestingly, microarray and RNA-seq commonly deregulated genes in *NOMO1*-KO cells showed to perturb the EMT process by the deregulation of *QSOX1*, *LUM*, *PCOLCE*, *CAPG*, *POSTN*, *CCN1*, *CCN2*, *CD44*, *SERPINE2*, *THBS1* and *VEGFA*. These genes are also involved in the alteration of cell migration(40–49). In addition, according to an over-representation analysis (ORA) that included the genes mentioned above, different biological processes were affected in relation to cell migration (Fig. S4C). Therefore, the alteration of these two processes could contribute to an increase in the metastatic capacity of *NOMO1*-KO cells.

For that reason, to identify whether *NOMO1* inactivation affected these processes, we first studied the protein expression of the typical EMT-associated markers E-Cadherin and Vimentin, and the expression of β -catenin due to its relationship with cell migration, invasion and metastasis. However, protein expression of these markers did not show significant changes when comparing *NOMO1*-KO and control clones (Fig. S4D).

Taken together, microarray and RNA-seq approaches showed a differential expression profile and a set of affected signaling pathways by *NOMO1* loss, specially EMT and cell migration.

Protein expression profiling in *NOMO1*-KO cell lines

To elucidate the protein expression profile changes after *NOMO1* inactivation in the two cell lines (HCT-116 and HS-5), a LC-IMS/MS analyze was performed. A total of 4 samples of a *NOMO1*-WT and *NOMO1*-KO clones were compared for the two cell lines using 0.5 μ g of total protein of each clone. The proteome quantification data of 3227 proteins revealed a total of 357 and 486 deregulated proteins ($p < 0.05$) in HCT-116 and HS-5 *NOMO1*-KO cell lines, respectively. A total of 205 up-regulated and 152 down-regulated proteins were found in the HCT-116 *NOMO1*-KO clone. In HS-5, 182 proteins were up-regulated and 304 down-regulated in the *NOMO1*-KO clone compared to WT. When we compared commonly deregulated proteins in all the *NOMO1*-KO clones vs. all control clones, we identified 12 up-regulated and 16 down-regulated proteins across the two cell lines (Table S6). Results from the GSEA revealed the existence of 2 pathways (apical junction and cholesterol homeostasis) enriched ($FDR < 0.05$) by the deregulated proteins associated with *NOMO1* loss.

In addition, differentially expressed cell migration associated-proteins in the LC-IMS/MS analysis (CTND1, LMNB1 and HMGA1) were detected and confirmed by WB (Fig. 5). These results justify the alteration of the processes observed in the transcriptome (EMT and cell migration) and suggest that *NOMO1* loss could affect to the migratory capability in CRC cells.

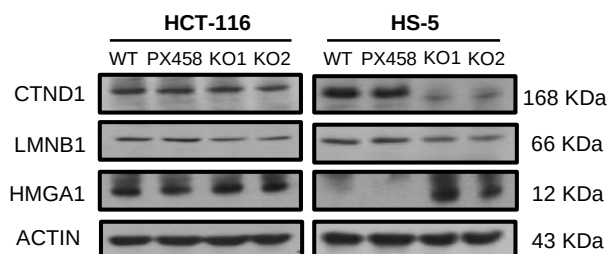


Fig. 5. Expression of differentially expressed cell migration associated-proteins (CTND1, LMNB1, HMGA1) in the LC-IMS/MS analysis of the HCT-116 and HS-5 cell lines.

Inactivation of *NOMO1* promote cell migration

To determine if *NOMO1* has an important role in cell migration, we aimed to test the migration capacity of colon cancer cells in the absence of this gene. Thus, we performed a wound healing and transwell migration assays to compare the cell migration ability of WT and *NOMO1*-KO clones in all two cell lines. The recolonized area was calculated at 24 and 48 hours for the HCT-116 cell line. For the HS-5 cell line it was calculated at 24 and 36 hours (in the wound healing assay), since at 36 hours all the wounds were closed. Consistent across all cell lines, our results showed that the percentage of migration was significantly higher in the *NOMO1*-KO clones compared to controls (t-test, $p<0.01$) (Fig. 6). Therefore, *NOMO1* loss promotes cell migration.

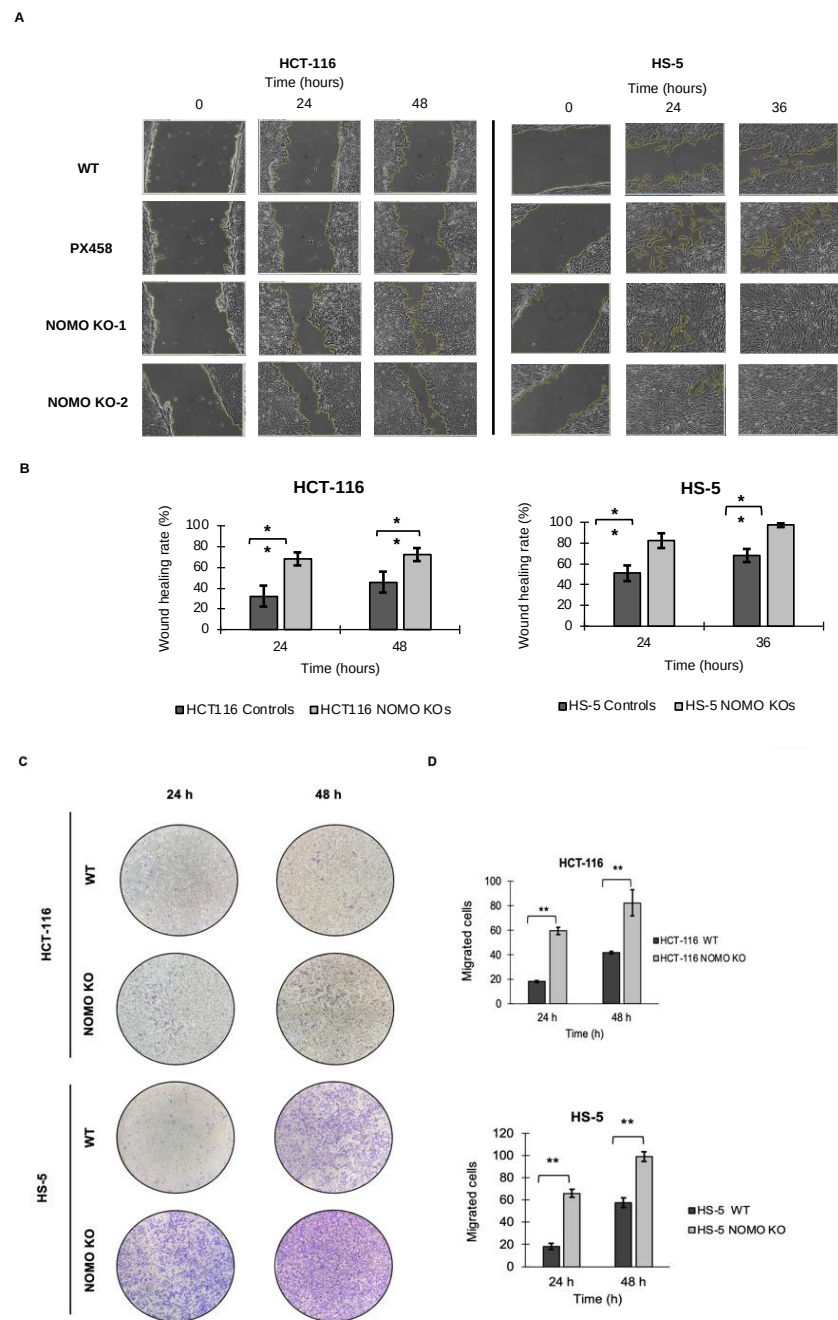


Fig. 6. Loss of *NOMO1* promote cell migration of HCT-116 and HS-5 cell lines. (A) Wound healing assay of *NOMO1*-KO and control clones analyzed at 0, 24 and 36 or 48 hours. (B) Wound healing rate (%)

representation for *NOMO1*-KO and control clones. Data are shown as the mean $\pm$ SD of the three replicates. Statistically significant differences are indicated by asterisks (** $p < 0.01$). (C) Transwell assay assessed at 24 and 48 hours of *NOMO1*-KO and control clones. (D) Quantification of cells with migration capability represented as migrated cells in the transwell assay to the two cell lines. Data are shown as the mean $\pm$ SD. Statistically significant differences are indicated by asterisks (** $p < 0.01$).

***NOMO1* deficiency in mouse colon cells does not modify the susceptibility of developing EOCRC.**

To see the effect of the somatic deletion of *NOMO1* in EOCRC, and considering that this gene could be a tumor suppressor, we decided to silence *NOMO1* expression in the mouse colon cells using a conditional mouse model(34). The selection of a conditional mouse model allowed to induce a *NOMO1* deletion at a specific time and cellular context. A total of 9 mice (8 weeks old) were colon injected with a Cre adenovirus: 6 *NOMO1*flox/flox, 2 *NOMO1*flox/+ and 1 *NOMO1*+/+. As a control group, we used 2 *NOMO1*flox/flox and 1 *NOMO1*flox/+. All mice were monitored for tumor development until 20 months after Cre activation, when they were euthanized for histological analysis of the gastrointestinal tract. *NOMO1* exon 3 deletion was checked by PCR in genomic DNA from mice of experimental and control groups. *NOMO1*flox/flox; CreAdV and *NOMO1*flox/+; CreAdV showed a fragment of 482 bp corresponding with the *NOMO1* exon 3 ablation (Fig. S5). The pathological study of both, transduced and control mice, shows no lesions in the digestive tract compatible with the development of CRC or other tumors. This result would indicate that *NOMO1* deficiency is not a driver for the development of CRC tumors in mice.

