



ESCUELA DE DOCTORADO  
INTERNACIONAL DE LA USC

Anael  
López Novo

Tesis doctoral

Integración de ómicas para la  
identificación de nuevos genes  
de susceptibilidad hereditaria a  
cáncer colorrectal

Santiago de Compostela, 2023

Programa de doctorado en Medicina Molecular



TESIS DE DOCTORADO

# **INTEGRACIÓN DE ÓMICAS PARA LA IDENTIFICACIÓN DE NUEVOS GENES DE SUSCEPTIBILIDAD HEREDITARIA A CÁNCER COLORRECTAL**

Anael López Novo

ESCUELA DE DOCTORADO INTERNACIONAL DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOCTORADO EN MEDICINA MOLECULAR



SANTIAGO DE COMPOSTELA

2023



D./Dña. **Anael López Novo**

Título de la tesis: **Integración de ómicas para la identificación de nuevos genes de susceptibilidad hereditaria a cáncer colorrectal**

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento y declaro que:

- 1) La tesis abarca los resultados de la elaboración de mi trabajo.
- 2) De ser el caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.
- 3) Confirmando que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.
- 4) La tesis es la versión definitiva presentada para su defensa y coincide la versión impresa con la presentada en formato electrónico.

Y me comprometo a presentar el Compromiso Documental de Supervisión en el caso que el original no esté depositado en la Escuela.

En **Santiago de Compostela, 14 de julio de 2023.**

**Firma electrónica**





Yo, Anael López Novo, con DNI 79340681L, declaro no tener conflictos de interés en relación a la presente tesis doctoral.

En Santiago de Compostela, 14 julio de 2023

Fdo. Anael López Novo



D./Dña. **Clara Ruiz Ponte**

En condición de: **Director/a**

Título de la tesis: **Integración de ómicas para la identificación de nuevos genes de susceptibilidad hereditaria a cáncer colorrectal**

INFORMA:

Que la presente tesis, se corresponde con el trabajo realizado por D/Dña **Anael López Novo**, bajo mi dirección/tutorización, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director/tutor de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

En **Santiago de Compostela, 14 de julio de 2023**

**Firma electrónica**



D./Dña. **Ángel María Carracedo Álvarez**

En condición de: **Tutor/a y director/a**

Título de la tesis: **Integración de ómicas para la identificación de nuevos genes de susceptibilidad hereditaria a cáncer colorrectal**

INFORMA:

Que la presente tesis, se corresponde con el trabajo realizado por D/Dña **Anael López Novo**, bajo mi dirección/tutorización, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director/tutor de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

En **Santiago de Compostela, 14 de julio de 2023**

**Firma electrónica**



Este trabajo ha sido financiado por el Instituto de Salud Carlos III – ISCIII – PI17/00509 (IP Clara Ruiz Ponte) – Cofinanciado por la Unión Europea – FEDER.



La doctoranda ha sido financiada por el Programa de ayudas a la etapa predoctoral 2018 (Axencia Galega de Innovación, Xunta de Galicia).







*“Argue for your limitations, and sure enough, they're yours”*

Richard Bach, «Illusions: The adventures of a Reluctant Messiah»  
(1977)



## AGRADECIMIENTOS

Todo el que me conoce sabe que no acostumbro a demostrar agradecimiento y cariño con las palabras, y prefiero los hechos. Y aunque estoy haciendo un gran esfuerzo para poder plasmar todo lo que tengo en mi cabeza, mientras escribo estas líneas siento que se quedarán escasas. Pero vamos allá, intentaré hacerlo lo mejor que sé.

Esta tesis doctoral es el resultado de la suma de muchas personas con las que me he cruzado a lo largo de estos años, y que de una u otra manera han participado en este trabajo.

En primer lugar, me gustaría agradecer a mis directores de tesis, Clara y Ángel, por darme la oportunidad de realizar este trabajo y la independencia para equivocarme un millón de veces y aprender de ello.

Gracias a los pacientes participantes en este proyecto, que en un momento tan delicado y de manera altruista han cedido sus muestras en favor de la ciencia y han hecho posible esta tesis. Espero de todo corazón que estéis bien.

Gracias al personal de los Servicios de Endoscopias Digestivas del Hospital Universitario Lucus Augusti, Complejo Hospitalario Universitario de Pontevedra y Complejo Hospitalario Universitario de Ourense, los Servicios de Cirugía y Digestivo del Complejo Hospitalario Universitario de Santiago de Compostela y Hospital Clínic de Barcelona, y los Servicios de Anatomía Patológica del Complejo Hospitalario Universitario de Santiago de Compostela, Complejo Hospitalario Universitario de Pontevedra y Hospital Clínic de Barcelona, que han colaborado en el reclutamiento de pacientes y toma de muestras y han hecho posible esta tesis.

Gracias al CNAG-CRG por su ayuda en la secuenciación de exoma completo y RNA-seq. Al centro encargado del servicio de genotipado, CEGEN-PRB3-ISCI3; subvencionado por PT17/00019, del PE I+D+i 2013-2016, financiado por ISCI3 y FEDER. Al Centro de Supercomputación de Galicia (CESGA) por permitir el uso de los servidores de FinisTerra II y FinisTerra III para llevar a cabo parte de los análisis que componen este trabajo. Los resultados aquí mostrados están basados en su totalidad o en parte en datos generados

por The Cancer Genome Atlas (TCGA) bajo la versión de dbGAP phs000178.v11.p8. TCGA está administrado por el NCI y NHGRI. La información sobre el TCGA puede encontrarse en <http://cancergenome.nih.gov>. El Proyecto Genotype-Tissue Expression está financiado por el Common Fund of the Office Director of the National Institutes of Health, y por NCI, NHGRI, NHLBI, NIDA, NIMH, y NINDS. Los datos usados para los análisis descritos en esta Tesis Doctoral fueron obtenidos de dbGaP accession number phs000424.v8.p2

Gracias a mis amigas. A Rita y María Cebro, por acogerme desde el primer momento con los brazos abiertos y haber sido casa en Santiago. Por compartir llantos, risas, alegrías y victorias, y porque sé que la finalización de esta tesis la recibís con los mismos nervios y las mismas ganas que yo. A Carla, la mejor compañera de piso que se puede tener, gracias por aguantarme y cuidarme. A Paula, quién sin saberlo, con su pasión por la genética y sus ganas constantes de aprender ha sido y es una inspiración. A Marinela, por su energía explosiva y contagiosa tan necesaria en los días malos, y por haber cambiado el café por el té (por esto último estoy muy agradecida). A María Regueira, que es una fuerza de la naturaleza, por repartir amor y alegría por donde pasa. A las de siempre, las de Cariño, por respetar mis ausencias y mis silencios sin reproches, solo deciros que ya vuelvo. En resumen, gracias a todas por ser presente, pero también porque seréis futuro.

Gracias a todas las personas que forman parte, o lo han hecho en estos años, de la Fundación Pública Galega de Medicina Xenómica por haberme tratado tan bien y haberme enseñado tanto. A los de la sala del café, Marga Formoso, Bego, Rocío, Sonia y Miguel, porque me habéis sacado muchas risas y sonrisas. A los “predocs”, los que lo son ahora y los que lo fueron, porque pocas cosas unen más que compartir esta etapa. A Jesús, quien me dijo hace unos años que sería un honor para él formar parte de estas líneas y la primera persona en la que pensé al comenzar a escribirlas; por ser un gran compañero, un investigador muy válido y una persona maravillosa. A Míriam, mi primera compañera aquí, por llevarme de la mano en mis primeros pasos. A Jorge, Juan y Ceres por las lecciones de bioinformática y los debates

científicos. A Silvia, por ser mi compañera de bata y una ayuda valiosísima. De verdad ha sido un placer haber coincidido y trabajado con todas vosotras y vosotros.

Y porque lo importante es siempre como se acaba, GRACIAS a mi familia. A mis padres, que con su esfuerzo y cariño han conseguido que hoy esté aquí. Gracias por darme siempre la libertad de escoger mi camino y acompañarme en él. A mi hermano, no podría haber tenido más suerte contigo, gracias por recordarme que el esfuerzo siempre merece la pena. A Timón, mi fiel compañero. Gracias a mis abuelos Jaime y Feli por las palabras de ánimo constantes. Y a mi abuela Carmen, que espero esté orgullosa.



## ABREVIATURAS

|                        |  |
|------------------------|--|
| <b>AFAP</b>            | Poliposis adenomatosa familiar atenuada (del inglés, <i>attenuated familial adenomatous polyposis</i> )                                      |
| <b>BER</b>             | Sistema de reparación por escisión de bases (del inglés, <i>base excision repair</i> )   |
| <b>CCR</b>             | Cáncer colorrectal   |
| <b>CESGA</b>           | Centro de Supercomputación de Galicia  |
| <b>CIMP</b>            | Fenotipo metilador de islas CpG (del inglés, <i>CpG island methylator phenotype</i> )  |
| <b>CIN</b>             | Inestabilidad cromosómica (del inglés, <i>chromosomal instability</i> )  |
| <b>CMMRD</b>           | Síndrome de Deficiencia Constitucional de Reparación de Errores de Apareamiento (del inglés, <i>constitutional MMR-deficiency syndrome</i> ) |
| <b>CMS</b>             | Subgrupos Moleculares Consenso (del inglés, <i>Consensus Molecular Subtypes</i> )  |
| <b>CNAG-CRG</b>        | Centro Nacional de Análisis Genómicos, Barcelona   |
| <b>COSMIC</b>          | <i>Catalogue Of Somatic Mutations In Cancer</i>  |
| <b>DBS</b>             | Firmas asociadas a sustituciones dinucleotídicas (del inglés, <i>doublet base substitutions</i> )  |
| <b>DMP</b>             | Posiciones diferencialmente metiladas (del inglés, <i>differentially methylated positions</i> )  |
| <b>DRAGEN</b>          | <i>Dynamic Read Analysis for Genomics</i>  |
| <b>ED</b>              | Dominio exonucleasa (del inglés, <i>exonuclease domain</i> )   |
| <b>EMT</b>             | Transición epitelio-mesenquimal (del inglés, <i>epithelial-mesenchymal transition</i> )  |
| <b>FAP</b>             | Poliposis adenomatosa familiar (del inglés, <i>familial adenomatous polyposis</i> )  |
| <b>GATK</b>            | <i>Genome Analysis Toolkit</i>   |
| <b>GSA</b>             | Análisis de conjunto de genes (del inglés, <i>gene set analysis</i> )  |
| <b>GTE<sub>x</sub></b> | <i>The Genotype-Tissue Expression</i>  |



|                               |  |
|-------------------------------|--|
| <b>GVCF</b>                   | <i>Genomic Variant Call Format</i>   |
| <b>ID</b>                     | Firmas asociadas a inserciones y deleciones de pequeño tamaño (del inglés, <i>small insertions and deletions</i> ) |
| <b>IGV</b>                    | <i>Integrative Genomics Viewer</i>   |
| <b>KEGG</b>                   | <i>Kyoto Encyclopedia of Genes and Genomes</i>   |
| <b>Mb</b>                     | Megabase   |
| <b>MMR</b>                    | Genes reparadores de errores de emparejamiento de bases (del inglés, <i>MisMatch Repair</i> )                      |
| <b>MSI</b>                    | Inestabilidad de microsatélites (del inglés, <i>microsatellite Instability</i> )                                   |
| <b>MSS</b>                    | Estabilidad de microsatélites (del inglés, <i>microsatellite stable</i> )  |
| <b>NGS</b>                    | Secuenciación masiva (del inglés, <i>Next-Generation Sequencing</i> )  |
| <b>pb</b>                     | Pares de bases   |
| <b>SBBR</b>                   | Síndrome de Bannayan-Riley-Ruvalcaba (del inglés, <i>Bannayan Riley-Ruvalcaba syndrome</i> )                       |
| <b>SBS</b>                    | Firmas asociadas a variantes de nucleótido único (del inglés, <i>single base substitutions</i> )                   |
| <b>SC</b>                     | Síndrome de Cowden   |
| <b>SNPs</b>                   | <i>Single nucleotide polymorphisms</i>   |
| <b>TCGA</b>                   | <i>The Cancer Genome Atlas</i>   |
| <b>TGF-<math>\beta</math></b> | Factor de crecimiento transformante- $\beta$ (del inglés, <i>transforming growth factor- <math>\beta</math></i> )  |
| <b>TNF</b>                    | Factor de necrosis tumoral (del inglés, <i>tumor necrosis factor</i> )   |
| <b>TSG</b>                    | Gen supresor de tumores (del inglés, <i>tumor suppressor gene</i> )  |
| <b>VCF</b>                    | <i>Variant Call Format</i>   |
| <b>WES</b>                    | Secuenciación de exoma complete (del inglés, <i>Whole-exome sequencing</i> )                                       |

# ÍNDICE

|            |   |           |
|------------|---|-----------|
| <b>1</b>   | <b>Resumen</b> .....  | <b>25</b> |
| <b>2</b>   | <b>INTRODUCCIÓN</b> .....   | <b>29</b> |
| <b>2.1</b> | <b>Epidemiología</b> .....  | <b>29</b> |
| <b>2.2</b> | <b>Tumorigénesis del cáncer colorrectal</b> .....                                     | <b>30</b> |
| 2.2.1      | Rutas de inestabilidad genómica .....   | 30        |
| 2.2.2      | Firmas mutacionales .....   | 32        |
| 2.2.3      | Subgrupos Moleculares Consenso.....   | 33        |
| <b>2.3</b> | <b>Arquitectura genética del cáncer colorrectal</b> .....                             | <b>36</b> |
| 2.3.1      | Síndromes de predisposición hereditaria a CCR .....                                   | 38        |
| 2.3.2      | Variantes de moderada penetrancia .....   | 44        |
| 2.3.3      | Variantes de baja penetrancia y estudios de asociación ....                           | 45        |
| <b>2.4</b> | <b>Genes candidatos de susceptibilidad hereditaria a CCR</b> ....                     | <b>46</b> |
| <b>3</b>   | <b>JUSTIFICACIÓN DEL ESTUDIO</b> .....  | <b>51</b> |
| <b>4</b>   | <b>OBJETIVOS</b> .....  | <b>55</b> |
| <b>5</b>   | <b>MATERIAL Y MÉTODOS</b> .....   | <b>59</b> |
| <b>5.1</b> | <b>Cohortes de estudio en la fase de descubrimiento</b> .....                         | <b>59</b> |
| 5.1.1      | Cohorte TOGETHER .....  | 59        |
| 5.1.2      | Cohorte de la base de datos <i>The Cancer Genome Atlas</i> (TCGA).....                | 61        |
| 5.1.3      | Cohorte de la base de datos <i>The Genotype-Tissue Expression</i> (GTEx).....         | 62        |
| <b>5.2</b> | <b>Secuenciación de exoma completo y procesamiento de datos</b> .....                 | <b>62</b> |
| 5.2.1      | Secuenciación de exoma completo de la cohorte TOGETHER .....                          | 62        |
| 5.2.2      | <i>Calling</i> y anotación de variantes: cohortes TOGETHER y TCGA (COAD y READ) ..... | 62        |

|   |            |
|---|------------|
| <b>5.3 RNA-seq y procesamiento de datos: cohortes TOGETHER, TCGA y GTEx.....</b>              | <b>65</b>  |
| <b>5.4 Endofenotipado tumoral.....</b>  | <b>66</b>  |
| 5.4.1 Firma mutacional tumoral.....   | 66         |
| 5.4.2 Carga mutacional tumoral.....   | 67         |
| 5.4.3 Identificación de genes driver en CCR.....  | 67         |
| 5.4.4 Análisis de mucosa colónica normal y tumoral.....                                       | 67         |
| 5.4.5 Subgrupos Moleculares Consenso o CMS.....   | 68         |
| 5.4.6 Perfil epigenético de los CMS.....  | 69         |
| <b>5.5 Integración de ómicas.....</b>   | <b>70</b>  |
| 5.5.1 Filtrado de variantes germinales.....   | 70         |
| 5.5.2 Aproximación <i>All vs One</i> .....  | 72         |
| <b>5.6 Cohorte de réplica.....</b>  | <b>72</b>  |
| <b>5.7 Aspectos éticos.....</b>   | <b>75</b>  |
| <b>6 RESULTADOS.....</b>  | <b>79</b>  |
| <b>6.1 Endofenotipado tumoral: caracterización molecular.....</b>                             | <b>79</b>  |
| 6.1.1 Firma y carga mutacional tumoral.....   | 79         |
| 6.1.2 Subgrupos Moleculares Consenso o CMS.....   | 83         |
| 6.1.3 Perfil epigenético de los CMS.....  | 88         |
| <b>6.2 Alteraciones somáticas en genes driver.....</b>  | <b>95</b>  |
| <b>6.3 Integración de ómicas: filtrado de variantes y aproximación <i>all vs one</i>.....</b> | <b>98</b>  |
| 6.3.1 Subgrupo Molecular Consenso 1 (CMS1).....   | 98         |
| 6.3.2 Subgrupo Molecular Consenso 2 (CMS2).....   | 99         |
| 6.3.3 Subgrupo Molecular Consenso 3 (CMS3).....   | 102        |
| 6.3.4 Subgrupo Molecular Consenso 4 (CMS4).....   | 104        |
| 6.3.5 Tumores no incluidos en Subgrupos Moleculares Consenso.....                             | 107        |
| <b>6.4 Cohorte de réplica.....</b>  | <b>109</b> |
| <b>7 DISCUSIÓN.....</b>   | <b>115</b> |
| <b>8 CONCLUSIONES.....</b>  | <b>141</b> |

|                                       |            |
|---------------------------------------|------------|
| <b>9 BIBLIOGRAFÍA .....</b>           | <b>145</b> |
| <b>10 MATERIAL SUPLEMENTARIO.....</b> | <b>167</b> |
| <b>11 ANEXOS .....</b>                | <b>257</b> |



# 1 RESUMEN

En los últimos años las tecnologías NGS (*Next-Generation Sequencing*) generaron muchas expectativas en cuanto a la identificación de variación genética que explicara la heredabilidad perdida en cáncer colorrectal (CCR). Sin embargo, no solo se identificó un número muy reducido de genes candidatos con fuerte evidencia científica, sino que se obtuvo un gran número de genes cuyo papel en la tumorigénesis colorrectal todavía no está bien establecido.

La finalidad de la presente Tesis Doctoral es descubrir nuevos genes de susceptibilidad a CCR, mediante la identificación de variantes germinales de riesgo a través de estudios de funcionalidad integrando los genomas germinal y de cáncer colorrectal de pacientes enriquecidos genéticamente por un diagnóstico temprano de la enfermedad.