DISCUSSION

The rising burden of EOCRC with etiologies unknown is a global epidemic—particularly given that EOCRC is characterized by more aggressive phenotypes and poorer prognostic outcomes compared with late-onset disease(10). Despite initial advances in understanding the distinct disease burden—with clinical, pathologic and molecular phenotypes—most notably, few studies have explored the mechanistic underpinnings of this malignancy(50–52). Consequently, such studies of EOCRC etiology may yield translation of research findings into clinical advances to improve outcomes specifically for this growing population. Here, our study is one of the first mechanistic studies of EOCRC biology—with a specific focus on the role of *NOMO1* in colorectal carcinogenesis among young individuals. We applied CRISPR-Cas9 technology to delete endogenous *NOMO1* in multiple independent cell lines and used a gut-specific conditional KO mouse model of *NOMO1* to study subsequent tumor development. While characterization of the *NOMO1*-KO clones revealed that loss of *NOMO1* did not affect Nodal signaling pathway activity nor cell proliferation, loss of *NOMO1* increased CRC cell migration. Deletion of *NOMO1* using in vivo models also yielded no subsequent tumor development. Together, these findings suggest that other signaling pathways deregulated by the loss of *NOMO1* may play a relevant role in the pathogenesis of EOCRC.

To date, three *NOMO* genes had been described: *NOMO1*, *NOMO2* and *NOMO3*, with a sequence homology of more than 96%. In this work, we observed that the amount of *NOMO* gene amplified by qPCR was the same as the amplified for different single copy genes. These results, together with: (i) an observed frequency of *NOMO1* WT alleles close to 100%, and (ii) the finding that across different species, including the mouse, there is only one *NOMO* gene; strongly suggest the existence of a single *NOMO* gene. Thus, it could be possible that alternative *NOMO* alleles could be either a rare mutation or a sequence annotation error.

In an earlier study, we reported that 70% of EOCRCs examined had a loss of the *NOMO1* gene(13). Here, we expanded upon these findings to investigate the mutational profile of *NOMO1* in heterozygous

NOMO1-KO EOCRCs. We found that 50% of the samples presented a mutation that generated a premature stop codon resulting in a truncated protein. In addition, 10 variants of unknown clinical significance were identified, 6 of which were classified as likely pathogenic. Thus, we demonstrate that this gene can be inactivated not only by deletion but also by mutation in EOCRC tumors, which highlights the recurrence of *NOMO1* inactivation in the pathogenesis or progression of EOCRC. These data show, for the first time, the existence of pathogenic mutations in *NOMO1* gene associated with EOCRC and suggests that this gene may function as a tumor suppressor.

NOMO1, together with NCLN and TMEM147, form a protein complex located in the endoplasmic reticulum that inhibits Nodal signaling—a signal transduction pathway that maintains pluripotency in human embryonic stem cells(14,39). Previous investigations have shown that the formation of the protein complex NCLN-*NOMO1*-TMEM147 are limited by a posttranscriptional regulation mechanism that originates from the assembly of the complexes(14). The incorporation of these monomeric components in the protein complex results in their stabilization, leading to a significant increase in their half-lives(39). Although the assembly mode of the complex is not well understood, Dettmer et al. (2010) showed the critical role that NCLN plays in the formation of the protein complex by controlling the cellular steady-state levels of *NOMO1* and TMEM147, which are synthesized in excess and their monomeric forms are subjected to rapid proteolytic degradation(14). It has been described that inactivation of any of the three proteins resulted in a strong reduction in the other two proteins. However, overexpression of *NOMO1* and TMEM147, alone or in combination, did not yield increased levels of NCLN expression, suggesting that NCLN may be the limiting factor in this protein complex formation(14,39).

In this work, we also observed that *NOMO1* inactivation strongly reduced NCLN protein expression. However, the levels of TMEM147 were not modified in any of the cell lines studied herein. These results confirm the destabilization of the protein complex by modifying the levels of some components. However, our results are inconsistent with the reported reduction of TMEM147 levels by *NOMO1* inactivation(14). We hypothesize that although the protein complex is not correctly assembled, the unbound TMEM147 is not degraded by the proteasome, potentially due to its stability in the monomeric form across the analyzed cell lines. We found that the destabilization of NCLN in the absence of *NOMO1* was not accompanied by a reduction of its mRNA, as indicated by microarrays, RNA-seq and qRT-PCR analysis. These results indicate that protein reduction is caused by a posttranscriptional mechanism, in agreement with the data published by Dettmer et al.(14).

After elucidated the effect of *NOMO1* loss in the NCLN-*NOMO1*-TMEM147 complex, we evaluated its role in the activation of the Nodal signaling pathway, essential for cellular differentiation during embryonic development(20) and reactivation in multiple tumor types(21), where it has been related to increased proliferation and invasion(18,53). We found that *NOMO1* inactivation did not perturb the expression of any proteins involved in the Nodal signaling pathway, including both p-Smad2 and p-Smad3, whose levels should be elevated upon signaling pathway activation. These results suggest that the possible carcinogenic effect of *NOMO1* loss is not related to its described role as an inhibitor of the Nodal pathway. Further investigation is essential to elucidate other biological pathways that might be regulated by the inactivation of *NOMO1* in EOCRC.

To determine the role of *NOMO1* in EOCRC, we subsequently interrogated the transcriptional and the protein expression profiles of WT and *NOMO1*-KO cell lines. Our results revealed a differentially expression profile of genes commonly deregulated across the HCT-116 and HS-5 cell lines due to *NOMO1* inactivation, with the involvement of several genes associated with cell migration, invasion and EMT processes. For that reason, we first studied the protein expression of E-Cadherin, Vimentin and B-Catenin. However, these markers not showed significantly changes after *NOMO1* inactivation, suggesting that other

proteins may be involved in the deregulation of EMT and cell migration. In this sense, the LC-IMS/MS analysis revealed a set of proteins that are communally deregulated and associated with migration, invasion and EMT processes and that could respond to the inactivation of *NOMO1* by contributing to the increased migratory capacity.

For example, CTND1 (underexpressed in all two cell lines) has been associated with the stabilization of E-cadherin and the maintenance of cell adherents junctions(54), consistent with the deregulated pathways associated with our proteome analysis. Moreover, CTND1 is involved in the Wnt/beta-catenin/CTNNB1 signaling pathway by regulating the proliferation, migration and cell differentiation of endothelial cells in tumor growth(55). Other studies have described that loss of CTND1 is associated with poor prognosis and metastasis in ductal breast cancer and is considered a tumor suppressor(56). In head and neck squamous carcinoma (HNSCC) and oral squamous cell carcinoma (OSCC) cell lines, loss of CTND1 causes increased migration and invasion and favors EMT(57,58). Lamin B1 (LMNB1), underexpressed in all two cell lines in our study, has been described as tumor suppressor in lung cancer. Lamin B1 knockdown in lung epithelial cells promoted EMT, cell migration, tumor growth and metastasis by activation of the RET/p38 signaling(59). HMGA1, up-regulated in *NOMO1*-KO cells, is overexpressed in cervical cancer tissues and positively correlate with lymph node metastasis and advanced clinical stage. In addition, HMGA1 enhances tumor growth and accelerate migration and invasion via targeting miR-221/222 in cervical cancer cell lines(60).

Previous studies have shown that Nodal expression can increase invasive and metastatic capacity of tumor cells(21). Although we did not find an increased activation of the Nodal signaling pathway upon *NOMO1* loss, we investigated the effect of *NOMO1* loss in cellular migration capacity due to the alteration of the cell migration associated-proteins (CTND1, LMNB1, HMGA1) discusses above. Strikingly, we found that inactivation of *NOMO1* increased the migration ability of all cell lines analyzed. Taken together, these data provide a possible explanation for the increased migration capacity exhibited by the *NOMO1*-KO clones and warrant further study in the role of these genes in EOCRC development. Although an increased migration induced by *NOMO1* loss could contribute to colorectal carcinogenesis, we cannot exclude the possibility that other signaling pathways deregulated by *NOMO1* loss may play a relevant role in the pathogenesis of the disease.

To study the *NOMO1* role in the colorectal carcinogenesis, a conditional *NOMO1* mouse model was used to analyze this gene function in a rigorous and specific way that avoids embryonic lethality(34). This animal model, which use the Cre-loxp system to delete a specific gene in a particular tissue, has been efficiently tested by other groups(61,62). For example, Huang et al. used this method to study the tumor formation in the colon(63). The results provided by our in vivo models indicate that loss of *NOMO1* does not directly contribute to the development of colorectal tumors in mice after 20 months of follow-up. In this case, we think that the cellular and environmental influence in the mouse colon cells differ from the cellular context in human colon during colorectal carcinogenesis.

Together, our data suggest that *NOMO1*, despite being deleted in 70% of EOCRC(13), plays a passenger role in the development of this disease. It also may suggest that other coding or non-coding genes, located in the same chromosomal region (16p13.11-13.12), could acts as a driver gene in early-onset colorectal carcinogenesis. Consequently, further mechanistic studies are needed to understand the role of this putative tumor suppressor in EOCRC and the distinct mechanistic underpinnings of early-onset colorectal carcinogenesis.

MATERIAL AND METHODS

Human tissue samples

Twenty formalin-fixed paraffin-embedded (FFPE) blocks of tumors isolated from EOCRC patients diagnosed at the “12 of October” University Hospital (Madrid) and at the “University Hospital of Salamanca” were included in this study. All tumors came from patients under 50 years who had previously showed heterozygous deletion of the *NOMO1* gene(13). Clinicopathological data, microsatellite instability (MSI) status and mismatch repair (MMR) gene mutations status were analyzed in all cases. All tumor samples were confirmed to be microsatellite stable (MSS)/sporadic and to carry WT MMR genes.