La identificación de nuevos genes de susceptibilidad contribuirá a explicar parte del riesgo heredado a desarrollar CCR, aumentando el conocimiento sobre las bases hereditarias de esta enfermedad.



# INTRODUCCIÓN





## 2 INTRODUCCIÓN

### 2.1 EPIDEMIOLOGÍA

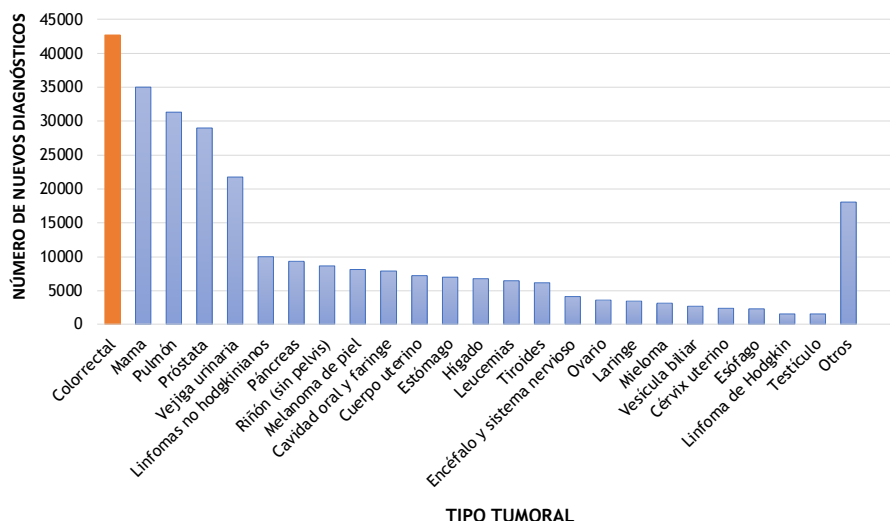
El cáncer colorrectal (CCR) es un conjunto de neoplasias que tienen lugar en el intestino grueso. A nivel mundial se trata de la tercera neoplasia con mayor incidencia y la segunda con mayor mortalidad en ambos sexos<sup>1</sup>, observándose en los países desarrollados una incidencia cuatro veces mayor con respecto a los países en vías de desarrollo<sup>2</sup>.

Así mismo, en España el CCR también ocupa los primeros puestos en número de nuevos diagnósticos anuales y mortalidad, estimándose que este año 2023 será la neoplasia más frecuentemente diagnosticada en nuestro país con 42.721 nuevos casos<sup>3</sup> (Figura 1). En nuestro país, la tasa de incidencia de CCR estandarizada por edad por cada 100.000 habitantes ha aumentado 16,14 y 4 puntos entre los años 1993 y 2010 en hombres y mujeres, respectivamente<sup>1</sup>. Por otro lado, la tasa de mortalidad estandarizada por edad por cada 100.000 habitantes de esta neoplasia en el mismo periodo de tiempo ha aumentado 2,05 puntos en hombres, mientras que en mujeres ha disminuido 0,80 puntos; sin embargo, en los últimos años se ha observado un descenso en el número de defunciones ocasionadas por estos tumores<sup>1</sup>.

Los últimos informes a nivel global indican un descenso en la incidencia de las neoplasias de colon y recto en individuos con edades iguales o superiores a los 50 años en países que presentan índices elevados de esta enfermedad<sup>2</sup>. Este descenso en la incidencia es debido a la existencia de programas de cribado poblacional a partir de los 50 años en estos países<sup>2</sup>. Por el contrario, se observa un aumento del número de nuevos casos diagnosticados en personas con edades inferiores a los 50 años, llegando en algunos países a presentar un incremento de la incidencia de hasta el 4% cada año<sup>2</sup>.

Dado que el mayor número de casos de CCR se localiza en países desarrollados, este podría considerarse un indicativo de la calidad de vida de los individuos de una región, pues estos países muestran mayor desarrollo socioeconómico<sup>2</sup>. Se asocian a la aparición de CCR factores

poblacionales (aumento y envejecimiento poblacional), estilo de vida (sedentarismo y tabaquismo, entre otros), dieta (aumento del consumo de alimentos de origen animal) y presencia de programas de cribado poblacional<sup>2,3</sup>.



**Figura 1. Estimación para ambos sexos del número de casos de cáncer en España para el año 2023 (excluidos los cánceres cutáneos no melanoma).** Las barras verticales representan el número de nuevos casos teniendo en cuenta ambos sexos para cada tipo tumoral. Figura de elaboración propia empleando datos procedentes de Las cifras del cáncer en España 2023<sup>3</sup>.

## 2.2 TUMOROGÉNESIS DEL CÁNCER COLORRECTAL

### 2.2.1 Rutas de inestabilidad genómica

El CCR progresa siguiendo una serie de cambios histológicos, epigenéticos y genéticos que provocan inestabilidad genómica y la activación de distintos oncogenes e inactivación de genes supresores de tumores (TSG), alterando así la morfología y función de las células epiteliales y del estroma que las rodea<sup>4-6</sup>. Inicialmente se describieron tres rutas tumorogénicas principales: (i) ruta de la inestabilidad cromosómica (CIN), (ii) la ruta de inestabilidad de microsatélites (MSI), y (iii) el fenotipo metilador de islas CpG (CIMP)<sup>7</sup>.

La ruta CIN se caracteriza por la secuencia adenoma convencional-CCR propuesta por Fearon y Vogelstein en 1990<sup>8</sup>. Esta ruta es la mayoritaria en CCR, y explica el 65-70% de los casos esporádicos<sup>7</sup>. Esta vía se caracteriza por ganancias o pérdidas de cromosomas completos (aneuploidía) o regiones cromosómicas, disfunción telomérica, errores en la maquinaria de respuesta a daños en el ADN, y alteraciones en genes supresores de tumores y oncogenes como *APC* y *KRAS* en etapas tempranas del proceso neoplásico y *TP53* en estadios más avanzados<sup>7,9-11</sup>. También es característica la pérdida de heterocigosidad a lo largo de todo el genoma<sup>7</sup>.

Los tumores que siguen la vía MSI se estima que representan el 15-20% de todos los casos de CCR<sup>12</sup>. Estos se caracterizan por presentar defectos en los genes reparadores de errores de emparejamiento de bases (genes MMR, del inglés *MisMatch Repair*), principalmente *MLH1*, *MSH2*, *MSH6* y *PMS2* necesarios para la correcta replicación del ADN<sup>6</sup>. La inestabilidad de microsatélites que caracteriza a estos CCR es el resultado de la imposibilidad de corregir estos errores, que se ve reflejada en estas secuencias repetitivas del ADN<sup>6,11</sup>. Este tipo de tumores presentan una localización preferencial en el colon derecho, poca diferenciación, abundante infiltración de células T e histología mucinosa o de células en anillo de sello<sup>12</sup>.

El tercer grupo formado por tumores que siguen la ruta CIMP se ha detectado en el 15-20% de los casos esporádicos y surge de lesiones polipósicas distintas de los adenomas clásicos, denominadas adenomas serrados sésiles<sup>4,6</sup>. Se caracteriza por presentar niveles aberrantes de hipermetilación en las islas CpG, regiones del genoma con elevada concentración de dinucleótidos CpG<sup>13</sup>. Las islas CpG se localizan normalmente en las regiones promotoras de los genes y su metilación provoca el silenciamiento transcripcional de estos, ocasionando una disminución de la expresión génica<sup>11,12,14</sup>. Aunque la hipermetilación de islas CpG afecta a cientos de promotores génicos, solo el descenso de la expresión de un subgrupo de TSG relacionados con la aparición de neoplasias colorrectales entre los que se encuentran los genes *CDKN2A*, *MLH1*, *CDH1* y *VHL*, parece ser responsable del proceso neoplásico<sup>13</sup>.

No obstante, estas tres vías no son exclusivas y en un mismo CCR pueden encontrarse características de múltiples rutas, como por ejemplo la presencia de inestabilidad de microsatélites en tumores con un elevado grado de hipermetilación de islas CpG<sup>7</sup>.

Además de estas tres rutas principales de tumorigénesis de CCR posteriormente se reportó la implicación de los microARNs y de la inflamación crónica en el inicio y progresión de CCR<sup>11</sup>.

### 2.2.2 Firmas mutacionales

El conjunto de mutaciones somáticas presentes en un genoma de cáncer es el reflejo de los distintos procesos mutagénicos que han provocado daños en el ADN, así como de la maquinaria de reparación que ha mitigado estos daños a lo largo del tiempo<sup>15</sup>. Estas alteraciones no se encuentran distribuidas de manera aleatoria en el genoma, sino que su localización parece relacionarse con la naturaleza del proceso que las ocasiona<sup>15,16</sup>.

Basándose en esta premisa, en el año 2013 Alexandrov y colaboradores propusieron un modelo teórico que proporcionó las bases para la caracterización de las firmas mutacionales extraídas del conjunto de mutaciones somáticas de cáncer<sup>16</sup>. Esto permitió comprender de manera individual los mecanismos patogénicos subyacentes a los distintos patrones mutacionales<sup>16,17</sup>. Para ello se estudiaron todo tipo de variación genómica (sustituciones nucleotídicas, inserciones-delecciones y reordenamientos) y otro tipo de características como los nucleótidos adyacentes en posiciones 3' y 5' a la alteración o la hebra transcripcional en la que sucede el cambio<sup>17</sup>. Esta primera aproximación reveló un total de 22 firmas mutacionales encontradas en genomas de cáncer, alguna de ellas asociada con procesos como la edad, errores en la maquinaria de reparación del ADN celular y factores mutagénicos tanto exógenos como endógenos<sup>17</sup>.

El aumento del número de datos genómicos tumorales permitió el descubrimiento de nuevas firmas y la descomposición de alguna de las ya existentes en otras con la misma o distinta etiología<sup>18</sup>. Actualmente en el *Catalogue Of Somatic Mutations In Cancer* (COSMIC)<sup>19</sup> se recogen un total de 132 firmas mutacionales (79 asociadas a variantes de nucleótido único o SBS, 11 a sustituciones dinucleotídicas o DBS,

18 a inserciones y deleciones de pequeño tamaño o ID y 24 a alteraciones en el número de copia, algunas con origen desconocido<sup>18,20,21</sup>.

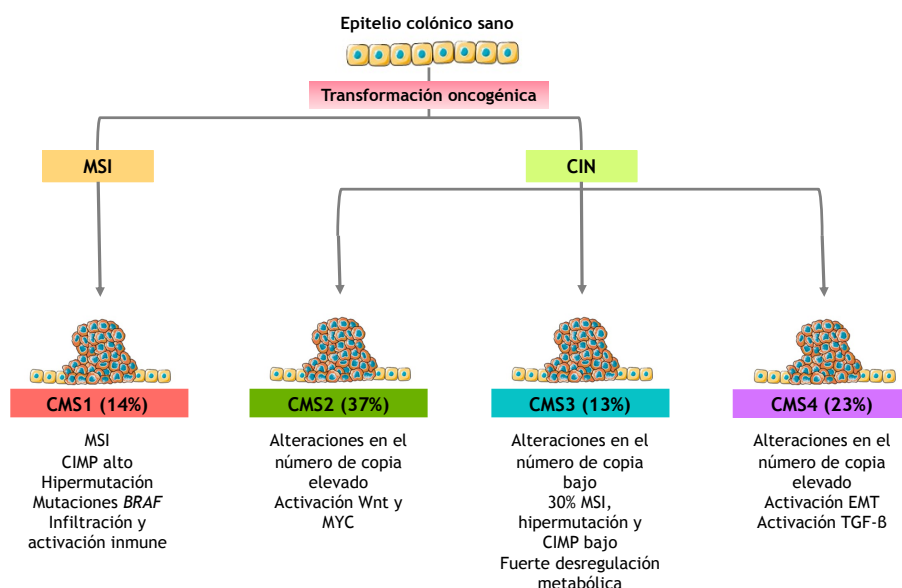
Existen una serie de firmas mutacionales que han sido relacionadas con agentes implicados en el proceso carcinogénico colorrectal. Catorce de estas se asocian a defectos en la maquinaria de reparación celular: defectos en el sistema de reparación MMR (firmas SBS6, SBS15, SBS21, SBS26, SBS44, DBS7, DBS10 e ID7)<sup>17,18,22-24</sup>, mutaciones en el dominio exonucleasa (ED) de las ADN polimerasas  $\epsilon$  y  $\delta 1$  (conjunto de firmas SBS10 y DBS3)<sup>18,25</sup>, el efecto combinado de los procesos anteriormente descritos (SBS14 y SBS20)<sup>17</sup>, y defectos en el sistema de reparación por escisión de bases (BER, del inglés *Base Excision Repair*) como las alteraciones bialélicas en *NTHL1* (SBS30)<sup>23,24</sup> y *MUTYH* (SBS36)<sup>26,27</sup>. También se han descrito firmas relacionadas con carcinógenos exógenos como el tabaco (SBS4)<sup>21,28</sup> o endógenos como la colibactina (SBS88 e ID18)<sup>29-31</sup>.

Recientemente, nuevas perspectivas han surgido con respecto al modelo empleado para la extracción de firmas mutacionales hasta ahora utilizado. Se ha propuesto que una aproximación órgano específica sería biológicamente más representativa de los procesos mutacionales que tienen lugar en un tejido concreto<sup>32</sup>. Aunque muchas de las firmas ya propuestas previamente se confirman por medio de esta aproximación, esto ha permitido describir 40 nuevas firmas SBS y 18 DBS con respecto al catálogo ya existente<sup>32</sup>. Hasta el momento el uso de este nuevo enfoque no está muy establecido, sin embargo, todo parece indicar que posiblemente sea la estrategia a seguir en los próximos años.

### 2.2.3 Subgrupos Moleculares Consenso

En el año 2015, Guinney y colaboradores<sup>33</sup> propusieron una nueva clasificación de CCR basada en el perfil molecular en lugar de histopatológico, los denominados Subgrupos Moleculares Consenso (CMS). Esta estrategia fue llevada a cabo con el fin de establecer un modelo de referencia para la clasificación de los CCR, eliminando así las inconsistencias presentes en estudios basados en patrones de expresión génica sin un protocolo consensuado de análisis<sup>33</sup>. Los datos

transcriptómicos representan el eje central de esta categorización, al estar más ligados al comportamiento tumoral, al que se suma información obtenida de otras tecnologías ómicas (mutaciones puntuales, alteraciones en el número de copia, epigenómica, microARNs y proteómica) para obtener una mayor interpretabilidad biológica y translación a la clínica<sup>33</sup>. El análisis bioinformático mostró cuatro grupos claramente definidos que abarcaban el 87% de las neoplasias: (i) MSI inmune o CMS1, (ii) canónico o CMS2, (iii) metabólico o CMS3, y (iv) mesenquimal o CMS4 (Figura 2), el 13% de los tumores restantes mostraba un fenotipo de transición o la existencia de heterogeneidad intratumoral<sup>33</sup>.



**Figura 2. Subgrupos Moleculares Consenso definidos por Guinney y colaboradores, 2015<sup>33</sup>.** Se muestran los cuatro subtipos transcriptómicos junto a las características moleculares más destacadas de cada uno de ellos. MSI: inestabilidad de microsátélites, CIN: inestabilidad cromosómica, CIMP: fenotipo metilador de islas CpG, EMT: transición epitelio-mesenquimal, TGF- $\beta$ : factor de crecimiento transformante  $\beta$ .

Del conjunto de tumores analizado, el 14% de los CCR se clasificaron como CMS1<sup>33</sup>. Este grupo de tumores se caracteriza por presentar hipermutación somática, baja prevalencia de alteraciones en el número de copia, MSI, un estado de hipermetilación generalizado,

fuerte infiltración y activación inmune, sobreexpresión de proteínas implicadas en la reparación de daños en el ADN como las proteínas MMR, y enriquecimiento de la mutación somática p.(Val600Glu) de *BRAF*<sup>33,34</sup>. Se ha observado que este tipo de tumores se presentan principalmente en mujeres y muestran una localización preferencial en el colon derecho y mayor grado histopatológico, además de que los pacientes muestran una menor supervivencia tras una recaída de la enfermedad<sup>33</sup>.

Los CMS restantes, al contrario que CMS1, muestran inestabilidad cromosómica<sup>33,34</sup>. CMS2 y CMS4 presentan niveles similares de alteraciones somáticas en el número de copia y también un espectro mutacional parecido, así como estabilidad de microsatélites (MSS) y bajos niveles de hipermetilación génica, sin embargo, se ven diferenciados por otras características moleculares<sup>34</sup>. CMS2, que representa el 37% del total de los CCR, se caracteriza por mostrar diferenciación epitelial y marcada activación de las dianas de las rutas de señalización Wnt y MYC, además de mayor expresión de oncogenes como *HNF4A* y de la familia de microARN miR-17-92<sup>33,34</sup>. Estos tumores se localizan preferencialmente en la región izquierda y es el grupo con mejores tasas de supervivencia de los pacientes<sup>33</sup>.

Los tumores clasificados como CMS4, 23%, muestran un fenotipo mesenquimal con influencia proangiogénica y estromagénica<sup>33</sup>. Se caracterizan por presentar una fuerte activación de rutas relacionadas con la transición epitelio-mesenquimal (EMT) y de genes de troncalidad (*stemness*), como la activación del factor de crecimiento transformante  $\beta$  (TGF- $\beta$ ), integrinas, remodelación de la matriz extracelular y el sistema del complemento<sup>33,34</sup>. Los pacientes con CCRs clasificados dentro de este grupo se diagnostican en estadios más avanzados de la enfermedad (III y IV) y presentan peor supervivencia<sup>33</sup>.

El CMS3 presenta un perfil genómico y epigenómico característico que lo diferencia de los otros subtipos con inestabilidad cromosómica, y en él se agrupan el 13% de los CCR<sup>33</sup>. En comparación con CMS2 y CMS4, la cantidad de alteraciones en el número de copia es menor y hasta un 30% de los tumores de este grupo presentan MSI, hipermutación somática y niveles intermedios de hipermetilación<sup>33,34</sup>. Los tumores CMS3 epiteliales se caracterizan por una fuerte

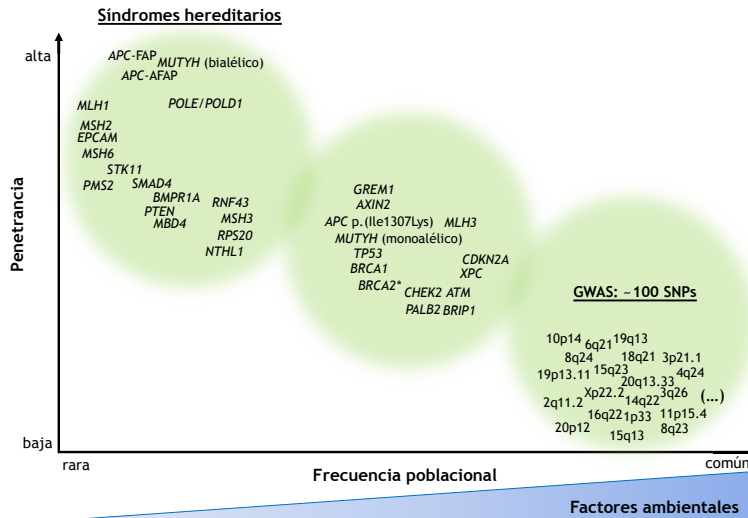


desregulación metabólica, lo que concuerda con el enriquecimiento de mutaciones activadoras en *KRAS*<sup>33-35</sup>.