Cell lines and culture conditions

The colon cancer cell line HCT-116 and the human stromal cell line HS-5 were acquired from ATCC (American Type Culture Collection). The cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich) supplemented with 10% of fetal bovine serum (FBS) (Sigma-Aldrich) and 1% penicillin/streptomycin (Gibco Life Technologies, Grand Island, NY). All cells were incubated at 37°C in a 5% CO₂. The presence of mycoplasma was routinely checked with MycoAlert kit (Lonza, Basel, Switzerland) and only mycoplasma free cells were used in all subsequent experiments.

DNA extraction and DNA quality evaluation

DNA was isolated from 10µm sections from each FFPE block. FFPE sections were treated with deparaffinization solution (Qiagen, Heidelberg, Germany) and DNA extraction was performed using the QIAamp DNA FFPE Tissue kit according to manufacturer's instructions (Qiagen, Heidelberg, Germany). The Illumina Custom DNA Panel Reference Guide (Illumina, San Diego, USA) was used to measure DNA quality by comparing FFPE-gDNA amplification potential with a reference non-FFPE gDNA (QCT). Using qPCR, delta Cq values was calculated to predict the dilution required for each sample.

Quantitative PCR (qPCR) and quantitative reverse transcription PCR (qRT-PCR)

For real-time quantification, 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.). The primers used for gene amplification are detailed in Table S1. For all cases a fragment of LEMD3 gene (single copy gene) was amplified from the same DNA sample for internal control. PCR experiments were performed using Applied Biosystems StepOnePlus® and RQ Manager® software for analysis. Three replicates were used for each PCR reaction. The comparative Ct method ($2^{-\Delta\Delta Ct}$) was applied to calculate the relative expression levels of each amplicon using LEMD3 as the reference gene for normalization. RT-PCR specificity of each PCR reaction was verified by melting curve analysis.

NOMO1 sequencing

DNA samples with optimal quality were used to sequence *NOMO1* within a panel of genes by Next Generation Sequencing (NGS), according to manufacturer's instructions (Illumina, San Diego, USA). Briefly, this process included following steps: hybridization with the oligo pool, removing unbound oligos, extension and ligation with bound oligos and amplification of the libraries. Finally, the libraries were cleaned up by magnetic beads. PCR products were quantified using a Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) and the libraries were normalized at 4nM in a final pool. Sequencing was performed by a MiSeq System (Illumina, San Diego, USA). Variant Studio Software was used for subsequent analysis (Illumina, San Diego,

USA). Somatic variants with >10% frequency, a quality score >500 in the bi-directional sequencing quality filter and met the software PASS filter were reported. In silico analyses to predict the pathogenicity of variants of uncertain significance (VUS) were performed using the prediction programs: PolyPhen (Polymorphism Phenotyping), Sift (Sorting Intolerant From Tolerant) and CADD (Combined Annotation Dependent Depletion).

CRISPR/Cas9-mediated generation of *NOMO1* knockout cells

Exon 3 and adjacent intronic regions of the *NOMO1* gene were selected as target sequences for the CRISPR-Cas9 design. This is because it is the first exon of the gene where the last nucleotide of its coding sequence encodes a codon with the first two nucleotides of the next exon. Therefore, removal this fragment from exon 3 could produce an early truncated *NOMO1* protein. Three sgRNAs (sgRNA1, sgRNA2 and sgRNA3) were designed with the Spanish National Biotechnology Centre (CNB)-CSIC web tool(24). Three complementary oligonucleotides corresponding to sgRNAs were designed, including two 4-bp overhang sequences (*NOMO1* UP and LOW oligonucleotides; Table S2). Each pair of oligonucleotides was phosphorylated, annealed, digested with Bpil enzyme (NEB) and ligated to the pSpCas9(BB)-2A-GFP (PX458) also digested with the same enzyme(25).

Subsequently, 50µl of *E. coli* DH5α cells were transformed with 2µl of ligated plasmids. Single colonies were grown, and plasmid DNA was extracted and purified using the Danagene plasmid miniprep kit per manufacturer protocol (DANAGEN-BIOTED S.L.). The correct insertion of the sgRNAs into the vector was confirmed by Sanger sequencing (data not shown). Cell lines were transfected with a combination of two different *NOMO1* CRISPR-Cas9 KO plasmids (4µg each) or 8µg of control CRISPR-Cas9 plasmid (empty PX458).

Transfection for the HCT-116 cell line was carried out using the Amaxa Cell Line Nucleofector Kit V, the Amaxa Nucleofector device (Lonza) with the T-016 program. HS-5 cell line was transfected by lipofection per manufacturer instructions (PolyJet™ In Vitro DNA Transfection Reagent, SignaGen). Single GFP+ cells were sorted into 96-well plates 48 hours after transfection using a BD FACSAria™ III flow cytometer. Isolated clones were expanded in culture over a period of three weeks and then genomic DNA was extracted by the phenol-chloroform method. We confirmed the *NOMO1* status of each clone by PCR using the *NOMO1* exon/intron 3 primers (Forward: 5'-CAGTGCTCAGTACCATGTAG-3'; Reverse: 5'-GGGAGGAATACAAACCTC-3'). At least two *NOMO1*-KO clones and two controls (WT cell line or clones generated after transfection with the empty plasmid) were assayed for each cell line. In addition, the *NOMO1*-KO clones were also confirmed by qPCR and Western blot (WB).

Western blotting

Cells were resuspended in RIPA buffer containing protease inhibitors (Complete, Roche Applied Science, Indianapolis), and protein concentration was measured using the Bradford assay (BioRad). 30µg of protein samples were separated on an 8% or 12% SDS-PAGE gel (depending on protein size), transferred to an Immobilon-P membrane (Millipore) and incubated with primary specific antibodies overnight (Table S3). The following day secondary antibodies were added, immunoblots were incubated for 1h at room temperature and developed using enhanced chemiluminescence WB detection reagents (Thermo Fisher Scientific). β-actin was used as a loading control.

Cell viability assay

MTT assay was used to quantify cell viability and metabolic activity. Metabolically active cells were marked using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (Sigma-Aldrich). A total of

5000 cells/well from the *NOMO1*-KO and control clones of the two cell lines were seeded in a 96-well plate. We measured the viability of cells by cellular metabolic function at 0 (control), 24, 48 and 72 hours, adding to each well 10 μ L of MTT and incubating for 1 hour at 37°C. The percentage of growth was determined by statistical analysis with SPSS-IBM software.

Wound healing assay

106 cells per well were seeded in 6-well plates containing 3 well silicone inserts with two defined cell-free gaps (Ibidi USA, Inc.). A minimal concentration of FBS was added to the culture medium to maintain survival but inhibit cell proliferation (2% for HCT-116 and HS-5 cell lines). After cell attachment (24 h), culture-inserts were removed creating a scratch. Each experiment was performed in triplicate. Photos of a determined region of each scratch (three wounds/replica) were taken every 10 minutes for 48 hours using a camera attached to a Mikon ECLIPSE TE-2000-E microscope. The ImageJ® program was used in combination with the MRI Wound Healing Tool to calculate the percentage of migration (shown on the Y-axis) for each time reference at 0, 12, 24 and 36 or 48 hours (shown on the X-axis). The SPSS/IBM software was used to calculate migration percentages for *NOMO1*-KO and control clones.

Transwell migration assay

For the transwell migration assay, 4x10⁴ cells of two HCT-116 and HS-5 *NOMO1*-KO clones and 2 control clones were suspended in 300 μ L of serum free DMEM. This cell suspension was added to the upper chamber of 24-multiwell insert system with 8 μ m pore (SARSTEDT) and 600 μ L of DMEM with 10% FBS was added to each lower well. After 24-hour incubation, migratory cells were fixed with 3.7% paraformaldehyde for 5 minutes and stained with 1% crystal violet solution for 15 minutes. Microscopy pictures were taken for each well with a 10x objective lens and the stained area was quantified with ImageJ® software.

RNA extraction, microarray data analysis and RNA sequencing

Total RNA was extracted from three *NOMO1*-KO clones and three control clones for HCT-116 and HS-5 using RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA was removed from the samples using RNase-free DNase Set (Qiagen, Hilden, Germany). RNA integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA).

For the microarray analysis, labelling and hybridizations were performed according to protocols from Affymetrix. Washing and scanning were performed using the Affymetrix GeneChip System (GeneChip Hybridization Oven 645, GeneChip Fluidics Station 450 and GeneChip Scanner 7G). Raw microarray data were quantile normalized, background corrected and log₂ transformed by the Robust Multi-array Average (RMA) procedure(26) in the oligo R package (v.1.54.1) using a custom cdf reference from BrainArray (v.25.0.0)(27). Differential gene expression analyses were carried out with the limma package (v.3.46.0). Microarray batch-effects were adjusted through the ComBat function from the sva package.

For RNA sequencing (RNA-seq), 0.5 μ g of total RNA was used to construct cDNA libraries with TruSeq stranded total RNA with Ribo-Zero kit (Illumina, San Diego, CA, USA). Then, cDNA libraries were sequenced with NovaSeq 6000 (paired-end 150bp x 2) with a range of 61.8-78.7M reads/sample according to the manufacturer instructions. Library construction and sequencing were performed by Macrogen Inc. (Macrogen, Seoul, Korea). Adapter sequences and low-quality bases were trimmed using Trimmomatic (v.0.39). Raw and trimmed reads quality was assessed with the FASTQC tool (v.0.11.9). The surviving paired reads after trimming were mapped to hg19 human genome with the STAR aligner (v.2.7.9a). The reference genome sequence (hg19, Genome Reference Consortium GRCh37) and annotation data (v.87)

were downloaded from the Ensembl website (<https://www.ensembl.org>). Gene level counts were calculated using the union mode from the HTSeq package (v.0.12.4). Gene expression count normalization through the median of ratios method(28) and differential gene expression analyses were performed with the DESeq2 R package (v.1.30.1). Batch-effects were adjusted through the ComBat_seq function from the sva R package (v.3.38.0).