Aunque los CMS actualmente son el sistema de clasificación con mayor interpretabilidad biológica y fácilmente trasladables a la práctica clínica, se espera un mayor refinamiento de estos a medida que se añadan datos moleculares e histopatológicos.

### **2.3 ARQUITECTURA GENÉTICA DEL CÁNCER COLORRECTAL**

El CCR es una enfermedad compleja que surge de la interacción entre factores de riesgo ambientales, epigenéticos y genéticos<sup>36,37</sup> (Figura 3). Se trata principalmente de una enfermedad esporádica de aparición tardía, la media de edad de los pacientes diagnosticados es de 70 años<sup>38</sup>. Sin embargo, estudios con gemelos han estimado la heredabilidad del CCR en hasta un 35%<sup>39</sup>. Esta heredabilidad se explica por la presencia en el genoma de variantes que van desde variantes muy raras con penetrancia elevada, responsables de los síndromes hereditarios de cáncer, hasta variantes comunes de baja penetrancia que de forma individual confieren un riesgo muy bajo para el desarrollo de la enfermedad, pero al interactuar entre ellas resultarían en el riesgo individual de cada persona<sup>38</sup>. La predisposición hereditaria y ciertos factores inherentes al individuo no son modificables, sin embargo, en la carcinogénesis de CCR existen una serie de factores modificables sobre los que se puede actuar.



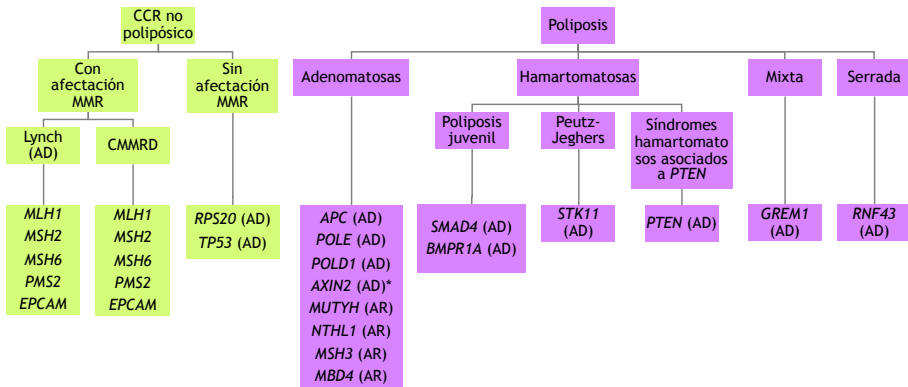
**Figura 3. Arquitectura genética del CCR.** Distribución de los distintos genes, variantes y SNPs de susceptibilidad a CCR descritos hasta el momento en función de su frecuencia poblacional y riesgo, y el efecto de los factores ambientales en función del tipo de variante presente. *BRCA2\**: reportes contradictorios respecto a riesgo. Imagen adaptada con permiso de BMJ Publishing Group Limited [Genetic architecture of colorectal cancer, Peters U, Bien S, Zubair N, 64(10), 1623-36. Gut, 2015]<sup>37</sup> (Anexo I).

Se estima que los factores ambientales, modificables y no modificables, representan un riesgo asociado a carcinogénesis colorrectal de entre el 80-90%<sup>40</sup>. Entre los factores modificables se encuentran los relacionados con la dieta y el estilo de vida del individuo, que pueden generar alteraciones en la microbiota intestinal<sup>41</sup>. Se ha visto que el consumo de carnes rojas y procesadas, granos refinados, bebidas azucaradas y también bebidas alcohólicas, elevada masa corporal y el hábito tabáquico son factores de riesgo de aparición de CCR<sup>41-44</sup>. Por el contrario, el consumo de fibra, grano integral, frutas y verduras, pescado, aves de corral, soja, productos lácteos bajos en grasa, vitamina D, el uso de aspirina y otros antiinflamatorios no esteroideos o la actividad física regular, desempeñan un papel protector frente a la tumorigénesis colorrectal<sup>41,42,45,46</sup>. También se ha observado que existe un grupo de factores ambientales que no son modificables y que aumentan el riesgo a padecer CCR, como son la edad, el sexo masculino, o enfermedades intestinales existentes como colitis ulcerativa o la enfermedad de Chron y la diabetes tipo 2<sup>40,44,47</sup>.

Los factores genéticos constituyen otro componente de riesgo a CCR que no es modificable. Hasta un 24% de los CCR diagnosticados ocurren en un contexto de agregación familiar<sup>48</sup>. A pesar de los esfuerzos realizados para conocer la heredabilidad que explique la aparición de CCR en individuos dentro de dicho contexto, solo un pequeño porcentaje de casos ha logrado explicarse por variantes que presentan alto, moderado y bajo riesgo para la aparición de la enfermedad<sup>49</sup>. Se definen así los síndromes de predisposición hereditaria a CCR caracterizados generalmente por un elevado riesgo de aparición de neoplasias colorrectales que se dividen en relación a la ausencia o presencia de lesiones polipósicas, genes de moderada penetrancia relacionados con rutas implicadas en CCR u otros síndromes de predisposición a cáncer, y variantes de bajo riesgo que son comunes en la población<sup>37,44,46,48,50</sup>.

### **2.3.1 Síndromes de predisposición hereditaria a CCR**

Estos síndromes siguen un patrón de herencia mendeliana y representan aproximadamente el 5% de todos los CCR diagnosticados<sup>48,49</sup>. Están causados por variantes patogénicas con frecuencia poblacional rara (<1%) y alta penetrancia, y generalmente la aparición del CCR en el individuo tiene lugar a una edad temprana<sup>48,49</sup>. Estos se dividen en síndromes no polipósicos y síndromes polipósicos clasificados en función del número y la histología de los pólipos (adenomatosos, hamartomatosos y serrados) (Figura 4).



**Figura 4. Genes de susceptibilidad a los síndromes hereditarios de CCR y poliposis propuestos hasta la fecha.** AXIN2 (AD)\*: presencia de CCR o pólipos de diversas histologías (adenomas, pólipos hiperplásicos) MMR: genes reparadores de errores de emparejamiento de bases (MisMatch Repair), AD: herencia autosómica dominante, AR: herencia autosómica recesiva.

### 2.3.1.1 Síndromes hereditarios no polipósicos

Incluyen la mayoría de los casos de CCR hereditario, y se dividen en función del fenotipo tumoral que presenten en MMR deficiente o MMR eficientes, según se haya detectado o no afectación de la expresión de las proteínas MMR (MLH1, MSH2, MSH6 y PMS2).

#### 2.3.1.1.1 Síndrome de Lynch

El síndrome de Lynch (OMIM #120435) está causado por la presencia de variantes patogénicas monoalélicas (heterocigotas) en línea germinal en uno de los genes MMR (*MLH1*, *MSH2*, *MSH6* o *PMS2*) y también por alteraciones epigenéticas como la delección en el extremo 3' de *EPCAM* que provoca el silenciamiento epigenético de *MSH2*<sup>46,48</sup>. Es el responsable del 2-3% de todos los casos de CCR, lo que lo convierte en el síndrome hereditario más común<sup>37</sup>. Las características principales de los individuos afectados son la aparición temprana de la enfermedad (media de 45 años) y la presencia de tumores con MSI elevada, tumores hipermutados ( $\geq 10$  mutaciones/Megabase (Mb)) y ausencia de expresión proteica por

inmunohistoquímica de las proteínas afectadas<sup>37,46,48</sup>. El colon y recto junto con endometrio son las principales localizaciones de aparición de neoplasias, sin embargo, también afecta a otros tejidos como ovario, tracto urinario, cerebro, páncreas, próstata, estómago, intestino delgado y tumores sebáceos de la piel<sup>46,48</sup>. Si incluye esta última manifestación extracolónica se denomina síndrome de Muir-Torre<sup>46,48</sup>. Debido al modo de acción de estos genes MMR, los individuos con variantes patogénicas que afectan a los genes *MLH1* y *MSH2* presentan fenotipos más severos que aquellos que portan variantes causales en *MSH6* o *PMS2*<sup>48</sup>.

Existe una variación poco frecuente del síndrome de Lynch, denominado síndrome de Deficiencia Constitucional de Reparación de Errores de Apareamiento (OMIM #276300) o CMMRD (del inglés *constitutional MMR-deficiency syndrome*) ocasionado por alteraciones germinales bialélicas patogénicas en uno de estos cuatro genes MMR<sup>50</sup>. Esta condición predispone a un elevado número de tumores malignos (glioblastomas, hematológicos, cáncer colorrectal), pólipos adenomatosos, que en la mayoría de los casos aparecen antes de la segunda década de vida; además en algunos casos el fenotipo de los pacientes puede solapar con el de la Neurofibromatosis por la presencia de manchas *café-au-lait*<sup>48,50</sup>.

### 2.3.1.1.2 CCR hereditario sin afectación en

#### MMR

Un elevado número de los casos diagnosticados con CCR no son debidos a alteraciones del sistema MMR, y han supuesto el foco de atención de muchas investigaciones. Los análisis de ligamiento, así como las aproximaciones metodológicas que emplean secuenciación masiva (NGS, del inglés *Next-Generation Sequencing*), han tenido poco éxito, reportándose hasta el momento dos genes con evidencia científica suficientes (*TP53* y *RPS20*)<sup>46,48,50</sup>.

El síndrome de Li-Fraumeni (OMIM #151623), ocasionado por variantes patogénicas heterocigotas a nivel germinal en el gen *TP53* es un síndrome hereditario de predisposición a cáncer muy severo con una penetrancia que oscila entre el 70% y 100%<sup>48</sup>. Aunque principalmente no se relaciona con predisposición a CCR, el aumento de la realización

de paneles de genes en la práctica clínica, así como otras aproximaciones NGS, ha permitido observar que entre el 1-14% de los pacientes con diagnóstico de CCR antes de los 50 años son portadores de variantes patogénicas en este gen<sup>48</sup>.

Dentro del grupo de tumores colorrectales sin deficiencia de la vía MMR destaca el denominado Cáncer Familiar de tipo X, que incluye familias que no presentan alteraciones germinales en los genes MMR, pero que cumplen criterios clínicos de Amsterdam I (tres individuos con CCR en más de dos generaciones y al menos uno diagnosticado antes de los 50 años), definidos para la identificación de pacientes con sospecha de Síndrome de Lynch<sup>46</sup>. En individuos pertenecientes a tres familias clasificadas como cáncer familiar de tipo X se identificaron variantes patogénicas germinales en el gen *RPS20*<sup>50</sup>. Este gen sigue un patrón de herencia dominante, no se han reportado manifestaciones extracolónicas y parece presentar baja prevalencia y alta penetrancia, aunque se necesitan más datos para definir la susceptibilidad a CCR<sup>46,50</sup>.

### 2.3.1.2 Síndromes hereditarios polipósicos

Estos síndromes, como su propio nombre indica, se caracterizan por la presencia de lesiones polipósicas ubicadas en el colon y recto. Su clasificación se basa en el número y la histología de las lesiones<sup>48</sup>. Clásicamente este grupo comprendía alteraciones patogénicas en *APC* y *MUTYH* en el caso de poliposis adenomatosa y en los genes *SMAD4*, *BMPRIA*, *STK11* y *PTEN* asociados a lesiones hamartomatosas<sup>49</sup>. Sin embargo, el abaratamiento de las tecnologías NGS y su uso extendido en el campo diagnóstico han permitido describir nuevos síndromes hereditarios polipósicos.

#### 2.3.1.2.1 Poliposis adenomatosa

Dentro de los síndromes hereditarios caracterizados por la presencia de lesiones adenomatosas encontramos distintos síndromes y genes implicados que se describen a continuación.

El primero de estos síndromes en ser descrito fue la poliposis adenomatosa familiar (FAP, del inglés *familial adenomatous polyposis*) (OMIM #175100), causada por variantes patogénicas

germinales en *APC*, un TSG implicado en la regulación de la ruta de señalización Wnt<sup>37,48</sup>. FAP sigue un patrón de herencia dominante, donde el 30% de los individuos diagnosticados no presentan historia familiar y son portadores de variantes *de novo*<sup>37,46</sup>. Se caracteriza por la presencia de cientos a miles de pólipos adenomatosos a lo largo de todo el epitelio colorrectal a partir de la primera década de vida y una penetrancia de casi el 100% a la edad de 40 años<sup>37</sup>. También se asocia a manifestaciones extracolónicas, como cáncer de tiroides papilar, tumores desmoides y pólipos gastrointestinales<sup>46</sup>. Existe una forma atenuada de este fenotipo denominada FAP atenuada o AFAP (por sus siglas en inglés de *attenuated familial adenomatous polyposis*) que se caracteriza por la presencia de un menor número de pólipos (20-100) debido a alteraciones patogénicas localizadas en los extremos 3' y 5' del gen y en el exón 9, la cual presenta un inicio más tardío y una penetrancia del 70%<sup>37,46,48</sup>.

Gracias al uso de la NGS y el estudio de individuos que presentan agregación familiar, se describió un nuevo síndrome polipósico que presenta herencia autosómica dominante, la poliposis adenomatosa asociada a la lectura de prueba de la polimerasa (PPAP, *polymerase proofreading-associated polyposis*) (OMIM #612591)<sup>50</sup>. Las variantes patogénicas localizadas en los ED de las ADN polimerasas  $\epsilon$  (gen *POLE*) y  $\delta 1$  (gen *POLD1*) dan lugar a fenotipos polipósicos solapantes con FAP y AFAP, CCR y otras manifestaciones extracolónicas como el cáncer de endometrio<sup>51</sup>. Los tumores ocasionados por este síndrome tienen un elevado número de variantes somáticas (tumores ultramutados, >100 mutaciones/Mb) y pueden mostrar expresión inmunohistoquímica ausente o conservada de las proteínas MMR<sup>48,50</sup>.

Existen además síndromes polipósicos que siguen un patrón de herencia recesivo. El más conocido es la poliposis asociada a *MUTYH* (MAP, del inglés *MUTYH-associated polyposis*) (OMIM #608456) debida a variantes patogénicas bialélicas a nivel germinal en el gen *MUTYH*, gen de la ruta BER implicado en la reparación de daños oxidativos en el ADN<sup>52</sup>. Estos pacientes presentan un riesgo del 80% de desarrollar CCR y su fenotipo clínico en ocasiones solapa con el de FAP y AFAP, también pueden presentar manifestaciones extracolónicas<sup>48</sup>. También se engloban en este apartado los síndromes

de poliposis asociada a *NTHL1* (OMIM #616415) y poliposis asociada a *MSH3* (OMIM #617100) que pertenecen a rutas previamente relacionadas con la tumorigénesis colorrectal, ruta BER y MMR respectivamente<sup>48,50</sup>. Recientemente se ha descrito el síndrome neoplásico asociado a *MBD4* (OMIM #619975), cuyo gen causal está implicado en la ruta BER<sup>53</sup>.

Dentro de este grupo de poliposis adenomatosas también se engloba el síndrome Oligodontia-CCR (OMIM #608615)<sup>54</sup>. Causado por variantes monoalélicas germinales en *AXIN2* (un gen de riesgo moderado) implicado en la ruta Wnt y caracterizado por la presencia de pólipos adenomatosos e hiperplásicos y/o CCR<sup>46,48,55</sup>.

#### 2.3.1.2.2 Poliposis hamartomatosas

Se caracterizan por la presencia de pólipos con histología hamartomatosa. Entre estos síndromes se encuentra la poliposis juvenil (OMIM #174900), ocasionada por variantes patogénicas en los genes *SMAD4* y *BMPRIA* implicados en la ruta TGF- $\beta$ /BMP, siguiendo un patrón de herencia dominante<sup>46</sup>. Se caracteriza por la presencia de pólipos juveniles a lo largo del tracto gastrointestinal<sup>56</sup>.

Las variantes germinales en *STK11*, inhibidor de la ruta mTOR, son responsables del síndrome de Peutz-Jeghers (OMIM #175200)<sup>48</sup>. Los pólipos hamartomatosos se distribuyen a lo largo de todo el tracto gastrointestinal lo que implica un aumento de riesgo de tumores CCR, gástricos y pancreáticos, pero también de otras manifestaciones extracolónicas como mama, pulmón y genitales<sup>46</sup>.

Los síndromes hamartomatosos asociados a *PTEN* (OMIM #158350) incluyen el síndrome de Cowden (SC), de Bannayan-Riley-Ruvalcaba (SBBR), de Proteus y Proteus-like<sup>56</sup>. SC y SBBR se consideran variantes fenotípicas de los mismos síndromes de poliposis hamartomatosas cuya expresión es variable y su penetrancia se relaciona con la edad<sup>48,57</sup>. *PTEN* es un TSG que está implicado en la ruta de señalización metabólica PI3K/Akt<sup>46,48,58</sup>.

#### 2.3.1.2.3 Poliposis mixta hereditaria

El síndrome de poliposis mixta hereditaria (OMIM #601228) sigue un patrón de herencia autosómico dominante<sup>46</sup>. Los



pacientes aquejados de esta patología presentan múltiples pólipos con más de un tipo histológico (adenomas, pólipos hiperplásicos y juveniles) y/o pólipos individuales con características histológicas solapantes (pólipos juveniles atípicos con mezcla de características histológicas)<sup>59</sup>. La aparición de la enfermedad se asocia principalmente a una duplicación que incluye el extremo 3' de *SCG5*, gen localizado aguasarriba de *GREM1*<sup>59,60</sup>.

#### 2.3.1.2.4 Poliposis serrada

Las variantes patogénicas heterocigotas germinales en *RNF43*, una ubiquitina ligasa E3 RING-type, dan lugar al síndrome de poliposis serrada (OMIM #617108)<sup>48</sup>. *RNF43* actúa como inhibidor de la ruta Wnt al inducir la endocitosis de sus receptores, y también ha sido propuesto como regulador de la respuesta a daños en el ADN mediada por ATM-ATR<sup>61,62</sup>.