Microarray and RNA-seq data have been deposited into the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/gds/>) with the accession number GSE198383. For both assays, the false discovery rate (FDR) was controlled by adjusting p-values with the Benjamini–Hochberg method. Genes with an absolute value of the fold change (FC) greater than 1.5 (over-expressed) and lower than -1.5 (under-expressed) were selected for further analysis. Gene expression values of the selected genes were visualized using a heat-map with the pheatmap R package (v.4.0.5). The WebGestalt tool (WEB-based Gene set analysis toolkit)(29) was used for pathway enrichment analyses. Hallmarks standards were used to find the biological processes that could be affected by the *NOMO1* inactivation.

Proteome analysis by liquid chromatography-mass spectrometry (LC-IMS/MS)

Changes in protein expression profiles after *NOMO1* inactivation were analyzed by liquid chromatography-mass spectrometry (LC-IMS/MS). A total of 4 samples of a *NOMO1*-WT and *NOMO1*-KO clones were compared for the HCT-116 and HS-5 cell lines. All purified cells (2000-3000 cells) were processed by lysis solution and phosphates-protease inhibitors(30). 0.5 µg of total protein was reduced with 10mM Dithiothreitol (DTT) + 55mM iodiacetamide at room temperature for 45 min. Protein was digested with trypsin (1:50 w/w) at 37 °C for 18h; then, the peptide mixture was acidified with 0.1% TFA & desalted with C18 StageTips. Samples were stored at -20 °C until LC-IMS/MS tests were performed. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD033636. For the LC-IMS/MS analysis, a nanoUPLC system (NanoElute, Bruker) was used with a C18 column 15cm x75µm, with 1.6µm C18 particles (Ion Optics Inc.) and 120 min. gradient (3-50% of ACN at 300nL/min.), which is coupled to a TimsTOF Pro (Bruker). The TimsTOF Pro was operated in PASEF mode (Parallel Accumulation Serial Fragmentation) using Compass Hystar 5.036.0. Settings for 11 samples per day method as follows: Mass Range 100 to 1700m/z, 1/K0 Start 0.6-End 1.6 V·s/cm², Ramp time 110.1ms, Lock Duty Cycle 100%, Capillary Vol. 1600V, Dry Gas 3 L/min, Dry Temp 180°C, PASEF settings: 10 MS/MS scans (total cycle 1.27sec), charge range 0-5, active exclusion 0.4 min, Schedul Target intensity 10000, Intensity threshold 2500, CID collision energy 42eV(31). All raw files were analyzed by MaxQuant v1.6.6.0 software using the integrated Andromeda search engine. Data search was against the Human Uniprot Reference Proteome with isoforms (latest version available) and a separated reverse decoy database using a strict trypsin specificity allowing up to 2 missed cleavages. The minimum peptide length set to 7αα'Carbamidomethylation of Cysteine as fixed modification, and N-acetylation and oxidation of Methionine as variable modifications. First search peptide tolerance set at 70 ppm-main search set at 30 ppm (other settings left as default). Single shot samples and fractions assign as separate parameter groups and matching between runs was turned on and set as and to" for single shot samples and "from" for fractions. Maximum peptide mass (Da) set at 8000. All other settings in group or global parameters were as default for MaxQuant. PTMS screening was carried out following PTMScan Direct PTMScan® m (Cell Signaling Technology, USA) with the slight modifications described(32,33).

Mouse strains, adenovirus injection and histological analysis.

Conditional mouse mutant for *NOMO1* has been previously described(34). In summary, to avoid *NOMO1* expression in C57BL/6J mouse intestine cells, *NOMO1*flox/flox and *NOMO1*flox/+ mice were infected by single injection of 300 MOI of Cre adenovirus (Ad5CMVCre-eGFP) into the colon area of 8 weeks old mice (named *NOMO1*flox/flox CreAdV and *NOMO1*flox/+ CreAdV). We used *NOMO1*flox/flox, *NOMO1*flox/+, and *NOMO1*+/+ uninfected mice as a control group. All groups were euthanized for CCR tumor observation 20 months old. PCRs using the m*NOMO1*-F and m*NOMO1*-R specific primers (Table S1) of genomic DNA from mice colons were performed to demonstrate whether Cre activity effectively removes *NOMO1* exon 3 achieving homozygous or heterozygous *NOMO1* null alleles. Specifically, mice were always housed in a temperature-controlled facility (specific pathogen free, spf) using individually ventilated cages, standard diet and a 12hr light/dark cycle, according to EU at the Servicio de Experimentación Animal of the University of Salamanca. For histological analysis, after the necropsy of the mice their digestive tracts were removed and fixed in formol 10% solution. All samples were processed into serial paraffin sections and stained with haematoxylin-eosin. Subsequently, a pathological analysis of the digestive tract of all mice was carried out at the Pathology Service of the University Hospital of Salamanca.

Statistical analysis

Data were analyzed using IBM/SPSS software v26 (SPSS Inc., Chicago, IL). All quantitative data are showed as the mean $\pm$ standard deviation (SD). Student's t-test was used to compared differences between two groups when they had a normal distribution (test Kolmogorov-Smirnov; p-value>0.05). A p-value $\leq$ 0.05 was considered statistically significant.

Study approval

The study with animals followed Spanish and European Union guidelines for animal experimentation (RD1201/05, RD 53/2013, and 86/609/CEE). Genetically modified animals were generated at Transgenic Facility, Nucleus, University of Salamanca and the generation was approved by the Bioethics Committee of the University of Salamanca and "Consejería de Agricultura y Ganadería de la Junta de Castilla y León" (ref 338).

Acknowledgements

The authors thank the Anatomy Pathology Service from the Hospital University of Salamanca and the Transgenic Service from the University of Salamanca for their assistance.

Conflict of Interest Statement

The authors have declared that no conflict of interest exists.

Availability of data and materials:

The datasets generated and/or analyzed during the current study are available in the GEO (ID GSE198383; <http://www.ncbi.nlm.nih.gov/gds/>) and PRIDE (PXD033636; <http://proteomecentral.proteomexchange.org>) repositories.

Funding:

This study was supported by the health research program of the "Instituto de Salud Carlos III" (Spanish Ministry of Economy and Competitiveness, PI20/01569 and PI20/0974) co-funded with FEDER funds, and "Mutua Madrileña Foundation" (FMM20/001). A. M.-M was supported by a predoctoral research

grant from the Dr. Moraza Foundation (FMoraza18/001). P.G.V and N.G.-U were supported by a predoctoral research grant from the Consejería de Educación—Junta de Castilla y León. A.N.H. was supported by the National Institutes of Health K12 HD043483 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Author's contributions

Conceptualization, R.G.-S and J.P; Methodology, J.P.G, A.M.-M, P.G.-V, N.G.-U and M.S.-M; Validation, R.G.-S; Formal analysis and investigation, J.L.G, L.C; Writing - original draft preparation, J.P.G and A.M.-M; Writing - review and editing, J.P.G, A.M.-M, AB.H, AN.H, J.P. and R.G.-S; Funding acquisition, J.P and R.G.-S; Resources, R.V.T, O.B, L.M, M.S.-M and M.F; Supervision, R.G.-S. All authors have read and agreed to the published version of the manuscript.

References:

1. H Sung, J Ferlay, RL Siegel, M Laversanne, I Soerjomataram, A Jemal, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* [Internet]. 2021 May [cited 2021 Sep 14];71(3):209–49. Available from: <https://pubmed.ncbi.nlm.nih.gov/33538338/>
2. M Arnold, MS Sierra, M Laversanne, I Soerjomataram, A Jemal, F Bray. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* [Internet]. 2017 Apr 1 [cited 2021 Sep 14];66(4):683–91. Available from: <https://pubmed.ncbi.nlm.nih.gov/26818619/>
3. J Perea, Y Rodríguez, D Rueda, JC Marín, J Díaz-Tasende, E Álvaro, et al. Early-onset colorectal cancer is an easy and effective tool to identify retrospectively Lynch syndrome. *Ann Surg Oncol* [Internet]. 2011 Nov [cited 2021 Jul 29];18(12):3285–91. Available from: <https://pubmed.ncbi.nlm.nih.gov/21590452/>
4. G Mauri, A Sartore-Bianchi, AG Russo, S Marsoni, A Bardelli, S Siena. Early-onset colorectal cancer in young individuals. *Mol Oncol* [Internet]. 2019 Feb 1 [cited 2021 Jul 29];13(2):109–31. Available from: <https://pubmed.ncbi.nlm.nih.gov/30520562/>
5. J Perea, D Rueda, A Canal, Y Rodríguez, E Álvaro, I Osorio, et al. Age at onset should be a major criterion for subclassification of colorectal cancer. *J Mol Diagn* [Internet]. 2014 Jan [cited 2021 Jul 29];16(1):116–26. Available from: <https://pubmed.ncbi.nlm.nih.gov/24184227/>
6. E Budinska, V Popovici, S Tejpar, G D'Ario, N Lapique, KO Sikora, et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* [Internet]. 2013 Sep 1 [cited 2021 Jul 29];231(1):63–76. Available from: <https://pubmed.ncbi.nlm.nih.gov/23836465/>
7. F Jasmine, R Rahaman, C Dodsworth, S Roy, R Paul, M Raza, et al. A genome-wide study of cytogenetic changes in colorectal cancer using SNP microarrays: opportunities for future personalized treatment. *PLoS One* [Internet]. 2012 Feb 20 [cited 2021 Jul 29];7(2). Available from: <https://pubmed.ncbi.nlm.nih.gov/22363777/>
8. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* [Internet]. 2012 Jul 19 [cited 2021 Jul 29];487(7407):330–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/22810696/>
9. AN Holowatyj, B Gigic, E Herpel, A Scalbert, M Schneider, CM Ulrich. Distinct Molecular Phenotype of Sporadic Colorectal Cancers Among Young Patients Based on Multiomics Analysis. *Gastroenterology* [Internet]. 2020 Mar 1 [cited 2021 Sep 14];158(4):1155–1158.e2. Available from: <https://pubmed.ncbi.nlm.nih.gov/31730769/>
10. M Arriba, C Sánchez, A Vivas, OA Nutu, D Rueda, S Tapial, et al. Intermediate-onset colorectal cancer: A clinical and familial boundary between both early and late-onset colorectal cancer. *PLoS One* [Internet]. 2019 May 1 [cited 2021 Sep 14];14(5). Available from: <https://pubmed.ncbi.nlm.nih.gov/31095598/>
11. M Di Leo, RA Zuppardo, M Puzzono, I Ditunno, A Mannucci, G Antoci, et al. Risk factors and clinical characteristics of early-onset colorectal cancer vs. late-onset colorectal cancer: a case-case study. *Eur J Gastroenterol Hepatol* [Internet]. 2021 [cited 2021 Sep 14];33(9):1153–60. Available from: <https://pubmed.ncbi.nlm.nih.gov/33208680/>
12. M Arriba, JL García, L Inglada-Pérez, D Rueda, I Osorio, Y Rodríguez, et al. DNA copy number profiling reveals different patterns of chromosomal instability within colorectal cancer according to the age of onset. *Mol Carcinog* [Internet]. 2016 May 1 [cited 2021 Jul 29];55(5):705–16. Available from: <https://pubmed.ncbi.nlm.nih.gov/25808986/>

13. J Perea, JL García, J Pérez, D Rueda, M Arriba, Y Rodríguez, et al. NOMO-1 gene is deleted in early-onset colorectal cancer. *Oncotarget* [Internet]. 2017 [cited 2021 Jul 29];8(15):24429–36. Available from: <https://pubmed.ncbi.nlm.nih.gov/28416736/>
14. U Dettmer, PH Kuhn, C Abou-Ajram, SF Lichtenthaler, M Krüger, E Kremmer, et al. Transmembrane protein 147 (TMEM147) is a novel component of the Nicalin-NOMO protein complex. *J Biol Chem* [Internet]. 2010 Aug 20 [cited 2021 Jul 29];285(34):26174–81. Available from: <https://pubmed.ncbi.nlm.nih.gov/20538592/>
15. C Haffner, M Frauli, S Topp, M Irmeler, K Hofmann, JT Regula, et al. Nicalin and its binding partner Nomo are novel Nodal signaling antagonists. *EMBO J* [Internet]. 2004 Aug 4 [cited 2021 Jul 29];23(15):3041–50. Available from: <https://pubmed.ncbi.nlm.nih.gov/15257293/>
16. F Francescangeli, P Contavalli, ML De Angelis, M Baiocchi, G Gambara, A Pagliuca, et al. Dynamic regulation of the cancer stem cell compartment by Cripto-1 in colorectal cancer. *Cell Death Differ* [Internet]. 2015 Oct 9 [cited 2021 Jul 29];22(10):1700–13. Available from: <https://pubmed.ncbi.nlm.nih.gov/26343543/>
17. JA Kelber, AD Panopoulos, G Shani, EC Booker, JC Belmonte, WW Vale, et al. Blockade of Cripto binding to cell surface GRP78 inhibits oncogenic Cripto signaling via MAPK/PI3K and Smad2/3 pathways. *Oncogene* [Internet]. 2009 Jun 18 [cited 2021 Jul 29];28(24):2324–36. Available from: <https://pubmed.ncbi.nlm.nih.gov/19421146/>
18. Y Gong, Y Guo, Y Hai, H Yang, Y Liu, S Yang, et al. Nodal promotes the self-renewal of human colon cancer stem cells via an autocrine manner through Smad2/3 signaling pathway. *Biomed Res Int* [Internet]. 2014 [cited 2021 Jul 29];2014. Available from: <https://pubmed.ncbi.nlm.nih.gov/24696849/>
19. W Duan, R Li, J Ma, J Lei, Q Xu, Z Jiang, et al. Overexpression of Nodal induces a metastatic phenotype in pancreatic cancer cells via the Smad2/3 pathway. *Oncotarget* [Internet]. 2015 [cited 2021 Jul 29];6(3):1490–506. Available from: <https://pubmed.ncbi.nlm.nih.gov/25557170/>
20. MM Shen. Nodal signaling: developmental roles and regulation. *Development* [Internet]. 2007 Mar [cited 2021 Jul 29];134(6):1023–34. Available from: <https://pubmed.ncbi.nlm.nih.gov/17287255/>
21. Quail Daniela F, Siegers Gabrielle M, Jewer Michael, Postovit Lynne Marie, DF Quail, GM Siegers, et al. Nodal signalling in embryogenesis and tumourigenesis. *Int J Biochem Cell Biol* [Internet]. 2013 Apr [cited 2021 Jul 29];45(4):885–98. Available from: <https://pubmed.ncbi.nlm.nih.gov/23291354/>
22. M Mancino, L Strizzi, C Wechselberger, K Watanabe, M Gonzales, S Hamada, et al. Regulation of human Cripto-1 gene expression by TGF-beta1 and BMP-4 in embryonal and colon cancer cells. *J Cell Physiol* [Internet]. 2008 Apr [cited 2021 Jul 29];215(1):192–203. Available from: <https://pubmed.ncbi.nlm.nih.gov/17941089/>
23. TM Bodenstine, GS Chandler, RE Seftor, EA Seftor, MJ Hendrix. Plasticity underlies tumor progression: role of Nodal signaling. *Cancer Metastasis Rev* [Internet]. 2016 Mar 1 [cited 2021 Jul 29];35(1):21–39. Available from: <https://pubmed.ncbi.nlm.nih.gov/26951550/>
24. Breaking-Cas: off-targets-free gRNAs for CRISPR/Cas technology [Internet]. [cited 2021 Jul 29]. Available from: <https://bioinfogp.cnb.csic.es/tools/breakingcas/>
25. FA Ran, PD Hsu, J Wright, V Agarwala, DA Scott, F Zhang. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* [Internet]. 2013 [cited 2021 Jul 29];8(11):2281–308. Available from: <https://pubmed.ncbi.nlm.nih.gov/24157548/>
26. Irizarry Rafael A, Hobbs Bridget, Collin Francois, Beazer-Barclay Yasmin D, Antonellis Kristen J, Scherf Uwe, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* [Internet]. 2003 [cited 2022 Mar 4];4(2):249–64. Available from: <https://pubmed.ncbi.nlm.nih.gov/12925520/>
27. Dai Manhong, Wang Pinglang, Boyd Andrew D, Kostov Georgi, Athey Brian, Jones Edward G, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* [Internet]. 2005 [cited 2022 Mar 4];33(20). Available from: <https://pubmed.ncbi.nlm.nih.gov/16284200/>
28. Anders Simon, Huber Wolfgang. Differential expression analysis for sequence count data. *Genome Biol* [Internet]. 2010 Oct 27 [cited 2022 Mar 4];11(10). Available from: <https://pubmed.ncbi.nlm.nih.gov/20979621/>
29. Liao Yuxing, Wang Jing, Jaehnig Eric J, Shi Zhiao, Zhang Bing. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* [Internet]. 2019 Jul 1 [cited 2022 Mar 4];47(W1):W199–205. Available from: <https://pubmed.ncbi.nlm.nih.gov/31114916/>
30. Díez Paula, Droste Conrad, Dégano Rosa M, González-Muñoz María, Ibarrola Nieves, Pérez-Andrés Martín, et al. Integration of Proteomics and Transcriptomics Data Sets for the Analysis of a Lymphoma B-Cell Line in the Context