Los pacientes con esta condición desarrollan pólipos hiperplásicos, adenomas serrados sésiles y adenomas serrados tradicionales a lo largo del colon, y presentan un incremento del riesgo a CCR del 16%<sup>50</sup>. Clínicamente este síndrome se define con una serie de criterios propuestos por la Organización Mundial de la Salud: (i) presencia de al menos 5 pólipos serrados próximos a sigma donde dos de ellos sean mayores de 10mm, (ii) cualquier número de pólipos serrados cercanos a sigma en un individuo con un familiar de primer grado diagnosticado de este síndrome, o (iii) más de 20 pólipos serrados distribuidos a lo largo del colon<sup>63</sup>.

### 2.3.2 Variantes de moderada penetrancia

El uso de paneles de genes en la práctica clínica ha permitido identificar una serie de variantes y genes que parecen exhibir una moderada penetrancia en la aparición de CCR, lo que supone un reto a la hora del manejo clínico de los pacientes en esta situación y también de sus familiares<sup>46</sup>.

Se han descrito variantes en genes implicados en síndromes hereditarios de CCR, como es el caso de la variante p.(Ile1307Lys) de *APC*, una variante fundadora en población Ashkenazi pero muy rara en el resto de poblaciones, que está asociada a un aumento de riesgo a CCR

del 2,17<sup>64</sup>. También ha sido descrito que las alteraciones monoalélicas en *MUTYH* provocan un incremento de 1,5-2 veces de padecer CCR, especialmente en aquellos individuos con un familiar de primer grado con diagnóstico de la enfermedad<sup>65</sup>.

La variante p.(Ser1188\*) homocigota en el gen *MLH3*, perteneciente a la ruta MMR, ha sido descrita en una familia finlandesa con fenotipo FAP/AFAP y manifestaciones extracolónicas<sup>66</sup>. Actualmente el grupo de expertos de ClinGen asocia a *MLH3* con riesgo moderado a padecer CCR<sup>54</sup>.

Genes implicados en otros síndromes hereditarios también parecen tener un papel en el desarrollo de CCR, sin embargo, se necesita un mayor número de datos para determinar si estos presentan un riesgo bajo o moderado a la enfermedad. La variante c.1100delC de *CHEK2*, principalmente relacionada con tumores mamarios ha sido reportada de manera frecuente en familias con diagnóstico combinado de cáncer de mama y CCR y se ha propuesto un aumento de dos veces el riesgo a padecer este último frente a la población general<sup>48</sup>.

Alteraciones patogénicas germinales en los genes *BRCA1* y *BRCA2* asociadas a cáncer de mama y ovario hereditario se han encontrado en individuos y co-segregando en familias con diagnóstico de CCR<sup>48,67</sup>. Análisis posteriores han propuesto que alteraciones en *BRCA1* podrían elevar el riesgo a padecer CCR en 1,48 veces en mujeres por debajo de los 50 años, sin embargo, esta asociación no se extiende a *BRCA2*<sup>68,69</sup>.

Otros estudios que han analizado la presencia de alteraciones en genes de predisposición a síndromes de cáncer hereditario han encontrado variantes patogénicas o probablemente patogénicas no relacionadas con CCR como es el caso de *PALB2*, *ATM*, *CDKN2A*, *BRIP1* o *XPC*, entre otros<sup>50,70,71</sup>.

### 2.3.3 Variantes de baja penetrancia y estudios de asociación

Los genes de moderada y alta penetrancia previamente descritos no han sido capaces de dar explicación a un gran porcentaje de casos de CCR que tienen lugar en grupos de agregación familiar a edades de diagnóstico tardías, en los cuales la aparición de este tipo de neoplasias es superior a la de la población general. Los *loci* genéticos que

predisponen a un bajo riesgo a la enfermedad parecen explicar el 12% de estos casos<sup>72</sup>.

La hipótesis “Enfermedad Común-Variante Común” estableció la importancia de estos *loci* en la herencia de una enfermedad común como es el CCR<sup>36,73</sup>. La estrategia seguida para la identificación de este tipo de variación común han sido los estudios de asociación (GWAS, *Genome-wide association studies*) caso-control de genoma completo<sup>74</sup>. La finalización del proyecto HapMap que demostró que la mayoría de las variantes del genoma se heredan en bloques de desequilibrio de ligamiento y no de forma independiente, junto con el desarrollo de plataformas de genotipado masivo de marcadores genéticos tipo SNPs (del inglés, *Single Nucleotide Polymorphisms*) han facilitado la realización de numerosos estudios GWAS y estrategias *fine-mapping* en los que se han identificado los aproximadamente 100 *loci* comunes de bajo riesgo (*Odds Ratio* de 1,04-1,23) que se asocian en la actualidad a CCR<sup>36,48,72</sup>.

Los *loci* individuales identificados se distribuyen a lo largo de todo el genoma, y en algunos casos han sido asociados a genes previamente relacionados con CCR, como los relacionados con la ruta TGF- $\beta$ <sup>75</sup>. Dada la baja magnitud de efecto que presentan estos SNPs hace que estos no sean informativos de manera individual<sup>76</sup>. Sin embargo, la información que proporcionan en conjunto puede ser de gran utilidad para la identificación de grupos de riesgo a nivel poblacional<sup>76</sup>. A la herencia conjunta de dos o más de estos alelos que ejercen un efecto aditivo se le conoce como herencia poligénica<sup>48</sup>. Esta herencia poligénica se ve fuertemente influenciada por el factor ambiental que será el que finalmente determine qué individuos desarrollarán la enfermedad<sup>48,77</sup>. A través de la herencia poligénica es posible estimar el riesgo a desarrollar CCR siguiendo el modelo de puntuación de riesgo poligénico<sup>48</sup>. Estos modelos de predicción todavía no se aplican a la práctica asistencial, pero en un futuro pueden ser determinantes para mejorar las estrategias poblacionales de cribado de CCR<sup>76,78</sup>.

## 2.4 GENES CANDIDATOS DE SUSCEPTIBILIDAD HEREDITARIA A CCR

Además de los genes de alta-moderada penetrancia descritos previamente y las variantes comunes de baja penetrancia, todavía existe

un gran porcentaje de heredabilidad perdida en CCR que permanece sin explicación. Con el fin de descubrir nuevos genes implicados en la predisposición a padecer CCR se han llevado a cabo técnicas de secuenciación de exoma completo (WES, *Whole-exome sequencing*) y genoma, no solo a nivel germinal sino también somático en cohortes fenotípicamente enriquecidas.

Estas aproximaciones han identificado un pequeño número de genes cuyo papel en la patología ha sido validado en cohortes externas, como por ejemplo los ya mencionados *POLE* y *POLD1* o *RPS20*, pero también han dejado una lista de posibles candidatos cuya implicación en la tumorigénesis colorrectal todavía es desconocida. Algunos de estos genes propuestos parecen presentar mayores evidencias que respaldan su participación en la susceptibilidad a CCR, como es el caso de *GALNT12* propuesto por varios estudios, o *BUB1* donde se detectaron 6 familias con enriquecimiento de casos de CCR que eran portadoras de alteraciones germinales en este<sup>48,50</sup>. Otros candidatos implicados en distintos procesos celulares han sido propuestos, sin embargo, su asociación con la enfermedad requiere más evidencias, como es el caso de genes implicados en la respuesta a daño del ADN (*WRN*, *ERCC6* y *FAN1*), adhesión focal (*FOCAD*), ruta de señalización Wnt (*LRP6* y *SETD6*) o la ruta TGF- $\beta$  (*BMPR2*), entre otros muchos<sup>48,50,79</sup>.

Por lo tanto, es necesario el estudio no solo de un mayor número de pacientes, sino también el empleo de nuevas aproximaciones que permitan añadir capas de conocimiento con el fin de conocer mejor los patrones moleculares que rigen esta susceptibilidad genética a CCR.



# **JUSTIFICACIÓN DEL ESTUDIO**



### 3 JUSTIFICACIÓN DEL ESTUDIO

La WES ha sido exitosa en la identificación de nuevos genes candidatos en síndromes raros de cáncer con fenotipos muy bien definidos. En CCR, a pesar de que se han identificado nuevos síndromes mendelianos de poliposis hereditarias (causados por variantes raras en *POLE*, *POLD1*, *NTHL1*, *MBD4*), muchos de los genes candidatos propuestos no han sido replicados en cohortes de validación. La identificación de variantes raras de alta-moderada penetrancia asociadas a CCR sigue siendo un reto, debido al número excesivo de variantes raras no causales que existen en el genoma, y a la heterogeneidad del CCR.

Además, las estrategias empleadas hasta ahora para la priorización tanto de variantes como de genes no parecen ser las más adecuadas en el estudio de enfermedades heterogéneas como el CCR, incluso aplicadas en cohortes enriquecidas genéticamente y aparentemente homogéneas, como CCR en pacientes jóvenes y/o con agregación familiar. Esto demuestra que estos fenotipos clínicos siguen siendo heterogéneos y es necesaria una mejor caracterización de los mismos a través, por ejemplo, del endofenotipado de los tumores, es decir, obteniendo su perfil mutacional a nivel genoma.

La hipótesis planteada es que el estudio de fenotipos CCR homogéneos facilitará el descubrimiento de variantes raras causales, ya que cabe esperar una sobrerrepresentación de estas en un mismo gen o genes pertenecientes a la misma vía. Al realizar estudios de WES y transcriptómica integrando los datos somáticos de CCR y germinales, se obtendrá información relevante sobre: la heterogeneidad/homogeneidad molecular del fenotipo definido de forma clínica; la funcionalidad de aquellas variantes germinales que supuestamente tienen una implicación oncogénica; y la aparición de segundos eventos mutacionales somáticos en genes candidatos y/o acúmulo de mutaciones somáticas en genes de la misma vía. Estas aproximaciones permitirán reducir el número final de variantes



germinales candidatas con respecto a los estudios WES clásicos y facilitar la identificación de nuevos genes de susceptibilidad a CCR.

# OBJETIVOS



## 4 OBJETIVOS

El objetivo de este estudio es la identificación de nuevos genes de susceptibilidad a cáncer colorrectal, mediante estudios de funcionalidad que integren la variación genética germinal y tumoral de cohortes molecularmente homogéneas de pacientes CCR, enriquecidas genéticamente por un diagnóstico temprano de cáncer. Se utilizarán las aproximaciones ómicas de RNA-seq, secuenciación de exoma completo y metilación.

Este objetivo global se llevará a cabo de acuerdo a los siguientes objetivos específicos:

1. Seleccionar pacientes enriquecidos genéticamente por diagnóstico de CCR a edad temprana y sin alteraciones en la vía de reparación de errores de emparejamiento de bases.
2. Definir los fenotipos tumorales tras obtención del perfil mutacional de los CCR, mediante secuenciación de exoma completo, RNA-seq y metilación.
3. Comparar los perfiles de expresión génica tumoral en las cohortes CCR (TOGETHER y TCGA) frente a los perfiles de expresión germinal en mucosa colónica normal (cohortes TOGETHER y GTEx).
4. Integrar los datos ómicos germinales y tumorales para obtener información adicional sobre genes o vías relevantes en CCR.



# MATERIAL Y MÉTODOS



## 5 MATERIAL Y MÉTODOS

### 5.1 COHORTES DE ESTUDIO EN LA FASE DE DESCUBRIMIENTO

Se seleccionaron pacientes con los siguientes criterios clínicos de inclusión: diagnóstico de CCR a una edad igual o inferior a los 50 años, sin fenotipo polipósico (<10 pólipos), sin tratamiento oncológico previo, con o sin historia familiar de cáncer, y con tumores sin alteraciones en la vía de los genes MMR.

Para llevar a cabo los análisis que se desarrollarán a lo largo de este apartado también se seleccionaron datos de mucosa colónica normal de pacientes sin antecedentes personales de CCR, a la edad de la toma de la muestra, procedentes de la base de datos *The Genotype-Tissue Expression* (GTEx), y datos de pacientes diagnosticados de la enfermedad con edades comprendidas entre los 51 y 65 años recogidos en 5.1.2 Cohorte de la base de datos *The Cancer Genome Atlas* (TCGA): cohortes COAD y READ.

#### 5.1.1 Cohorte TOGETHER

Estudio multicéntrico en el que participaron los Servicios de Endoscopias Digestivas (Hospital Universitario Lucus Augusti (HULA), Complejo Hospitalario Universitario de Pontevedra (CHOP) y Complejo Hospitalario Universitarios de Ourense (CHUO), los Servicios de Cirugía y Digestivo del Complejo Hospitalario Universitario de Santiago de Compostela (CHUS) y Hospital Clínic (Barcelona), y los Servicios de Anatomía Patológica del CHUS, CHOP y Hospital Clínic (Barcelona).

Se seleccionó un grupo de 20 pacientes, procedentes de familias no relacionadas, diagnosticados de CCR a la edad de 50 años o inferior, con y sin historia familiar de cáncer, y con tumores sin alteración de la vía MMR (confirmada por expresión conservada de las proteínas MLH1, MSH2, MSH6 y PMS2 por inmunohistoquímica). Ninguno de los pacientes recibió tratamiento oncológico previo para esta patología (Tabla suplementaria 1).



En todos los casos el paciente fue debidamente informado de los objetivos y finalidad del estudio, así mismo se firmó por triplicado el consentimiento informado propio del estudio (Anexo II). Además, el personal médico se encargó de cubrir el cuaderno de recogida de datos con la información clínica, personal y familiar de cáncer requerida para de cada uno de los pacientes.

#### 5.1.1.1 Obtención de muestras biológicas

De cada uno de los pacientes se obtuvo remanente de muestras de mucosa colónica normal y de CCR, recogidas a distancias de entre 10 y 15 centímetros y dentro del mismo segmento colónico.

En los Servicios de Endoscopias Digestivas de los centros participantes se tomaron entre 6 y 8 biopsias colónicas de 2mm tanto para mucosa colónica normal como tumoral. Las muestras se conservaron en la solución de conservación *RNAlater RNA Stabilization Reagent* (Qiagen, Hilden, Alemania) hasta su procesamiento al día siguiente.

Las piezas de mucosa colónica normal y tumoral obtenidas a través de los Servicios de Cirugía fueron seleccionadas por un patólogo y congeladas de manera inmediata en un tiempo no superior a 20 minutos tras su extracción. Estas muestras se almacenaron a  $-80^{\circ}\text{C}$  hasta su procesamiento.

#### 5.1.1.2 Extracción de ADN y ARN

Las muestras de mucosa colónica normal y CCR, independientemente del procedimiento de obtención de la muestra, se incluyeron en un tubo de microcentrífuga de 1,5 ml (Eppendorf, Hamburgo, Alemania) que contenía una solución de lisis compuesta de 98% Buffer RLT (Qiagen, Hilden, Alemania) y 2% ditiotritol 2M (VWR, West Chester, Pensilvania, Estados Unidos) y se homogeneizaron manualmente y por separado empleando una micromano de mortero (Eppendorf, Hamburgo, Alemania). El homogeneizado obtenido se pasó a través de una columna de homogeneización QIAshredder (Qiagen, Hilden, Alemania).

La extracción tanto del ADN como del ARN procedentes de la mucosa colónica normal y tumoral se realizó con el kit AllPrep

DNA/RNA/miRNA Universal (Qiagen, Hilden, Alemania) siguiendo las especificaciones del fabricante.

A continuación, se evaluó mediante espectrofotometría la concentración y pureza de estos, para lo que se emplearon los sistemas *Nanodrop® 1000 Spectrophotometer* (Thermo Fisher Scientific, Massachusetts, Estados Unidos), *SimpliNano UV-Vis Spectrophotometer* (GE Healthcare, Illinois, Estados Unidos). También se midió la concentración de ácidos nucleicos empleando el método de cuantificación por fluorimetría *Qubit™ fluorometer*, concretamente *Qubit™ dsDNA BR Assay Kit* para la cuantificación del ADN y *Qubit™ RNA BR Assay Kit* para ARN (Thermo Fisher Scientific, Massachusetts, Estados Unidos).

Por último, se determinó la integridad de los ácidos nucleicos empleando el sistema de electroforesis digital *2200 TapeStation System* (Agilent Technologies, California, Estados Unidos) y los sistemas *Genomic DNA ScreenTape* y sus reactivos (Agilent Technologies, California, Estados Unidos) y *RNA ScreenTape* y sus reactivos (Agilent Technologies, California, Estados Unidos), para medir las integridades de ADN y ARN respectivamente.

El material genético obtenido de este proceso se almacenó a  $-80^{\circ}\text{C}$  para mantener su integridad.

### **5.1.2 Cohorte de la base de datos *The Cancer Genome Atlas* (TCGA)**

De la base de datos TCGA se obtuvo información a nivel de exoma, secuenciación de ARN y metilación de un total de 26 pacientes diagnosticados de cáncer colorrectal ( $\leq 50$  años) procedentes de los conjuntos de datos de adenocarcinomas de colon y recto COAD y READ (versión dbGAP phs000178.v11.p8).

Según la información recogida en la base de datos TCGA los 26 pacientes eran europeos no finlandeses y ninguno había sido sometido a tratamiento previo de la enfermedad. No se dispone de datos sobre la alteración de las proteínas MMR en el tumor de 15 pacientes, mientras que 9 individuos presentaban tumores sin alteración de la vía MMR y 2 con tumores con pérdida de expresión de las proteínas reparadoras MMR y/o inestabilidad de microsatélites. Los datos de estos 2 pacientes

también se incluyeron en el estudio con la finalidad de obtener una cohorte heterogénea y obtener resultados más fiables en el análisis de endofenotipado tumoral. Los datos de la cohorte TCGA se recogen en la tabla suplementaria 2.

### **5.1.3 Cohorte de la base de datos *The Genotype-Tissue Expression (GTEx)***

Se obtuvo información de secuenciación de ARN de tejido colónico normal procedente de colon transversal de un total de 47 individuos (versión dbGAP phs000424.v8.p2) con edades no superiores a 50 años y sin historia personal de cáncer colorrectal (Tabla suplementaria 3).