- of the Chromosome-Centric Human Proteome Project. *J Proteome Res* [Internet]. 2015 Sep 4 [cited 2022 Jan 13];14(9):3530–40. Available from: <https://pubmed.ncbi.nlm.nih.gov/26216070/>
31. Meier Florian, Brunner Andreas David, Koch Scarlet, Koch Heiner, Lubeck Markus, Krause Michael, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* [Internet]. 2018 Dec 1 [cited 2022 Jan 13];17(12):2534–45. Available from: <https://pubmed.ncbi.nlm.nih.gov/30385480/>
 32. Lawrence Robert T, Searle Brian C, Llovet Ariadna, Villén Judit. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat Methods* 2016 135 [Internet]. 2016 Mar 28 [cited 2022 Jan 13];13(5):431–4. Available from: <https://www.nature.com/articles/nmeth.3811>
 33. Stokes Matthew P, Farnsworth Charles L, Moritz Albrecht, Silva Jeffrey C, Jia Xiaoying, Lee Kimberly A, et al. PTMScan Direct: Identification and Quantification of Peptides from Critical Signaling Proteins by Immunoaffinity Enrichment Coupled with LC-MS/MS. *Mol Cell Proteomics*. 2012 May 1;11(5):187–201.
 34. García-Tuñón Ignacio, Vuelta Elena, Lozano Laura, Herrero María, Méndez Lucía, Palomero-Hernandez Javier, et al. Establishment of a conditional Nomo1 mouse model by CRISPR/Cas9 technology. *Mol Biol Rep* [Internet]. 2020;47(2):1381–91. Available from: <https://doi.org/10.1007/s11033-019-05214-7>
 35. NOMO1 Gene - GeneCards | NOMO1 Protein | NOMO1 Antibody [Internet]. [cited 2021 Jul 29]. Available from: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=NOMO1>
 36. TD Schmittgen, KJ Livak. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* [Internet]. 2008 May [cited 2021 Jul 29];3(6):1101–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/18546601/>
 37. Home - SNP - NCBI [Internet]. [cited 2021 Jul 29]. Available from: <https://www.ncbi.nlm.nih.gov/snp/>
 38. ExAC browser [Internet]. [cited 2021 Jul 29]. Available from: <http://exac.broadinstitute.org/>
 39. C Haffner, U Dettmer, T Weiler, C Haass. The Nicastrin-like protein Nicalin regulates assembly and stability of the Nicalin-nodal modulator (NOMO) membrane protein complex. *J Biol Chem* [Internet]. 2007 Apr 6 [cited 2021 Jul 29];282(14):10632–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/17261586/>
 40. Geng Yibo, Xu Cheng, Wang Yi, Zhang Liwei. Quiescin Sulfhydryl Oxidase 1 Regulates the Proliferation, Migration and Invasion of Human Glioblastoma Cells via PI3K/Akt Pathway. *Onco Targets Ther* [Internet]. 2020 [cited 2022 Apr 21];13:5721–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/32606784/>
 41. Malinowski Mariusz, Pietraszek Katarzyna, Perreau Corinne, Boguslawski Mateusz, Decot Véronique, Stoltz Jean François, et al. Effect of lumican on the migration of human mesenchymal stem cells and endothelial progenitor cells: involvement of matrix metalloproteinase-14. *PLoS One* [Internet]. 2012 Dec 7 [cited 2022 Apr 21];7(12). Available from: <https://pubmed.ncbi.nlm.nih.gov/23236386/>
 42. Wang Shang, Zhong Li, Li Yin, Xiao Desheng, Zhang Ruhua, Liao Dan, et al. Up-regulation of PCOLCE by TWIST1 promotes metastasis in Osteosarcoma. *Theranostics* [Internet]. 2019 [cited 2022 Apr 21];9(15):4342–53. Available from: <https://pubmed.ncbi.nlm.nih.gov/31285765/>
 43. Yun Da Peng, Wang Yu Qi, Meng De Long, Ji Yuan Yuan, Chen Ju Xiang, Chen Hong Yan, et al. Actin-capping protein CapG is associated with prognosis, proliferation and metastasis in human glioma. *Oncol Rep* [Internet]. 2018 Mar 1 [cited 2022 Apr 21];39(3):1011–22. Available from: <https://pubmed.ncbi.nlm.nih.gov/29399702/>
 44. Wu Ziqiang, Dai Wenyong, Wang Pei, Zhang Xiaozhen, Tang Yi, Liu Lin, et al. Periostin promotes migration, proliferation, and differentiation of human periodontal ligament mesenchymal stem cells. *Connect Tissue Res* [Internet]. 2018 Mar 4 [cited 2022 Apr 21];59(2):108–19. Available from: <https://pubmed.ncbi.nlm.nih.gov/28301220/>
 45. Löbel Madlen, Bauer Sandra, Meisel Christian, Eisenreich Andreas, Kudernatsch Robert, Tank Juliane, et al. CCN1: a novel inflammation-regulated biphasic immune cell migration modulator. *Cell Mol Life Sci* [Internet]. 2012 Sep [cited 2022 Apr 21];69(18):3101–13. Available from: <https://pubmed.ncbi.nlm.nih.gov/22527715/>
 46. Kiwanuka Elizabeth, Andersson Lauren, Caterson Edward J, Junker Johan PE, Gerdin Bengt, Eriksson Elof. CCN2 promotes keratinocyte adhesion and migration via integrin $\alpha 5 \beta 1$. *Exp Cell Res* [Internet]. 2013 Nov 15 [cited 2022 Apr 21];319(19):2938–46. Available from: <https://pubmed.ncbi.nlm.nih.gov/23988606/>
 47. Nam Kee Soo, Oh Sunhwa, Lee Kyung min, Yoo Seung ah, Shin Incheol. CD44 regulates cell proliferation, migration, and invasion via modulation of c-Src transcription in human breast cancer cells. *Cell Signal* [Internet]. 2015 Sep 1 [cited 2022 Apr 21];27(9):1882–94. Available from: <https://pubmed.ncbi.nlm.nih.gov/25979842/>

48. Shen Yuan, Wang Xiaoyu, Xu Jianping, Lu Lin. SerpinE2, a poor biomarker of endometrial cancer, promotes the proliferation and mobility of EC cells. *Cancer Biomark* [Internet]. 2017 [cited 2022 Apr 21];19(3):271–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/28453461/>
49. Oommen Saji, Gupta Shiv K, Vlahakis Nicholas E. Vascular endothelial growth factor A (VEGF-A) induces endothelial and cancer cell migration through direct binding to integrin $\alpha_9\beta_1$: identification of a specific $\alpha_9\beta_1$ binding site. *J Biol Chem* [Internet]. 2011 Jan 14 [cited 2022 Apr 21];286(2):1083–92. Available from: <https://pubmed.ncbi.nlm.nih.gov/21071450/>
50. N Akimoto, T Ugai, R Zhong, T Hamada, K Fujiyoshi, M Giannakis, et al. Rising incidence of early-onset colorectal cancer - a call to action. *Nat Rev Clin Oncol* [Internet]. 2021 Apr 1 [cited 2021 Jul 29];18(4):230–43. Available from: <https://pubmed.ncbi.nlm.nih.gov/33219329/>
51. AN Burnett-Hartman, JK Lee, J Demb, S Gupta. An Update on the Epidemiology, Molecular Characterization, Diagnosis, and Screening Strategies for Early-Onset Colorectal Cancer. *Gastroenterology* [Internet]. 2021 Mar 1 [cited 2021 Jul 29];160(4):1041–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/33417940/>
52. AN Willauer, Y Liu, AAL Pereira, M Lam, JS Morris, KPS Raghav, et al. Clinical and molecular characterization of early-onset colorectal cancer. *Cancer* [Internet]. 2019 Jun 15 [cited 2021 Jul 29];125(12):2002–10. Available from: <https://pubmed.ncbi.nlm.nih.gov/30854646/>
53. RL Elliott, GC Blobe. Role of transforming growth factor Beta in human cancer. *J Clin Oncol* [Internet]. 2005 [cited 2021 Jul 29];23(9):2078–93. Available from: <https://pubmed.ncbi.nlm.nih.gov/15774796/>
54. Venhuizen Jan Hendrik, Jacobs Femke JC, Span Paul N, Zegers Mirjam M. P120 and E-cadherin: Double-edged swords in tumor metastasis. *Semin Cancer Biol*. 2020;60(June 2019):107–20.
55. Duñach Mireia, Del Valle-Pérez Beatriz, García de Herreros Antonio. p120-catenin in canonical Wnt signaling. *Crit Rev Biochem Mol Biol* [Internet]. 2017;52(3):327–39. Available from: <http://dx.doi.org/10.1080/10409238.2017.1295920>
56. Kurley Sarah J, Tischler Verena, Brier Brian, Novitskiy Sergey V., Noske Aurelia, Varga Zsuzsanna, et al. A requirement for p120-catenin in the metastasis of invasive ductal breast cancer. *J Cell Sci*. 2021;134(6).
57. Kidacki Michal, Lehman Heather L, Green Michelle V., Warrick Joshua I, Stairs Douglas B. p120-catenin downregulation and PIK3CA mutations cooperate to induce invasion through MMP1 in HNSCC. *Mol Cancer Res*. 2017;15(10):1398–409.
58. Zhong C, Zuo Z, Ji Qing, Feng D. P120ctn may participate in epithelial-mesenchymal transition in OSCC. *Indian J Cancer* [Internet]. 2016 Jan 1 [cited 2022 Mar 2];53(1):20–4. Available from: <https://pubmed.ncbi.nlm.nih.gov/27146732/>
59. Jia Yanhan, Vong Joaquim Si Long, Asafova Alina, Garvalov Boyan K, Caputo Luca, Cordero Julio, et al. Lamin B1 loss promotes lung cancer development and metastasis by epigenetic derepression of RET. *J Exp Med*. 2019;216(6):1377–95.
60. Fu Fangfang, Wang Tian, Wu Zhangying, Feng Yourong, Wang Wenwen, Zhou Su, et al. HMGA1 exacerbates tumor growth through regulating the cell cycle and accelerates migration/invasion via targeting miR-221/222 in cervical cancer article. *Cell Death Dis* [Internet]. 2018;9(6). Available from: <http://dx.doi.org/10.1038/s41419-018-0683-x>
61. Kim Hyeonhui, Kim Minki, Im Sun-Kyoung, Fang Sungsoon. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res* [Internet]. 2018 [cited 2022 Apr 21];34(4):147. Available from: <https://pubmed.ncbi.nlm.nih.gov/30671100/>
62. Loesch Robin, Caruso Stefano, Paradis Valérie, Godard Cecile, Gougelet Angélique, Renault Gilles, et al. Deleting the β -catenin degradation domain in mouse hepatocytes drives hepatocellular carcinoma or hepatoblastoma-like tumor growth. *J Hepatol* [Internet]. 2022 Mar [cited 2022 Apr 21]; Available from: <https://pubmed.ncbi.nlm.nih.gov/35257829/>
63. Huang XD, Zheng YB, Yang YJ, Yang C, Li HL, Cheng HR. [Mouse models for human colorectal cancer with liver metastasis]. *Zhonghua Yi Xue Za Zhi* [Internet]. 2019 Sep 10 [cited 2022 Apr 21];99(34):2701–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/31505723/>

Artículo 2

Title: Telomere length as a new marker of risk of developing early-onset colorectal cancer.

Authors: Abel Martel-Martel^{1,2,3}, Spanish Early-onset Colorectal Cancer Consortium (SECOC), José Perea^{1,*}, Manuel A. Gómez-Marcos¹, Juan J. Cruz-Hernández^{1,2}, Rogelio González-Sarmiento^{1,2,4,*}

Affiliations: 1) Institute of Biomedical Research of Salamanca (IBSAL-SACYL-USAL-CSIC), Salamanca, Spain. 2) Molecular Medicine Unit, Department of Medicine, University of Salamanca, Salamanca, Spain. 3) Medical Oncology Service, University Hospital of Salamanca, Salamanca, Spain. 4) Institute of Molecular and Cellular Biology of Cancer (IBMCC), University of Salamanca-CSIC, Salamanca, Spain.