## **5.2 SECUENCIACIÓN DE EXOMA COMPLETO Y PROCESAMIENTO DE DATOS**

### **5.2.1 Secuenciación de exoma completo de la cohorte TOGETHER**

Sobre el ADN obtenido de las 20 parejas de mucosa colónica normal y mucosa colónica tumoral de la cohorte TOGETHER se llevó a cabo WES. La secuenciación de exoma germinal (tejido normal) y CCR se realizó en el Centro Nacional de Análisis Genómico (CNAG-CRG), Barcelona. Se empleó el kit SureSelect Human All Exon v5 (Agilent Technologies, California, Estados Unidos) y la plataforma de secuenciación HiSeq 3000 (Illumina Inc., California, Estados Unidos) siguiendo los protocolos propios del centro, y de acuerdo a los estándares de calidad establecidos para NGS por el CNAG-CRG. La profundidad de lectura media de los exomas fue de 150X y la secuenciación de extremo emparejado 2x101 pares de bases (pb).

### **5.2.2 *Calling* y anotación de variantes: cohortes TOGETHER y TCGA (COAD y READ)**

Todos los procesos bioinformáticos recogidos en este apartado se llevaron a cabo utilizando los servidores del Centro de Supercomputación de Galicia (CESGA).

### 5.2.2.1 WES germinal: *calling* de variantes germinales y anotación

#### 5.2.2.1.1 Cohorte TOGETHER

Los procesos enumerados a continuación se realizaron de acuerdo a las recomendaciones de buenas prácticas vigentes en el momento de procesar los datos y recogidas en la web del *Genome Analysis Toolkit* (GATK)<sup>80</sup> del Broad Institute. La versión de GATK empleada en todo momento fue la 4.1.4.1.

Se evaluó la calidad de los archivos *FASTQ* resultantes de la secuenciación de exoma completo empleando la herramienta *FastQC*<sup>81</sup>. Seguidamente, estos archivos fueron alineados contra el genoma humano de referencia GRCh37/hg19 empleando la herramienta *BWA-MEM*<sup>82</sup>. Posteriormente se realizó el marcaje de duplicados y ordenamiento de las secuencias (herramientas de Picard<sup>83</sup> *MarkDuplicates* y *SortSam*) y se recalibraron las puntuaciones de calidad de las bases (herramientas de GATK<sup>80</sup> *BaseRecalibrator* y *ApplyBQSR*) para obtener los archivos *BAM* alineados.

A continuación, se empleó la herramienta de GATK<sup>80</sup> *HaplotypeCaller* para identificar variantes de nucleótido único e inserciones y deleciones de pequeño tamaño en línea germinal. Se seleccionó el modo multimuestra para aumentar la sensibilidad y reducir el número de errores que tienen lugar durante el procesamiento bioinformático de los datos. Para ello, a partir de los archivos *BAM* individuales de mucosa colónica normal se generó un archivo intermedio en formato *GVCF* (del inglés, *Genomic Variant Call Format*). Seguidamente, se unificaron los archivos *GVCF* individualizados y se generó un archivo *GVCF* multimuestra empleando la herramienta de GATK<sup>80</sup> *CombineGVCF* y se realizó el genotipado conjunto de todas las muestras de la cohorte TOGETHER con la herramienta de GATK<sup>80</sup> *GenotypeGVCFs* obteniendo un archivo *VCF* (del inglés, *Variant Call Format*) multimuestra con los registros de variantes de nucleótido único e inserciones y deleciones de cada una de las muestras a nivel germinal de los distintos pacientes. Sobre estos archivos *VCF* se realizó un filtrado estricto (*hard-filtering*) tanto para variantes de nucleótido único como inserciones y deleciones de

pequeño tamaño siguiendo los parámetros recomendados por el código de buenas prácticas de GATK<sup>80</sup> y empleando la herramienta *VariantFiltration*. A continuación, se generó un archivo *VCF* individual de cada paciente, se realizó una división de los sitios multialélicos y una normalización (*left-normalization*) de las posiciones empleando la herramienta *bcftools*<sup>84,85</sup>.

Por último, se realizó la anotación de variantes empleando las herramientas SnpEff 5.0e<sup>86</sup> y ANNOVAR 2020Jun<sup>787</sup> para agregar información actualizada de distintas bases de datos a nivel genético, clínico, poblacional y de impacto funcional.

#### 5.2.2.1.2 Cohorte TCGA (COAD y READ)

Los datos de WES germinal procedentes de sangre periférica y pertenecientes a los conjuntos de datos COAD y READ se habían secuenciado en una plataforma Illumina (Illumina Inc., California, Estados Unidos) empleando el kit SeqCap EZ HGSC VCRome (Roche, NimbleGen Inc., California, Estados Unidos). Los exomas presentan coberturas medias que oscilan entre 53X y 304,7X.

El procesamiento bioinformático de los datos se realizó empleando una *pipeline* desarrollada en la Fundación Pública Galega de Medicina Xenómica. Para ello se empleó de manera combinada la herramienta de Illumina *Dynamic Read Analysis for Genomics* (DRAGEN) y GATK v4.3.0.0<sup>80</sup> siguiendo el código de buenas prácticas del Broad Institute. La anotación de variantes se llevó a cabo con las herramientas SnpEff 5.0e<sup>86</sup> y ANNOVAR 2020Jun<sup>787</sup>.

#### 5.2.2.2 WES Tumor: *calling* de variantes somáticas y anotación

Los datos de exoma tumoral procedentes de la cohorte TOGETHER secuenciados empleando el kit *SureSelect Human All Exon v5* (Agilent Technologies, California, Estados Unidos) y los conjuntos de datos COAD y READ secuenciados empleando el kit SeqCap EZ HGSC VCRome (Roche, NimbleGen Inc., California, Estados Unidos) se procesaron en los servidores del CESGA siguiendo el código de buenas prácticas de GATK (versión 4.1.4.1)<sup>80</sup>. Todos los exomas presentaban una cobertura media superior a 100X.

La obtención de los archivos *BAM* alineados se realizó siguiendo el mismo protocolo descrito previamente en el apartado 5.2.2.1.1. Para el *calling* de variantes de nucleótido único se empleó la herramienta *Mutect2* y se incluyó el par de mucosa colónica normal en cada uno de los casos. Se controló la presencia de artefactos y el porcentaje de contaminación de mucosa colónica normal a través de las herramientas *LearnReadOrientationModel* y *CalculateContamination*. Por último, se llevó a cabo el filtrado de variantes con *FilterMutectCalls* para obtener un archivo *VCF* propio de cada tumor, se realizó una división de los sitios multialélicos y una normalización (*left-normalization*) de las posiciones empleando la herramienta *bcftools*<sup>84,85</sup>.

Para la anotación de variantes se emplearon las herramientas SnpEff 5.0e<sup>86</sup> y ANNOVAR 2020Jun7<sup>87</sup>.

### 5.3 RNA-SEQ Y PROCESAMIENTO DE DATOS: COHORTES TOGETHER, TCGA Y GTEx

El procesamiento de datos de RNA-seq se llevó a cabo utilizando muestras de tejido colónico normal y tumoral procedentes de las tres cohortes:

**Cohorte TOGETHER:** secuenciación del ARN extraído de muestras pareadas de mucosa colónica normal y CCR (n= 40) empleando la librería TruSeq Stranded mRNA (Illumina Inc., California, Estados Unidos) en la plataforma de secuenciación HiSeq3000 (Illumina Inc., California, Estados Unidos). Este proceso se llevó a cabo en el CNAG-CRG de acuerdo a los protocolos de calidad específicos del centro. Para cada muestra biológica se obtuvieron más de 50 millones de lecturas de extremo emparejado 2x76pb.

**Cohorte TCGA:** los datos de ARN procedentes de COAD y READ (TCGA) pertenecían a muestras congeladas de CCR (n= 26), según la información disponible estos fueron secuenciados en la plataforma Illumina (Illumina Inc., California, Estados Unidos) empleando una librería TruSeq no específica de hebra (Illumina Inc., California, Estados Unidos). Además, se seleccionaron datos de ARN de un total de 55 individuos diagnosticados de cáncer colorrectal con edades comprendidas entre los 51 y 65 años (Tabla suplementaria 4), secuenciados con la misma tecnología (versión dbGAP

phs000178.v11.p8), con el fin de aumentar el tamaño muestral para el análisis de Subgrupos Molecular Consenso que se explicará en el apartado 5.4.5. El número mínimo de lecturas detectado en las muestras fue de 38 millones de extremo único (*single-end*) de 48pb.

Cohorte GTEx: los datos de ARN de muestras congeladas de mucosa colónica normal corresponden a 47 individuos sanos sin antecedentes personales de CCR. De acuerdo a lo indicado en la base de datos, para su secuenciación se empleó el protocolo de librería TruSeq no específica de hebra (Illumina Inc., California, Estados Unidos) y la plataforma Illumina (Illumina Inc., California, Estados Unidos). La media de profundidad de lectura a la que las muestras fueron secuenciadas fue de 82,6 millones de lecturas de extremo único de 76pb.

El procesamiento bioinformático de los datos de RNA-seq llevado a cabo fue el mismo en todos los casos de las tres cohortes. A partir de los archivos *FASTQ* obtenidos de la secuenciación de ARN se estudiaron los parámetros de calidad de estos empleando *FastQC*<sup>81</sup>. A continuación, se empleó la herramienta STAR v2.7.0e<sup>88</sup> en la modalidad de dos pasos (*2-pass*) para llevar a cabo el alineamiento de los archivos *FASTQ* con el genoma humano de referencia GRCh37/hg19 y la anotación génica con GENCODE v19, en el momento del procesamiento se tuvo en cuenta el tipo de secuenciación y librería empleados en cada una de las cohortes. Posteriormente se realizó el ordenamiento de las secuencias con *samtools*<sup>85</sup> y marcaje de duplicados con la herramienta de Picard<sup>83</sup> *MarkDuplicates*. Por último, se analizó la calidad de los archivos resultantes y se obtuvieron las lecturas génicas para cada una de las muestras empleando la herramienta RNA-SeQC v2.4.2<sup>89</sup>.

Las secuencias de ARN alineadas fueron visualizadas en el visor genómico IGV (*Integrative Genomics Viewer*)<sup>90-93</sup>.

## 5.4 ENDOFENOTIPADO TUMORAL

### 5.4.1 Firma mutacional tumoral

A partir de los datos de WES se realizó el análisis de firma mutacional tumoral sobre los 46 tumores procedentes de las cohortes

TOGETHER (n= 20) y TCGA COAD y READ (n= 26). Para ello se emplearon los archivos *VCF* procedentes del *calling* de variantes somáticas (tumoraes) y se seleccionaron aquellas variantes exclusivas del tumor (Filter= PASS).

Se utilizó la herramienta *SigProfilerAssignment*<sup>94-96</sup> en Python v3.8. De las 46 muestras tumorales se extrajo la firma mutacional de nucleótido único empleando la base de datos COSMIC v3.3<sup>19</sup> y la versión del genoma GRCh37/hg19.

#### 5.4.2 Carga mutacional tumoral

A partir de los datos de WES se calculó la carga mutacional de los 46 tumores, procedentes de la cohorte TOGETHER y la base de datos TCGA, partiendo de los archivos *VCF* del *calling* somático y seleccionando aquellas variantes exclusivas de tumor (Filter= PASS).

Se empleó el paquete *pyTMB* v1.3.0<sup>97</sup> de Python v3.8 usando la versión del genoma GRCh37/hg19 y siguiendo las recomendaciones de los desarrolladores. El cálculo de mutaciones somáticas por megabase (Mb) se realizó en función del tamaño de la librería empleada en la secuenciación de cada una de las muestras.

#### 5.4.3 Identificación de genes driver en CCR

A partir de los datos de WES de CCR anotados se seleccionaron aquellas variantes exclusivas de tumor (Filter= PASS) que provocan un cambio en la secuencia aminoacídica o localizadas en las regiones de *splicing*, y que presentaban al menos un 5% de frecuencia del alelo variante (VAF). Los datos obtenidos para cada uno de los CCR se compararon con la lista de genes driver recogida en la base de datos IntOGen (última actualización del 1 de febrero de 2020)<sup>98</sup> empleando la herramienta *dffiltering* v0.2.7<sup>99</sup> en entorno Python v3.8.

Se consultaron las bases de datos de *hotspots* en cáncer Cancer Hotspots<sup>100</sup> y Cancer Genome Interpreter<sup>101</sup>.

Las variantes fueron visualizadas en el visor genómico IGV<sup>90-93</sup>.

#### 5.4.4 Análisis de mucosa colónica normal y tumoral

Se examinaron los datos de ARN de mucosa colónica normal (cohortes TOGETHER y GTEEx) y CCR (cohortes TOGETHER y



TCGA) de los individuos analizados a lo largo de este estudio para conocer cómo se distribuían de acuerdo a sus patrones de expresión. Los datos de ARN total se normalizaron por el método TMM y se aplicó una transformación  $\log_2$  a través del paquete de R *edgeR* v3.38.4 (versión de R 4.2.2)<sup>102-104</sup>. Se eliminaron también los efectos de las covariables asociadas a cohorte de procedencia (TOGETHER o TCGA), el método de conservación (RNA*later* o congelado), el sexo del individuo y el tipo de librería de secuenciación (hebra específica o no). Este análisis se centra en el estudio de las diferencias de expresión génicas, y no de isoformas, debido a la imprecisión que podría ocasionar el uso de secuenciación no específica de hebra (cohortes TCGA y GTEx en el análisis de estas últimas)<sup>105</sup>.

Los datos normalizados y corregidos se presentaron en una gráfica de componentes principales empleando el paquete de R *PCAtools* v2.8.0 (versión de R 4.2.2)<sup>106</sup>.

#### 5.4.5 Subgrupos Moleculares Consenso o CMS

Se estudió la clasificación CMS empleando los datos de ARN procedentes de muestras tumorales de 34 individuos diagnosticados de cáncer colorrectal a una edad igual o inferior a 50 años, que cumplieran los criterios de inclusión en el estudio y procedentes de las cohortes TOGETHER y TCGA (COAD y READ).

Para alcanzar un mínimo de 40 muestras y evitar un sesgo en las predicciones resultantes de la clasificación CMS, tal como recomienda la herramienta empleada para ello, se incluyeron datos de pacientes que no cumplieran los criterios de inclusión en el estudio. Se añadieron datos de ARN de cinco CCR de individuos con edad de diagnóstico igual o inferior a los 50 años, con firmas mutacionales indicativas de defectos en el sistema de reparación MMR y en el ED de la polimerasa  $\epsilon$  (202, COAD-14, COAD-16, COAD-18, READ-7). También se incluyó un grupo de muestras adicional compuesto por datos de ARN de CCR de 55 individuos pertenecientes a la base de datos TCGA (cohortes COAD y READ) con edades de diagnóstico comprendidas entre los 51 y 65 años.

Los datos de expresión totales de ARN se procesaron empleando el paquete de R *edgeR* v3.38.4<sup>102-104</sup> (versión de R 4.2.2). La

normalización de las muestras se llevó a cabo por el método TMM y transformación  $\log_2$ , y se corrigieron los efectos de *batch* asociados a la cohorte de procedencia (TOGETHER o TCGA), el método de conservación (RNA*later* o congelado) y el sexo del individuo. Los valores de expresión normalizados se emplearon para la clasificación en los subtipos moleculares consenso o CMS propuesta por Guinney y colaboradores<sup>33</sup> a través del paquete de R *CMSCaller* v0.99.2<sup>107</sup> (versión de R 3.4.4), con 1000 permutaciones y  $FDR < 0,05$ .

La distribución de los tumores se representó gráficamente con el paquete de R *PCATools* v2.8.0<sup>106</sup> (versión de R 4.2.2) a partir de los datos de expresión de ARN normalizados.

Además, se llevó a cabo un análisis de conjunto de genes (GSA, del inglés *Gene Set Analysis*) empleando la herramienta *CMSgsa* del paquete de R *CMSCaller* v0.99.2<sup>107</sup> y teniendo en cuenta la previa agrupación por CMS.

#### 5.4.6 Perfil epigenético de los CMS

Los perfiles epigenéticos tumorales pertenecientes a los subgrupos moleculares se estudiaron empleando los *arrays* de metilación de Illumina Infinium HumanMethylation450 BeadChip (cohorte TCGA) y Infinium MethylationEPIC BeadChip (cohorte TOGETHER). En primer lugar, a partir de los archivos no procesados de metilación (archivos idat) y empleando el paquete de R *ChAMP* 2.26.0<sup>108,109</sup> en la versión de R 4.2.2, se llevó a cabo un filtrado de sondas eliminando aquellas con un p-valor de detección  $>0,01$  en más del 10% de las muestras, se filtraron también sondas con  $<3$  *beads* en más del 5% de las muestras, sondas no ubicadas en regiones no CpG, relacionadas con SNPs empleando la lista de Zhou y colaboradores<sup>110</sup>, sondas que alineaban en múltiples posiciones recogidas por Nordlund y colaboradores<sup>111</sup> y aquellas localizadas en los cromosomas sexuales. Se generaron las matrices de valores  $\beta$  (valores comprendidos entre 0 y 1) para obtener los niveles de metilación de los distintos sitios CpG para cada uno de los *arrays*. A continuación, se combinaron ambos *arrays* empleando la función *combineArrays* del paquete de R *minfi* versión 1.42.0<sup>112</sup>, para obtener aquellas sondas comunes en ambos *arrays* que incluyen  $>90\%$  de las presentes en Illumina Infinium

HumanMethylation450 BeadChip<sup>113</sup>. Las diferencias técnicas entre los tipos de sondas de los *arrays* (sondas tipo I y sondas tipo II) se normalizó empleando el método BMIQ (*beta-mixture quantile normalization*)<sup>114</sup> implementado en *ChAMP*<sup>108,109</sup>. Después del procesamiento, los posteriores análisis se realizaron sobre las 344.826 sondas restantes.

Los efectos de *batch* asociados a la cohorte de procedencia (TOGETHER o TCGA), método de conservación (*RNA later* o congelado), localización tumoral, sexo y edad de los pacientes, así como los asociados al *barcode* se corrigieron empleando la función *champ.runCombat* de *ChAMP*<sup>108,109</sup>.

El análisis de las posiciones diferencialmente metiladas (DMP, *differentially methylated position*) para cada uno de los CMS se llevó a cabo con la función *ChAMP.DMP* del paquete de R *ChAMP*<sup>108,109</sup>. Se empleó la corrección de Benjamini-Hochberg para ajuste de test múltiple con un valor de significancia de p-valor ajustado <0,05.

Finalmente, para cada uno de los CMS, se realizó un análisis de enriquecimiento funcional en rutas KEGG (*Kyoto Encyclopedia of Genes and Genomes*)<sup>115</sup> para aquellos genes obtenidos de los análisis de posiciones diferencialmente metiladas que presentaban valores de  $|\log \text{fold change}| \geq 0,10$ . Se empleó para ello la herramienta *g:Profiler*<sup>116</sup>. Se consideraron significativas aquellas rutas KEGG<sup>115</sup> con un p-valor ajustado <0,05 después de la corrección por el método Benjamini-Hochberg.