* Authors share co-senior authorship.

Estado: sometido a publicación

INTRODUCTION

Colorectal cancer (CRC) is the third most diagnosed tumor and the second leading cause of cancer death worldwide(1). Over the past few decades, there has been an alarming increase in early-onset CRC (EOCRC, age<50 years at diagnosis) incidence. Nowadays, EOCRC accounts for approximately 10% of the total CRC cases diagnosed each year and it is also being shown a significant rise in the associated mortality(2). The pathogenesis of EOCRC is well characterized among individuals with hereditary CRC. However, the remaining 80% does not carry a germline mutation associated with any cancer predisposition, thus we could defined as sporadic EOCRC(3,4). These sporadic early-onset tumors have worse prognosis than late-onset CRC tumors, being characterized by developing early metastasis and more aggressiveness features. Apart from this, early-onset tumors are diagnosed at very advanced stages because current screening programs do not consider this age subgroup in average-risk population, and only for high-risk individuals with family history of CRC this is reduced. In fact, the European Society of Gastrointestinal Endoscopy (ESGE) only recommends screening for CRC before 50 (beginning at 40 y/o) when criteria for familial CRC are fulfilled (at least two CRC in FDR, or at least one FDR with CRC younger than 50 years old)(5). Thus, the need of developing new markers that can be included in average-risk population CRC prevention strategies to detect individuals in risk for developing EOCRC is undeniable.

In this sense, telomerase activity has been postulated as a key factor in cancer development, in terms of the cellular immortalization through telomeric elongation, as well as aging and senescence through telomere shortening. Telomeres cap linear chromosomes to maintain stability and shorten with DNA replication during cell division. In healthy cells, erosion of telomere length (TL) eventually leads to regulated cell senescence and apoptosis. However, in abnormal cells, continued cell division results in telomeric shortening that can lead to end-to-end fusion of chromosomes and chromosomal instability. Therefore, telomere shortening is a process of aging associated with genetic instability and oncogenesis(6). Previous studies have shown that the depletion of constitutional telomere structure end sequences is associated with an increased risk for some cancers, such head and neck, urinary bladder, renal, lung, esophageal, and CRC. However, studies on CRC telomere dynamics in CRC have reported controversial results, with a recent study indicating that both long and short telomeres are associated with an increased risk of CRC(7). For this reason, we want to explore TL as a useful tool in CRC screening programs, particularly in the population under 50 years of age.

RESULTS

Telomeric shortening predisposes to the development of EOCRC.

First, we set out to analyze whether TL variability could be associated with the development of EOCRC. To this end, the absolute TL of leukocytes from 87 sporadic EOCRC patients and 109 healthy controls (HC) was quantified by RT-qPCR. Interestingly, we observed that TL was significantly shorter in the EOCRC group than in the control group (EOCRC mean: 122 kb vs HC mean: 296 kb; t-test, $p < 0.001$). Therefore, telomeric shortening could be associated with individuals with EOCRC, in comparison with healthy controls. Moreover, we studied in our cohort of patients the status of genes involved in the maintenance of TL (telosome/shelterin complex) by whole exome sequencing (WES). However, we did not find any relevant pathogenic mutation that could explain the telomeric shortening observed in our series.

***hTERT* rs2853669 polymorphism is associated with the risk of developing EOCRC.**

In addition, we studied whether the genotypic distribution of the polymorphism of the *hTERT* gene (rs2853669; A>G), related to telomeric shortening(8) and not reached by WES because its location at -245 bp from *hTERT* ATG site, confer a higher risk of developing EOCRC. In our cohort, the status of the minority allele (G) was associated with the risk of developing EOCRC (p -value=0.028; OR: 1.326 (95% IC: 1.018-1.728)) (Table 1). In this case, we observed that G allele was less frequent in EOCRC than in control population (26% EOCRC vs 36% in control group). These data suggest that the minority allele may play a protective role against the development of EOCRC. Therefore, the loss of this protective factor would increase the risk of developing EOCRC through telomerase dysfunction.

Comparative analysis focused on clinical and familial features.

When we compared telomere lengths with the different clinico-pathological and familial features, gender and BMI showed statistical differences (data not shown): shorter telomeres were associated with males, as well as with higher BMI. Regarding the comparison between all the features and the different forms of *hTERT* polymorphism, no association showed statistical significance. No analysis was carried out focused on ESGE screening criteria in our cohort as none of the cases fulfilled the need of reducing the age of screening.

Table 1. *hTERT* rs2853669 SNP showing statistical differences between patients with sporadic EOCRC and controls.

SNP*	Genotype	Controls	Sporadic EOCRC	P-value	OR* (95% CI)
<i>hTERT</i> (rs2853669)	AA	43 (39%)	47 (54%)	0.082	
	AG	53 (49%)	35 (40%)		
	GG	13 (12%)	5 (6%)		
	AA+AG	96 (88%)	82 (94%)	0.137	
	GG	13 (12%)	5 (6%)		
	GG+GA	66 (61%)	40 (46%)	0.042	1.803 (1.020-3.190)
	AA	43 (39%)	47 (54%)		
	A	139 (64%)	129 (74%)	0.028	1.326 (1.018-1.728)
	G	79 (36%)	45 (26%)		

* SNP: Single Nucleotide Polymorphism, OR: Odd Ratio.

DISCUSSION

A study on germline telomere shortening by Lisa Boardman *et al.* showed that longer telomeres increased the risk of developing EOCRC, whereas shorter telomeres were associated with the development of CRC in patients older than 50 years(9). Although these findings contrast with our results, we should note that our case/control study is only focused on the population younger than 50 years, since the main objective is to establish the analysis of TL as a predictive marker for risk of develop EOCRC. In addition, leukocyte telomeres systematically shorten over the years and have been positioned as a clear factor of cellular aging, which could explain the greater shortening observed by Lisa's group in LOCRC patients. Patients older than 50 years could have a significant telomeric shortening due to age that leads to telomere crisis and subsequent chromosomal instability, enabling cells to advance toward malignancy. Therefore, if we consider cancer as a disease of the older adult, the existence of a germline telomeric shortening in people under 50, could generate a premature cellular aging situation that increases the risk of developing EOCRC.

In addition, the rs2853669 polymorphism of the *hTERT* gene (A>G) has been associated with the development of different types of tumors such as gliomas, melanoma, breast cancer, hepatocellular carcinoma, and lung cancer(10). In these cases, there was an association between GG genotype and tumor development through a gain of telomerase function that promotes cell immortalization by telomeric elongation. In our cohort, the lower frequency of the G allele would be related to a loss of telomerase activity leading to telomeric shortening and, therefore, to premature cellular aging that predisposes to the development of EOCRC.

In conclusion, we consider the measurement of germline TL at early ages as a non-invasive method that could facilitate the early identification of individuals in risk of developing EOCRC, but without any familial criteria that could lead in an advance of the starting age of the CRC screening. Thus, this could justify the need of including TL as a new marker in CRC prevention strategies, after consequent validation cohort in other young populations.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-249.
2. Sinicrope FA. Increasing Incidence of Early-Onset Colorectal Cancer. Longo DL, ed. *N Engl J Med* 2022;386:1547-1558.
3. Perea J, Rodríguez Y, Rueda D, et al. Early-onset colorectal cancer is an easy and effective tool to identify retrospectively Lynch syndrome. *Ann Surg Oncol* 2011;18:3285-3291.
4. Mauri G, Sartore-Bianchi A, Russo AG, et al. Early-onset colorectal cancer in young individuals. *Mol Oncol* 2019;13:109-131.
5. van Leerdam ME, Roos VH, van Hooft JE, et al. Endoscopic management of Lynch syndrome and of familial risk of colorectal cancer: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2019;51:1082-1093.
6. Vertecchi E, Rizzo A, Salvati E. Telomere Targeting Approaches in Cancer: Beyond Length Maintenance. *Int J Mol Sci* 2022;23:3784.
7. Cui Y, Cai Q, Qu S, et al. Association of leukocyte telomere length with colorectal cancer risk: nested case-control findings from the Shanghai Women's Health Study. *Cancer Epidemiol Biomarkers Prev* 2012;21:1807-1813.
8. Goswami A, Huda N, Yasmin T, et al. Association study of leukocyte telomere length and genetic polymorphism within hTERT promoter with type 2 diabetes in Bangladeshi population. *Mol Biol Rep* 2021;48:285-295.
9. Boardman LA, Litzelman K, Seo S, et al. The association of telomere length with colorectal cancer differs by the age of cancer onset. *Clin Transl Gastroenterol* 2014;5.
10. Shen N, Lu Y, Wang X, et al. Association between rs2853669 in TERT gene and the risk and prognosis of human cancer: a systematic review and meta-analysis. *Oncotarget* 2017;8:50864-50872.
11. Perea J, Marti M, Espin E, et al. Cohort profile: the Spanish Early-onset Colorectal Cancer (SECOC) cohort: a multicentre cohort study on the molecular basis of colorectal cancer among young individuals in Spain. *BMJ Open* 2021;11:e055409.

Acknowledgments:

This study has been funded by Instituto de Salud Carlos III (ISCIII) through the projects PI20/0974 and PI20/01569, and co-funded by the European Union.

Conflicts of interest statement:

The authors have declared that no conflict of interest exists.

Author contributions:

Study concept and design, AMM, JP, RGS; acquisition of data, SECOC; analysis and interpretation of data, AMM; drafting of the manuscript, all authors; critical revision of the manuscript for important intellectual content, JP, RGS; statistical analysis, AMM; obtained funding, JP, RGS; technical, or material support, AMM; study supervision, JP, RGS.