## 5.5 INTEGRACIÓN DE ÓMICAS

### 5.5.1 Filtrado de variantes germinales

De las variantes resultantes del procesamiento bioinformático de los datos WES de mucosa colónica normal de las cohortes de descubrimiento TOGETHER y TCGA se seleccionaron aquellas que cumplían los filtros de calidad (Filter= PASS), con una profundidad de lectura mínima de 20X, que presentasen al menos el 20% del alelo alterado y localizadas en regiones solapantes de ambos diseños de librerías (SureSelect Human All Exon v5 de Agilent Technologies y SeqCap EZ HGSV VCRome de Roche presentan un porcentaje de

solapamiento del 73,21%, calculado mediante la herramienta *bedops* v2.4.41<sup>117</sup>). Además, se priorizaron aquellas variantes con frecuencia poblacional rara (gnomAD v2.1.1\_genome\_NFE  $\leq 0,1\%$  para variantes heterocigotas y gnomAD v2.1.1\_genome\_NFE  $\leq 1\%$  para variantes homocigotas y posibles heterocigotas compuestas) presentes en la región exónica (tipo: *startloss*, *stopgain*, *stoploss*, *frameshift*, *inframe*, *missense* y sinónimas) y en posiciones intrónicas hasta  $\pm 20$ . También se estudió la existencia de variantes patogénicas germinales presentes en genes de predisposición hereditaria a cáncer colorrectal (Tabla 1). El filtrado de variantes se llevó a cabo con la herramienta *dffiltering* v0.2.7<sup>99</sup> en el entorno Python v3.8. Las variantes fueron visualizadas en el visor genómico IGV<sup>90-93</sup>.

**Tabla 1. Genes relacionados con susceptibilidad hereditaria a cáncer colorrectal.** CCR: cáncer colorrectal, CMMRD: síndrome de Deficiencia Constitucional de Reparación de Errores de Apareamiento, AD: herencia autosómica dominante, AR: herencia autosómica recesiva.

| Gen           | Síndrome   | Herencia |
|---------------|--|----------|
| <i>MLH1</i>   | Síndrome de Lynch  | AD       |
|               | CMMRD  | AR       |
| <i>MSH2</i>   | Síndrome de Lynch  | AD       |
|               | CMMRD  | AR       |
| <i>MSH6</i>   | Síndrome de Lynch  | AD       |
|               | CMMRD  | AR       |
| <i>PMS2</i>   | Síndrome de Lynch  | AD       |
|               | CMMRD  | AR       |
| <i>EPCAM</i>  | Síndrome de Lynch  | AD       |
|               | CMMRD  | AR       |
| <i>APC</i>    | Poliposis adenomatosa familiar   | AD       |
| <i>AXIN2</i>  | Síndrome oligodontia-CCR   | AD       |
| <i>MUTYH</i>  | Poliposis asociada a <i>MUTYH</i>                                      | AR       |
| <i>NTHL1</i>  | Poliposis asociada a <i>NTHL1</i>                                      | AR       |
| <i>POLE</i>   | Poliposis adenomatosa asociada a la lectura de prueba de la polimerasa | AD       |
| <i>POLD1</i>  | Poliposis adenomatosa asociada a la lectura de prueba de la polimerasa | AD       |
| <i>GREM1</i>  | Poliposis mixta hereditaria  | AD       |
| <i>MSH3</i>   | Poliposis asociada a <i>MSH3</i>                                       | AR       |
| <i>STK11</i>  | Síndrome de Peutz-Jeghers  | AD       |
| <i>SMAD4</i>  | Poliposis juvenil  | AD       |
| <i>BMPR1A</i> | Poliposis juvenil  | AD       |
| <i>RNF43</i>  | Poliposis serrada  | AD       |
| <i>PTEN</i>   | Síndromes hamartomatosos asociados a <i>PTEN</i>                       | AD       |
| <i>TP53</i>   | Síndrome de Li-Fraumeni  | AD       |
| <i>MBD4</i>   | Síndrome neoplásico asociado a <i>MBD4</i>                             | AR       |
| <i>RPS20</i>  | Cáncer colorrectal familiar tipo X                                     | AD       |

### 5.5.2 Aproximación *All vs One*

Se realizó el análisis de expresión diferencial de manera individual para cada una de las 34 muestras tumorales que cumplían los criterios de inclusión del estudio (pacientes con diagnóstico de cáncer colorrectal  $\leq 50$  años, sin alteración del sistema de reparación MMR). Para este análisis se enfrentaron los datos de ARN de cada uno de los tumores al conjunto de datos de mucosa colónica normal (67 muestras de mucosa colónica normal procedentes de la cohorte TOGETHER y la base de datos GTEx). Se empleó el paquete de R *edgeR* v3.38.4<sup>102-104</sup> (versión de R 4.2.2), la normalización TMM y se controlaron las covariables de cohorte de procedencia (TOGETHER, TCGA o GTEx), método de conservación (RNA*later* o congelado), sexo del individuo, tipo de librería de secuenciación (hebra específica o no). Para el análisis de expresión diferencial se utilizó la aproximación estadística de estimación por cuasiverosimilitud (*quasi-likelihood*) sobre los datos de ARN total. Se seleccionaron aquellos genes con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  resultantes de la estimación por cuasiverosimilitud.

La lista de genes con expresión diferencial significativa obtenida para cada una de las muestras se comparó con la lista de variantes germinales filtradas para cada uno de los individuos. También se compararon las muestras pertenecientes a cada CMS entre sí.

Además, se llevó a cabo la búsqueda de segundos eventos mutacionales a nivel somático en aquellos genes con expresión diferencial significativa y que presentaban variantes germinales tras aplicar los filtros de priorización de variantes. Para obtener la lista de variantes somáticas coincidentes en cada uno de los individuos se empleó la herramienta *dffiltering* v0.2.7<sup>99</sup> en Python v3.8.

## 5.6 COHORTE DE RÉPLICA

Los 10 genes candidatos seleccionados en la fase de descubrimiento se estudiaron en datos de WES germinal de sangre periférica de un total de 39 pacientes (20 procedentes de la base de datos TCGA y 19 pertenecientes a una base de datos interna) con diagnóstico de cáncer colorrectal a una edad igual a los 50 años o inferior, de los que no se dispone de datos de RNA-seq.

Cohorte TCGA: 20 pacientes no relacionados con y sin historia familiar oncológica procedentes de las cohortes COAD y READ (versión dbGAP phs000424.v8.p2) (Tabla suplementaria 5). Sobre los tumores colorrectales de estos pacientes se llevó a cabo el procesamiento de datos de secuenciación descrito previamente en el apartado 5.2.2.2 y el estudio de firma mutacional tumoral y carga mutacional tumoral (ver apartados 5.4.1 y 5.4.2, respectivamente) con el fin de validar las características fenotípicas de los CCR reportadas en la base de datos TCGA y conocer las características moleculares de aquellos tumores de los que no se disponía de información.

Cohorte interna: 19 pacientes pertenecientes a 18 familias no relacionadas, con y sin historia familiar oncológica y que presentan tumores colorrectales con expresión conservada de las proteínas reparadoras MMR (Tabla suplementaria 5).

Los datos de WES germinal pertenecientes a los conjuntos de datos COAD y READ, al igual que en la fase de descubrimiento, se secuenciaron en una plataforma Illumina (Illumina Inc., California, Estados Unidos) con el kit SeqCap EZ HGSC VCRome (Roche, NimbleGen Inc., California, Estados Unidos). Los datos se alinearon contra el genoma humano de referencia GRCh37/hg19, se emplearon las herramientas DRAGEN y GATK v4.3.0.0<sup>80</sup> para el *calling* de variantes de nucleótido único e inserciones y deleciones de pequeño tamaño y se llevó a cabo la anotación con las herramientas SnpEff 5.0e<sup>86</sup> y ANNOVAR 2020Jun7<sup>87</sup>, de acuerdo al código de buenas prácticas de Broad Insitute y la *pipeline* desarrollada en la Fundación Pública Galega de Medicina Xenómica. Los valores de cobertura media oscilaban entre los 50X y 300X.

La secuenciación de los datos germinales procedentes de la base de datos interna se realizó en un secuenciador 5500xl SOLiD<sup>TM</sup> (Thermo Fisher Scientific, Massachusetts, Estados Unidos) empleando el kit SureSelect Human All Exon v5 (Agilent Technologies, California, Estados Unidos). El alineamiento de las secuencias contra el genoma humano de referencia GRCh37/hg19 se realizó empleando una suite completa de análisis de Lifescope, el *calling* de variantes de pequeño tamaño se llevó a cabo a través de los *softwares* LifeScope y GATK 3.0 y se anotaron las variantes detectadas con las herramientas SnpEff

5.0e<sup>86</sup> y ANNOVAR 2020Jun7<sup>87</sup>. El procesamiento bioinformático de los datos se realizó utilizando una *pipeline* desarrollada en la Fundación Pública Galega de Medicina Xenómica. La cobertura media en todos los casos fue superior a 50X.

Posterior al procesamiento bioinformático se seleccionaron aquellas variantes que cumplieran los filtros de calidad, con una profundidad de lectura mínima de 20X y al menos el 20% del alelo alterado y que se encontrasen localizadas en las regiones solapantes de ambos diseños de librerías (73,21% de solapamiento, al igual que en la fase de descubrimiento, pues se emplean las mismas librerías). Además de los genes candidatos seleccionados en la fase de descubrimiento se estudió la existencia de variantes patogénicas o (probablemente) patogénicas en genes previamente relacionados con la tumorigénesis colorrectal recogidos en la tabla 1. Se priorizaron las variantes que presentaron una frecuencia poblacional en gnomAD v2.1.1\_genome\_NFE  $\leq 0,1\%$  en heterocigotas y  $\leq 1\%$  para homocigotas y posibles heterocigotas compuestas presentes en las regiones exónicas (tipo: *startloss*, *stopgain*, *stoploss*, *frameshift*, *inframe*, *missense* y sinónimas) y en posiciones intrónicas hasta  $\pm 20$ . La selección de variantes *missense* se realizó en base a su impacto funcional, de acuerdo a los siguientes predictores *in silico*: SIFT *Deleterious* (D), PolyPhen-2 *Probably Damaging* (D) o *Possibly Damaging* (P), MutationTaster *Disease Causing Automatic* (A) o *Disease Causing* (D), PROVEAN  $< -2.5$  y CADD\_pred  $\geq 20$ , cuya combinación presenta un elevado grado de concordancia<sup>118</sup>. Se priorizaron aquellas variantes *missense* en las cuales al menos cuatro de los predictores indicaron una posible afectación a la funcionalidad proteica. También se consultaron las predicciones de SpliceAI de las distintas variantes seleccionadas para conocer su posible afectación a este proceso. SIFT, PolyPhen-2, MutationTaster y PROVEAN predicen el efecto funcional del cambio ocurrido<sup>119-122</sup>, CADD (*Combined Annotation-Dependent Depletion*) es un metapredictor que integra métricas de conservación, impacto funcional, información sobre elementos reguladores, transcritos y niveles de expresión<sup>123</sup>. SpliceAI consiste en una red neuronal residual que predice la probabilidad de cada posición de ser un sitio donador de *splicing*, aceptor de *splicing* o ninguno<sup>124</sup>.

El filtrado de variantes se llevó a cabo empleando la herramienta *dffiltering* v.0.2.7<sup>99</sup> en el entorno de Python v3.8. Todas las variantes se visualizaron en el visor genómico IGV<sup>90-93</sup>.

### 5.7 APECTOS ÉTICOS

Este estudio respeta los principios fundamentales establecidos en la Declaración de Helsinki y el Acuerdo de Oviedo y otras normativas que aplican en el ámbito de la investigación biomédica, la protección de datos de carácter personal y la bioética, especialmente la ley de Investigación Biomédica y las regulaciones sobre Biobancos (41/2001, (LIB) 14/2007, R(2006)4, entre otras). El protocolo de estudio cumple con el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de esos datos, así como con la legislación nacional. También se respetan los principios establecidos por el *International Cancer Genome Consortium* (ICGT) y el *International Rare Diseases Research Consortium* (IRDiRC). Asimismo, el proyecto “Integración de datos ómicos germinales y tumorales para la identificación de nuevos genes de susceptibilidad hereditaria a cáncer colorrectal (TOGETHER) (PI17/00509)” fue aprobado por el Comité Ético de Investigación Clínica de Galicia (código de registro 2017/355) (Anexo III).





# RESULTADOS



## 6 RESULTADOS

### 6.1 ENDOFENOTIPADO TUMORAL: CARACTERIZACIÓN MOLECULAR

#### 6.1.1 Firma y carga mutacional tumoral

La firma mutacional de variantes de nucleótido único de los tumores obtenida a través de la herramienta de Python v3.8 *SigProfiler Assignment*<sup>94-96</sup> empleando la base de datos COSMIC v3.3<sup>19</sup>, mostró la existencia de heterogeneidad tumoral dentro de las cohortes de estudio TOGETHER y TCGA de la fase de descubrimiento (Figura 5 y tabla suplementaria 6).

Se observó la presencia de firmas mutacionales planas (SBS1, SBS5 y SBS40) en la totalidad de las muestras de cáncer colorrectal, que en la mayoría de los casos representaba de forma conjunta un porcentaje elevado, oscilando entre el 8,46% en el tumor READ-7 y el 100% (tumores 108, 209, COAD-2, COAD-7 y COAD-11). También destacó la existencia de patrones mutacionales asociados a defectos en el sistema de reparación MMR (firmas SBS6, SBS14, SBS15, SBS26 y SBS44) en un total de dieciocho tumores, cuatro de los cuales presentaban un porcentaje superior al 30% (202, COAD-14, COAD-16 y COAD-18), mientras que en los restantes los valores se situaban entre el 5,76% y el 24,86%. Se observaron otros patrones mutacionales compatibles con defectos en sistemas de reparación celulares en diecisiete tumores, como las firmas que indican defectos en el funcionamiento de distintas polimerasas (SBS9, SBS10a, SBS10b, SBS10d y SBS14), APOBEC (SBS2 y SBS13) o la firma SBS30, asociada a defectos en el sistema de reparación por escisión de bases debido a mutaciones en el gen *NTHL1*, en un rango porcentual bastante amplio que va del 3,41% de la firma SBS2 en el CCR de la paciente COAD-17 al 41,67% de SBS10b del CCR del paciente READ-7.

Veintitrés de los tumores presentaron firmas mutacionales relacionadas con factores ambientales. Tres CCR de la cohorte COAD (COAD-4, COAD-10 y COAD-15) y dos de la cohorte READ (READ-

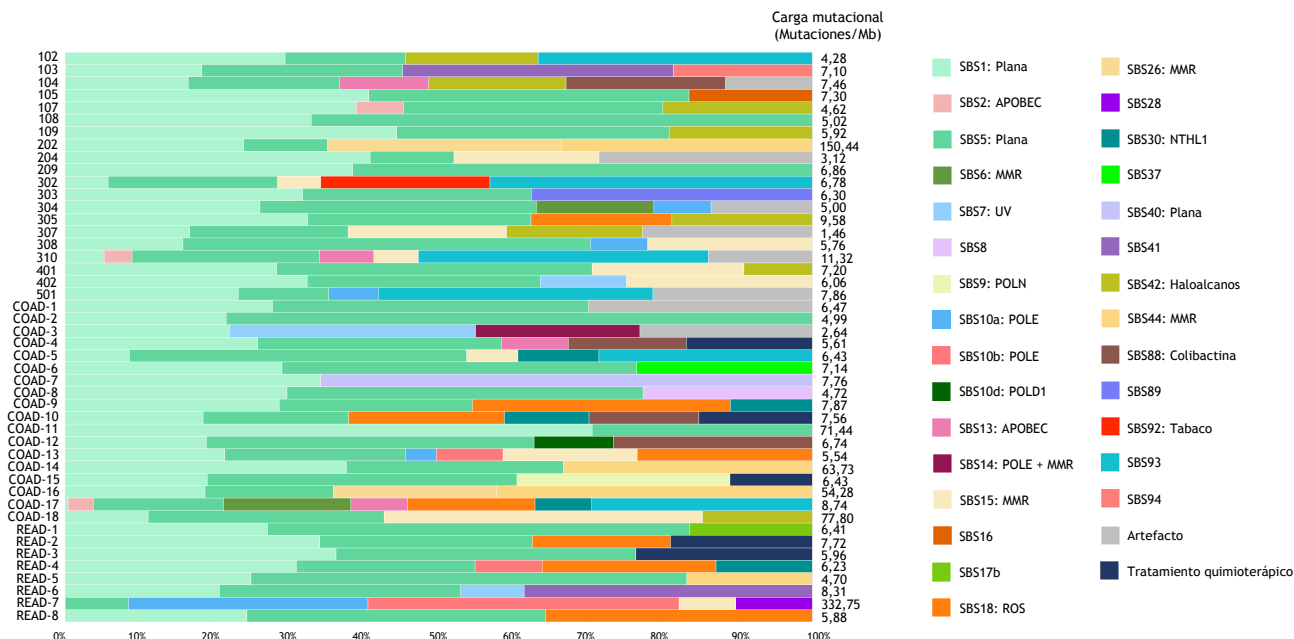
2 y READ-3) mostraron patrones mutacionales asociados con el tratamiento con agentes quimioterápicos con porcentajes que oscilaban entre el 11,03% y el 23,62%. Se encontraron también firmas con etiologías asociadas a otros factores ambientales externos: SBS7 (que comprende las firmas SBS7a, SBS7b y SBS7d) asociada a la exposición a la luz ultravioleta en los tumores de los pacientes 402 (11,56%), COAD-3 (32,97%) y READ-6 (8,52%); exposición a haloalcanos (SBS42) en siete de los CCR pertenecientes a la cohorte TOGETHER (102, 104, 107, 109, 305, 307 y 401) y uno de la cohorte COAD (COAD-18) entre el 9,16% y 20%; y en un caso (302) se detectó la firma mutacional SBS92 asociada al hábito tabáquico en un 22,66%. El patrón mutacional asociado a la colibactina (SBS88), un factor ambiental endógeno, estaba presente en los CCR de los pacientes 104, COAD-4, COAD-10 y COAD-12, con valores entre el 14,69% y 26,67%. (Figura 5 y tabla suplementaria 6). Ocho de los CCR (305, COAD-9, COAD-10, COAD-13, COAD-17, READ-2, READ-4 y READ-8) presentaron la firma mutacional SBS18 asociada a daños en el ADN causados por especies reactivas del oxígeno, en un rango porcentual del 17,05% al 35,68% (Figura 5 y tabla suplementaria 6).

Además, en catorce de los tumores se manifestó la existencia de firmas mutacionales cuya etiología es desconocida hasta el momento (SBS8, SBS16, SBS17b, SBS28, SBS37, SBS41, SBS89, SBS93 y SBS94), en la mayoría de los casos con porcentajes superiores al 20%. La más común de estas fue la SBS93 detectada en seis de los CCR (102, 302, 310, 501, COAD-5 y COAD-17), en todos los casos su porcentaje era superior al 28%. En el tumor del paciente 103 destacó la presencia de dos firmas de etiología desconocida, SBS41 (36,20%) y SBS94 (18,64%) (Figura 5 y tabla suplementaria 6).

La carga mutacional de los tumores expresada como el número de mutaciones por Mb existentes en el exoma tumoral, obtenida a través del paquete de Python v3.8 *pyTMB* versión 1.3.0<sup>97</sup>, mostró una mayor homogeneidad dentro de la cohorte. En la mayoría de los tumores colorrectales este valor se encontraba por debajo de las 10 mutaciones/Mb. Siete de los tumores presentaban una carga mutacional por encima de este umbral, y únicamente los casos 202 y READ-7 superaron las 100 mutaciones/Mb con 150,44 y 332,75 mutaciones/Mb,

respectivamente. Destacan también los tumores 204, 307 y COAD-3 cuya carga mutacional es inferior a 4 mutaciones/Mb (Figura 5).

Se eliminaron de posteriores análisis aquellas muestras tumorales que presentaron firmas mutacionales acordes a la existencia de tratamiento quimioterápico (COAD-4, COAD-10, COAD-15, READ-2 y READ-3), pues la ausencia de este tipo de profilaxis era un requisito de selección de los pacientes a estudio. Se mantuvieron aquellos tumores que presentaban firmas y cargas mutacionales acordes con la existencia de síndromes de predisposición hereditaria a cáncer, como es el caso de los patrones mutacionales que indican defectos en el sistema de reparación MMR y defectos en el ED de la polimerasa  $\epsilon$  en tumores con elevada carga mutacional (tumores 202, COAD-14, COAD-16, COAD-18, READ-7), con el fin de obtener una clasificación CMS más precisa (de acuerdo a las recomendaciones de la herramienta utilizada para tal fin).



**Figura 5. Patrón de firmas mutacionales de nucleótido único (COSMIC v3.3<sup>19</sup>) y carga mutacional tumoral de los 46 tumores que componen la fase de descubrimiento procedentes de las cohortes TOGETHER y TCGA (COAD y READ). Firmas mutacionales presentes: firmas planas (SBS1, SBS5, SBS40), actividad de familia APOBEC de citidina desaminasas (SBS2, SBS13), defectos en sistema MMR (SBS6, SBS15, SBS26, SBS44), exposición a luz UV (SBS7: SBS7a, SBS7b y SBS7d), hipermutación somática de polimerasa  $\mu$  (SBS9), mutación en ED polimerasa  $\epsilon$  (SBS10a, SBS10b) y polimerasa  $\delta$ 1 (SBS10d), mutación en ED polimerasa  $\epsilon$  y defectos en sistema MMR (SBS14), daño por especies reactivas del oxígeno (SBS18), defectos en BER por alteraciones bialélicas en NTHL1 (SBS30), exposición a haloalcanos (SBS42), exposición a colibactina (SBS88), hábito tabáquico (SBS92) y etiología desconocida (SBS8, SBS16, SBS17b, SBS28, SBS37, SBS41, SBS89, SBS93, SBS94).**

### 6.1.2 Subgrupos Moleculares Consenso o CMS

Previo al estudio de los Subgrupos Moleculares Consenso definidos por Guinney y colaboradores<sup>33</sup> se llevó a cabo el estudio de los datos brutos de expresión procedentes de tejido normal (cohortes TOGETHER y GTEx) y de CCR (cohortes TOGETHER y TCGA COAD y READ) para conocer la distribución de los tejidos procedentes de mucosa colónica normal y CCR a estudio de los individuos con edades hasta los 65 años. Los datos de ARN totales se normalizaron y se eliminaron los efectos de la covariables. Por último, se representó la distribución de los datos en un gráfico de componentes principales, observándose que dos de las muestras inicialmente señaladas como CCR se agrupaban más próximas al conjunto de datos de mucosa colónica normal (Figura 6). Estos casos se corresponden con los tumores de los pacientes de la cohorte TOGETHER 204 y 307, los cuales fueron eliminados para posteriores análisis.

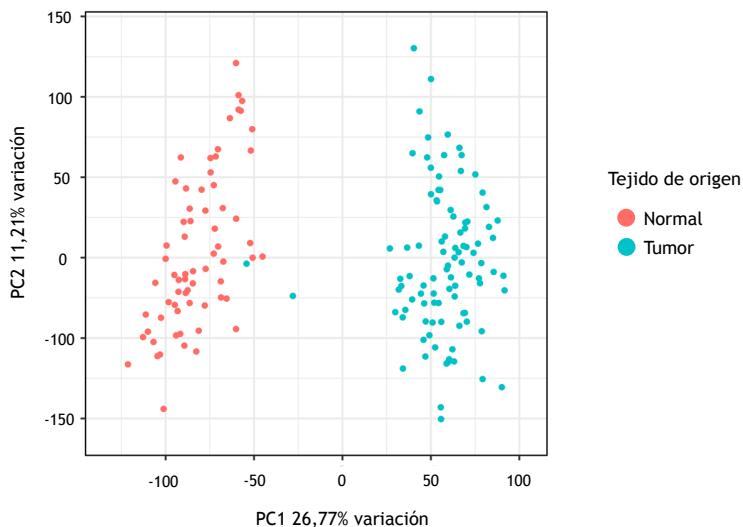


Figura 6. Gráfica de componentes principales que muestra la distribución de las muestras de mucosa colónica normal y CCR y su agrupación atendiendo a patrones de expresión.

Con los datos de ARN propios de CCR se procedió al análisis de los CMS empleando la herramienta *CMSCaller* v0.99.2<sup>107</sup>. Los

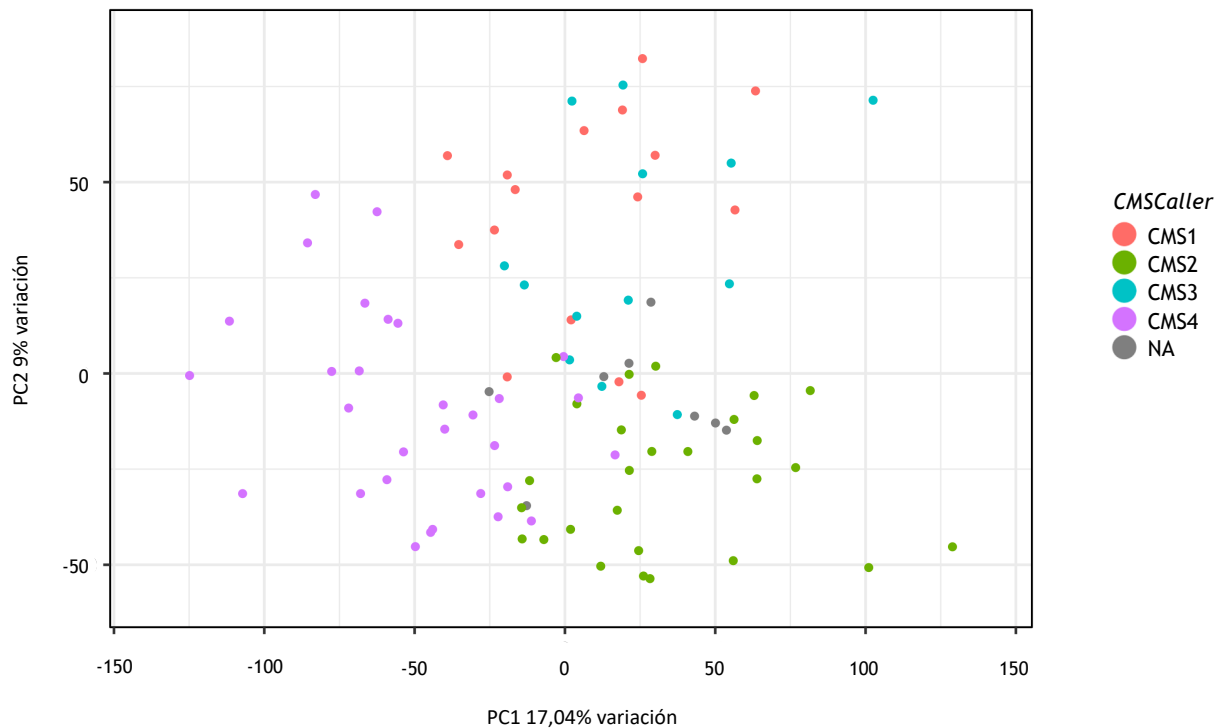


distintos CCR seguían una distribución heterogénea: 15 de los tumores fueron clasificados como CMS1, 26 como CMS2, 13 como CMS3, 30 como CMS4, y 10 de los tumores no fueron adjudicados a ninguno de los Subgrupos Moleculares Consenso (Tabla suplementaria 7). La disposición de los tumores en una gráfica de componentes principales mostró que los CCR no se dividían en grupos estancos, sino que se distribuían a lo largo del espacio mostrando una ligera agrupación. Se observó también que los tumores pertenecientes a los CMS1 y CMS3 se encuentran más próximos entre sí, al igual que ocurre con los tumores de los subgrupos CMS2 y CMS4 (Figura 7).

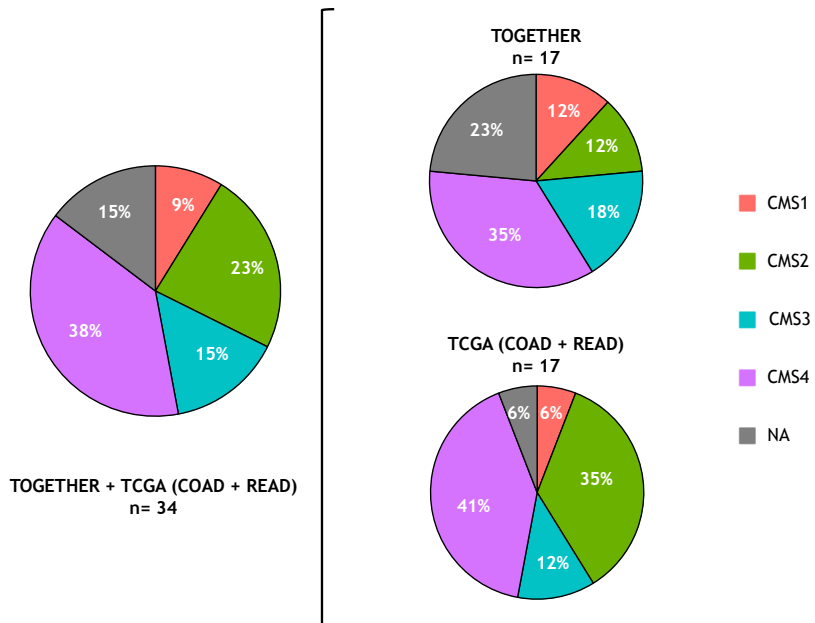
La clasificación de los tumores de los pacientes en los que se centra este trabajo (individuos con diagnóstico de CCR  $\leq 50$  años que cumplían criterios de inclusión en el estudio,  $n = 34$ ) reveló una distribución heterogénea (Figura 8). El 9% de los tumores se clasificó como CMS1 (CCR de los pacientes 105, 401 y READ-1), el 23% como CMS2 (tumores de los pacientes 303, 310, COAD-1, COAD-2, COAD-12, COAD-13, COAD-17 y READ-5), un 15% de los tumores pertenecía al CMS3 (CCR de los pacientes 104, 107, 108, COAD-6 y READ-6), y el 38% de los tumores se clasificó como CMS4 (tumores de los pacientes 102, 209, 302, 304, 305, 308, COAD-3, COAD-5, COAD-7, COAD-8, COAD-9, COAD-11 y READ-8). El 15% de los tumores restantes (tumores de los pacientes 103, 109, 402, 501 y READ-4) no se clasificó dentro de ninguno de los Subgrupos Moleculares Consenso. El CMS4 es el subgrupo mayoritario tanto en los tumores de la cohorte TOGETHER (35%) como en los procedentes de TCGA (41%) (Figura 8). Excepto en CMS3, la distribución del resto de subgrupos entre ambas cohortes es más variable (Figura 8).

Se realizó también un análisis GSA para cada uno de los CMS utilizando la función *CMSgsa* dentro del paquete de R *CMSCaller* v0.99.2<sup>107</sup>. En este se destacó que en los CCR del subgrupo CMS1 se encuentran sobreexpresados los conjuntos de genes relacionados con MSI, reparación del ADN, *MYC*, ciclo celular, glicólisis y ácidos grasos, mientras que la expresión de los genes asociados a *HNF4A*, *MSS* y *CDX2* se encontraba reducida (Figura 9 y tabla suplementaria 8). En el CMS2 se detectó principalmente un aumento de expresión de los genes relacionados con *MSS*, por el contrario, se redujo en el

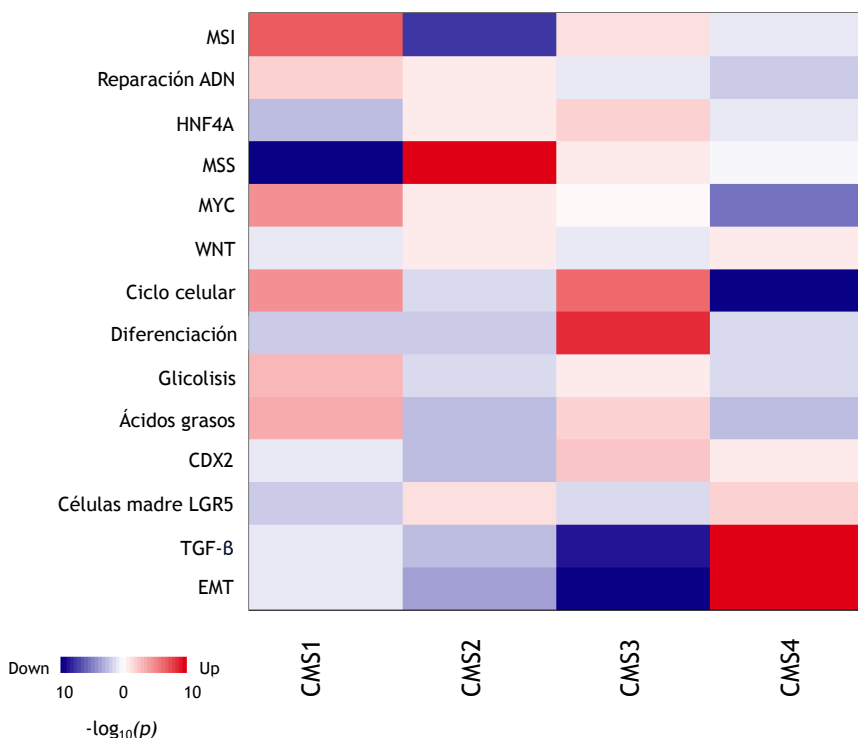
conjunto de genes asociados a MSI (Figura 9 y tabla suplementaria 9). Los grupos de genes relacionados con TGF- $\beta$  y EMT se mostraron con expresión reducida de manera significativa en el caso del grupo de tumores asignados como CMS3, donde aumentó la expresión de los genes asociados a diferenciación y ciclo celular (Figura 9 y tabla suplementaria 10). Por el contrario, los genes relacionados con TGF- $\beta$  y EMT están sobreexpresados en los tumores CMS4, mientras que los asociados a *MYC* y ciclo celular presentan una expresión reducida de manera significativa (Figura 9 y tabla suplementaria 11).



**Figura 7. Distribución de acuerdo a expresión génica de los CCR pertenecientes a distintos CMS en una gráfica de componentes principales. CMS1: Subgrupo Molecular Consenso 1, CMS2: Subgrupo Molecular Consenso 2, CMS3: Subgrupo Molecular Consenso 3, CMS4: Subgrupo Molecular Consenso 4, NA: no clasificado.**



**Figura 8.** Distribución por CMS de los CCR de individuos  $\leq 50$  años en el conjunto de datos (cohortes TOGETHER y TCGA) y en cada una de las cohortes de manera individual. CMS1: Subgrupo Molecular Consenso 1, CMS2: Subgrupo Molecular Consenso 2, CMS3: Subgrupo Molecular Consenso 3, CMS4: Subgrupo Molecular Consenso 4, NA: no clasificado.



**Figura 9. Representación en mapa de calor (*HeatMap*) de los resultados del análisis GSA de cada CMS.** El color rojo representa sobreexpresión de los genes de esa ruta y el color azul descenso en la expresión de los genes de una ruta, la saturación de color indica el aumento de la significancia estadística. MSI: inestabilidad de microsatélites; MSS: estabilidad de microsatélites; EMT: transición epitelio-mesenquimal, CMS1: Subgrupo Molecular Consenso 1, CMS2: Subgrupo Molecular Consenso 2, CMS3: Subgrupo Molecular Consenso 3, CMS4: Subgrupo Molecular Consenso 4.

### 6.1.3 Perfil epigenético de los CMS

Para investigar los factores epigenéticos asociados a cada uno de los cuatro CMS, se llevó a cabo un análisis de metilación diferencial entre los tumores de los pacientes pertenecientes a cada uno de los Subgrupos Moleculares Consenso.

En los tumores pertenecientes al CMS1, se observaron 32.256 DMPs de manera significativa. El 35,59% de estas DMPs estaban hipermetiladas mientras que el 64,41% de las DMPs estaban

hipometiladas. De las 11.479 posiciones que presentaron hipermetilación (con un valor máximo de metilación detectado de  $\log_{FC}= 0,66$ ), 5.922 posiciones se localizaban en islas CpG a lo largo del genoma. El análisis de enriquecimiento mostró que los genes con posiciones diferencialmente metiladas estaban implicados en rutas KEGG relacionadas con cáncer, como es el caso de hsa05200 (*Pathways in cancer*), hsa04151 (*PI3K-Akt signaling pathway*) o hsa04010 (*MAPK signaling pathway*), entre otras (Figura 10 y tabla suplementaria 12).