Contributors:

Spanish EOCRC (SECOC) consortium. Members and Affiliations:

Belvitge University Hospital; Bellvitge Research Institute (IDIBELL), Barcelona: Principal Investigator: Cristina Santos. Associated Investigators: Mercedes Martínez Villacampa, Víctor Moreno, Jose Carlos Ruffinelli. **Complutense University, School of Medicine, Madrid, Biostatistics Department:** Principal Investigator: Lucía Inglada-Pérez. **Fundación Alcorcón Hospital, Madrid:** Principal Investigator: Sirio Melone. Associated Investigator: José A. Rueda Orgaz. **Fundación Jiménez Díaz University Hospital, Madrid:** Principal Investigator: Damián García-Olmo. Associated Investigators: Víctor Castellano Megías, Sergio Hernández-Villafranca, Manuel Escanciano. **Galdakao-Usansolo Hospital, Vizcaya:** Principal Investigator: Fernando Jiménez. Associated Investigators: Adriana Cavero, Vicente Portugal. **Gregorio Marañón University Hospital, Madrid:** Principal Investigator: Elena Hurtado Caballero. Associated Investigators: María Arriba Domenech, Luis M. Jiménez Gómez, Isabel Peligros Gómez, Cristina Rey Valcárcel, Jaime Zorrilla Ortúzar. **Department of Gastroenterology, Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain:** Principal Investigator: Francesc Balaguer. Associated Investigators: María Daca, Teresa Ocaña, Miriam Cuatrecasas, Sabela Carballal, Leticia Moreira, Lorena Moreno, María Pellisé, Ariadna Sánchez. **Hospital del Mar, Barcelona:** Principal Investigator: Marta Jiménez Toscano. Associated Investigators: Mar Iglesias Coma. **Molecular Oncology IMDEA Food Institute, Madrid:** Principal Investigator: Ana Ramírez de Molina. Associated Investigators: Gonzalo Colmenarejo, Isabel Espinosa-Salinas, Lara P Fernández, Marta Gómez de Cedrón. **Biomedical Research Institute of Salamanca**

(IBSAL): Principal Investigator: José Perea. Associated Investigators: Luis Corchete; Juan L. García; Paula García Vallés; Rogelio González-Sarmiento, Ana B. Hernández; Abel J. Martel Martel, Jéssica Pérez.

Infanta Leonor University Hospital, Madrid: Principal Investigator: Edurne Álvaro. Associated Investigators: Ana Burdaspal, Maria Luisa de Fuenmayor. **La Paz University Hospital, Madrid:** Principal Investigator: Isabel Prieto (Surgery Department). Associated Investigators: Alejandro Forero, Inés Rubio.

León University Hospital, León: Principal investigator: Jesús Fernández Fueyo (Surgery Department). Associated Investigators: Enrique Pastor, Amaya Villafañe. **MD Anderson Spain, Madrid:** Principal Investigator: Irene López Rojo. Associated Investigators: Oscar Alonso, Sara Encinas, Ana Teijo. **Navarra University Clinic, Madrid:** Principal Investigator: Carlos Pastor. Associated Investigators: Jorge Arredondo, Jorge Baixauli Fons, Lucia Cenicerós Paredes, Javier Rodríguez Rodríguez Carlos Sánchez Justicia.

Ramón y Cajal University Hospital: Principal Investigator: Araceli Ballesteró Pérez. Associated Investigators: Javier Die Trill, José M^a Fernández Cebrián, Juan Ocaña Jiménez. **San Carlos University Hospital, Madrid:** Principal Investigator: Gonzalo Sanz (Surgery Department). Associated Investigators: Jana Dziakova, Sara Picazo Marín, Rodrigo Sanz López, María Suárez Solís. **Salamanca University Hospital:** Principal Investigator: José A. Alcazar. Associated Investigators: Jacinto García, Rosario Vidal Tocino. **Spanish National Cancer Research Center. Clinical Cancer Unit. Human Genetics Group:** Principal Investigator: Miguel Urioste. **Spanish National Cancer Research Center. Genetic & Molecular Epidemiology Group:** Principal Investigator: Núria Malats. Associated Investigators: Lidia Estudillo.

Secugen, Madrid: Principal Investigator: Julián Pérez-Pérez. **Vall d'Hebron University Hospital, Barcelona:** Principal Investigator: Marc Martí Gallostra. Associated Investigators: Eloy Espín, Franco Marinello, Miquel Kraft, Stefania Landolfi. **Villalba University Hospital, Madrid:** Principal Investigator: Lorena Brandáriz. Associated Investigators: Ignacio Valverde. **12 de Octubre University Hospital, Madrid:** Principal Investigator: Alfredo Vivas. Associated Investigators: Cristina Narváez.

SUPPLEMENTARY METHODS**Patient's selection:**

For this study, we firstly selected a total of 96 consecutive cases of EOCRC, from which cases with pathogenic mutations in any of the CRC susceptibility genes were excluded, leaving the definitive study sample in 87 EOCRC cases. They were selected from the prospective multicenter study defined as the Spanish Early-Onset Colorectal Cancer Consortium (SECOC)(11). All patients were diagnosed before age 50 years, without history of inflammatory bowel disease and with a histopathological diagnosis of adenocarcinoma. Clinicopathological features were determined from a detailed review of the medical records. Variables included gender, age at CRC diagnosis, Body Mass Index (BMI) at diagnosis, tumor stage at diagnosis, tumor location, histological features (grade of differentiation, mucinous component, "signet ring" cells), multiple primary neoplasms (synchronous and metachronous CRC) and familial cancer history (Supplementary Table 1). Peripheral blood (PB) samples from all patients included in the study were collected before any type of neoadjuvant (chemotherapy or radiotherapy), adjuvant or surgical treatment.

DNA isolation, genotyping, and telomere length analysis by Real Time Quantitative PCR (RT-qPCR):

DNA was obtained from leukocytes PB tube using the phenol-chloroform method. All DNA samples were stored in Eppendorf tubes at 20°C to prevent its progressive degradation and potential contamination.

The rs2853669 single nucleotide polymorphism (SNP) of *hTERT* gene was selected according to the following criteria: literature association with shortened telomeres, and >5% minor allele frequency in Caucasian population. Genotyping was performed by Real-Time Quantitative PCR (RT-qPCR) using the TaqMan® allelic discrimination assay (Applied Biosystems, Foster, CA), including specific oligonucleotides to amplify the SNP sequences and two labelled probes with the fluorochrome VIC and FAM to detect both alleles of each SNP. For that, 40 ng of DNA sample were added to 5 µL of TaqMan® Universal PCR Master Mix and combined with the specific forward and reverse primers, and the VIC/FAM labelled probes.

The leukocytes telomere length (TL) from each sample was measured using RT-qPCR using the Absolute Human Telomere Length Quantification qPCR Assay Kit (ScienCell, Catalog #8918, CA, USA). This technique allows the initial amount of DNA coding for telomerase (TEL) and a single copy reference gene (used as an endogenous control) to be quantified simultaneously. The difference in the amount of DNA quantified represents the relative TL of each sample. To analyze these relative changes, a reference fragment of known TL (provided by the manufacturer) was added to each assay, allowing the absolute

quantification of the TL of each sample, for a diploid cell and/or chromosome end to be determined. Triplicate reactions were carried to minimize variability. The TEL and SCR fragments were amplified using 10 ng in 2 μ L of DNA, 1 μ L of each specific primer and 10 μ L of the FastStart SYBR Green Master Mix. The amplification program was as follows: 10 min at 95 °C, followed by 40 cycles at 95 °C for 15 s, 52 °C for 30 s and 60 °C for 1 min. Finally, the Ct ($2^{-\Delta\Delta C_t}$) comparative method was used to calculate the relative DNA amount of each amplicon.

These assays were performed in a 96 well plate and the detection was carried out in the Step-One Plus Real-Time PCR system (Applied Biosystems).

Statistical analyses:

Categorical variables were expressed as number of cases and percentage and were compared using Pearson's Chi Square (χ^2) test. Comparing continuous with categorical variable, Student's t-Test was used for independent samples and the Mann–Whitney U test was used for continuous variables (all of which were expressed as mean values plus/minus standard deviation (SD)). Analysis of variance (ANOVA) was also used to compare more than two groups following normal distributions and the Kruskal–Wallis test for those following nonparametric distributions. SPSS version 23.0 (IBM) was used for statistical analyses. A p-Value < 0.05 was established to consider differences as statistically significant.

Supplementary Table 1. Baseline data from Early-onset colorectal cancer cases.

<i>Characteristic</i>	Total (%)
Total	87
Sex	
Female	36 (41)
Male	51 (59)
Age at Diagnosis	
<30 years	4 (5)
30-39 years	15 (17)
40-49 years	68 (78)
Mean (std), years	42.7 (6.2)
Median (range), years	44 (23-49)
Body Mass Index (BMI)	
Underweight (<18.5 kg/m ²)	2 (2)
Normoweight (18.5-<25 kg/m ²)	33 (38)
Overweight (25-<30 kg/m ²)	24 (28)
Obese (30+ kg/m ²)	11 (13)
Unknown	17 (19)
Mean (std), kg/m ²	25.15 (5.4)
Median (range), kg/m ² 3	24.9 (15.6-43)
Tumor Stage at diagnosis	
I	12 (14)
II	17 (20)
III	30 (34)
IV	20 (23)
Unknown	8 (9)
Tumor Site	
Right colon	20 (23)
Left colon	27 (31)
Rectosigmoid junction/Rectum	40 (46)
Histological features*	
Mucinous	13/59 (22)
"Signet ring" cells	4/59 (7)
Unknown	18 (21)
Grade of differentiation*	
Low	9 (10)
Medium	23 (26)
High	27 (31)
Unknown	18 (21)
Multiple Primary Neoplasms	
Synchronous colorectal cancer	3 (3)
Other Multiple Primary Neoplasms	8 (9)
Familal Cancer History	
Colorectal Cancer	19 (22)
Unknown	8 (9)
Sporadic cases	
Yes	39 (45)
No	43 (49)
Unknown	5 (6)

* The remaining cases are those with only biopsy due to palliative conditions.