En el grupo de tumores CMS2 se detectaron 4.253 DMPs diferencialmente metiladas, el 49,50% de ellas hipermetiladas y el 50,50% hipometiladas. Del total de posiciones que mostraron hipermetilación (2.105), 986 se localizaban en islas CpG con un valor máximo de metilación de  $\log_{FC}= 0,30$ . El análisis de rutas mostró de manera significativa el enriquecimiento de DMPs diferencialmente metiladas localizadas en genes relacionados con la ruta de señalización de las MAPK (ruta KEGG hsa04010), como única ruta previamente implicada en la tumorigénesis colorrectal (Figura 11 y tabla suplementaria 13).

En los tumores agrupados bajo el CMS3 12.353 DMPs presentaron alteraciones significativas en la metilación, el 75,76% de las posiciones estaban hipermetiladas y el 24,24% hipometiladas. Del total de posiciones que presentaron hipermetilación (9.358), 3.006 estaban localizadas en islas CpG a lo largo del genoma (valor de metilación máximo detectado:  $\log_{FC}= 0,50$ ). El análisis de enriquecimiento de rutas mostró la presencia de genes alterados en la ruta KEGG de señalización de las MAPK, asociada a CCR (Figura 12 y tabla suplementaria 14).

Por último, en el caso del CMS4, 29.930 DMPs mostraron una alteración significativa en los niveles de metilación, al igual que en el resto de subgrupos se detectó un mayor número de posiciones hipermetiladas (64,92%) y un menor porcentaje de DMPs hipometiladas (35,08%). Solo 1.506 posiciones hipermetiladas se localizaban en islas CpG, con un valor máximo de metilación detectado de  $\log_{FC}= 0,26$ . El análisis de enriquecimiento mostró que los genes con posiciones diferencialmente metiladas estaban implicados en rutas

KEGG relacionadas CCR como hsa04151 (*PI3K-Akt signaling pathway*), hsa04010 (*MAPK signaling pathway*) y hsa05200 (*Pathways in cancer*) (Figura 13 y tabla suplementaria 15).

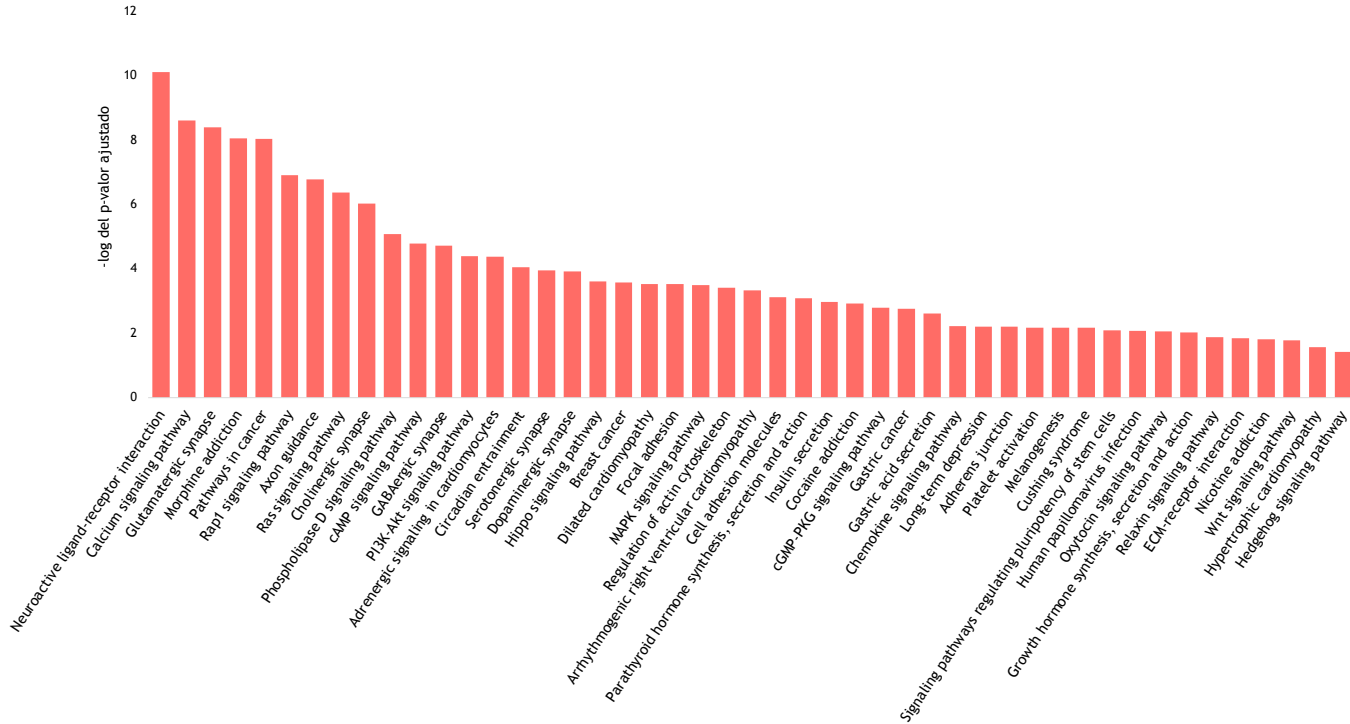


Figura 10. Gráfico de barras del análisis de enriquecimiento sobre rutas de señalización KEGG de las DMPs diferencialmente metiladas en los tumores CMS1.



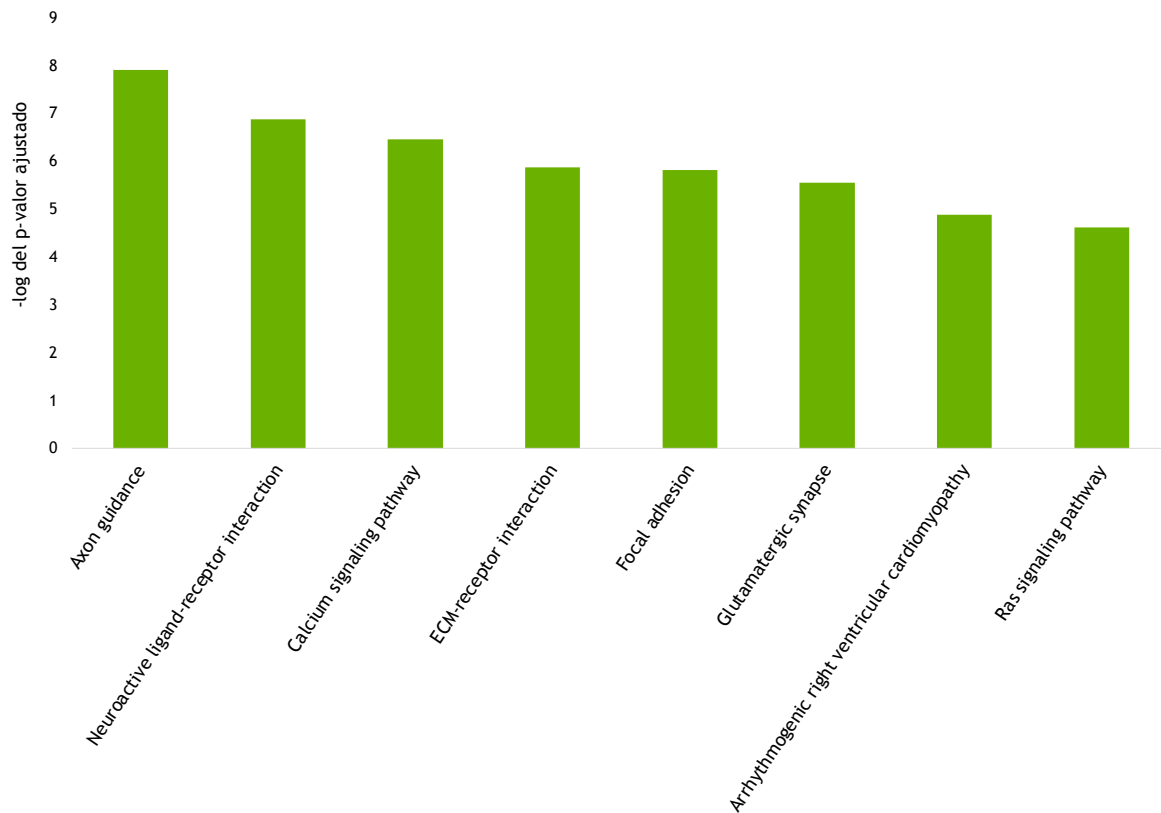


Figura 11. Gráfico de barras del análisis de enriquecimiento sobre rutas de señalización KEGG de las DMPs diferencialmente metiladas en los tumores CMS2.

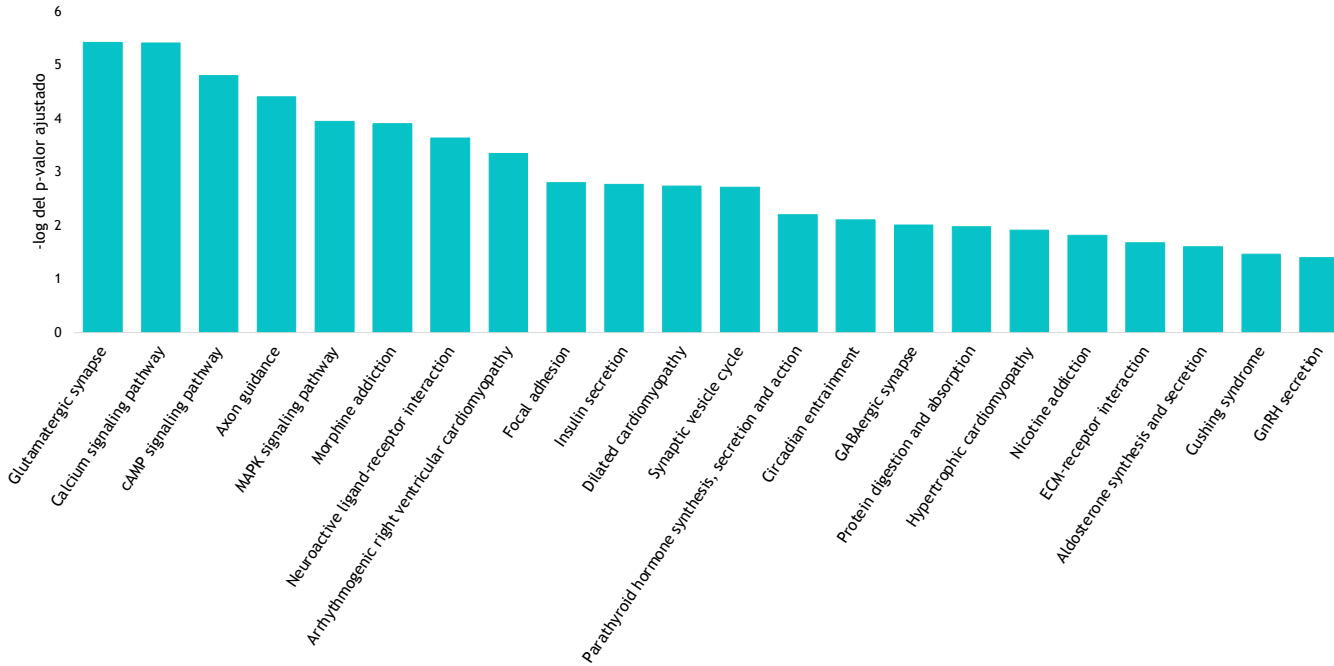


Figura 12. Gráfico de barras del análisis de enriquecimiento sobre rutas de señalización KEGG de las DMPs diferencialmente metiladas en los tumores CMS3

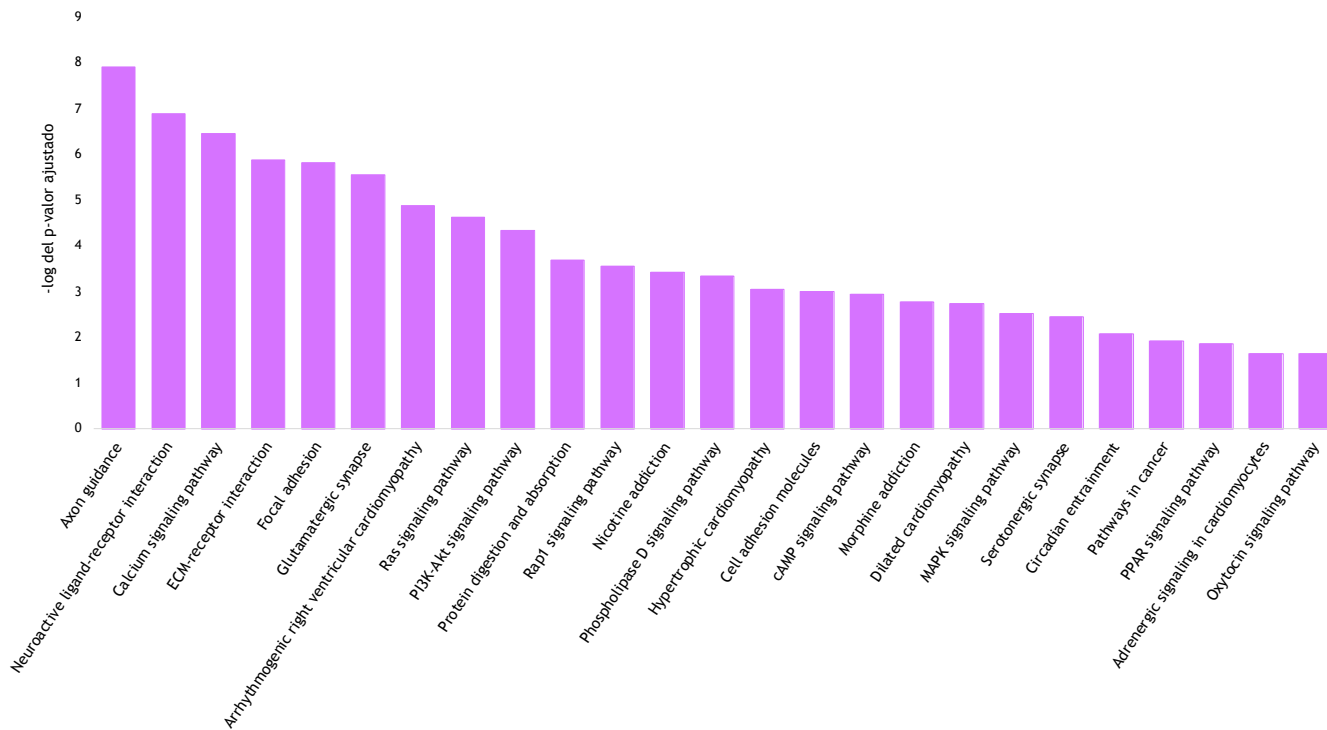
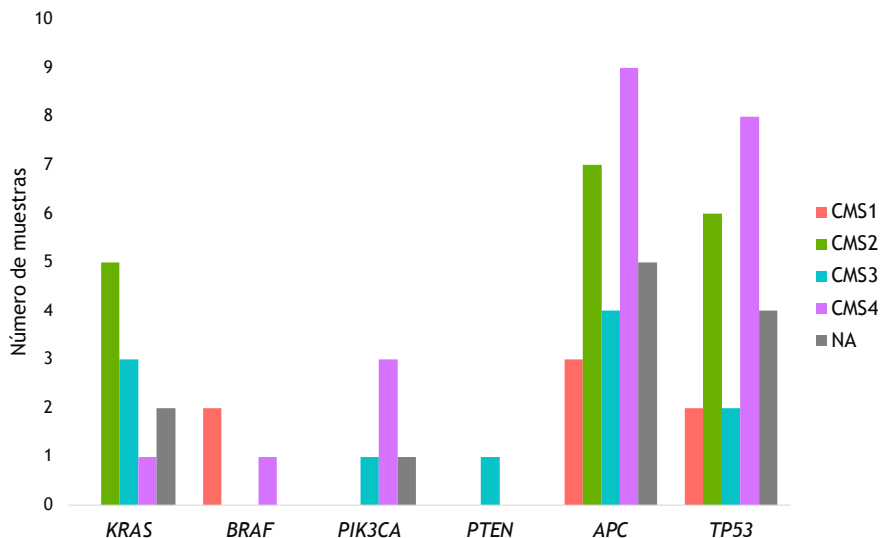


Figura 13. Gráfico de barras del análisis de enriquecimiento sobre rutas de señalización KEGG de las DMPs diferencialmente metiladas en los tumores CMS4.

## 6.2 ALTERACIONES SOMÁTICAS EN GENES DRIVER

Se realizó el filtrado de genes *driver* sobre los datos de exoma tumoral. Las muestras presentaron una media de 9,23 variantes que implican un cambio en la secuencia aminoacídica en este tipo de genes (Tablas suplementarias 16-49). Destaca el CCR del individuo COAD-11 con un total de 57 variantes en genes *driver* (Tabla suplementaria 41).

Se analizó la distribución de las alteraciones somáticas en los oncogenes *KRAS*, *BRAF* y *PIK3CA*, así como de los genes supresores tumorales *PTEN*, *APC* y *TP53* en los distintos CMS y en los tumores no adscritos a ningún Subgrupo Molecular Consenso (Figura 14 y tablas suplementarias 16-49). Las alteraciones en *APC* y *TP53* se encontraron presentes en CCRs pertenecientes a los cuatro CMS así como en tumores no clasificados en ninguno de estos. Se detectaron alteraciones en *KRAS* en once de los CCR, ninguno de ellos perteneciente al CMS1. El oncogen *PIK3CA* no presentó alteraciones en tumores pertenecientes a los subgrupos CMS1 ni CMS2, mientras que la variante p.(Val600Glu) de *BRAF* se detectó en dos de los tres tumores clasificados como CMS1 y en uno de los clasificados como CMS4. Solo se detectó una variante somática en el gen *PTEN* en el CCR del paciente 108, clasificado como CMS3.



**Figura 14. Distribución de alteraciones en los genes *driver* comunes en CCR para los distintos CMS.** Distribución de variantes que implican un cambio en la secuencia aminoacídica en los oncogenes *KRAS*, *BRAF* y *PIK3CA* y los genes supresores tumorales *PTEN*, *APC* y *TP53*. CMS1: Subgrupo Molecular Consenso 1, CMS2: Subgrupo Molecular Consenso 2, CMS3: Subgrupo Molecular Consenso 3, CMS4: Subgrupo Molecular Consenso 4, NA: no clasificado.

También se detectaron dobles eventos somáticos de nucleótido único o inserciones y deleciones de pequeño tamaño en distintos genes que presentan una función supresora asociada. En seis CCR se detectaron dobles eventos somáticos en el gen *APC* (tumores de los individuos 304, 310, 401, 501, READ-1 y READ-6). En dos CCR clasificados como CMS1 se detectaron dos variantes *frameshift* en *APC* (p.(Leu519Tyrfs\*19) y p.(Thr1556Asnfs\*3)) en el tumor del paciente 401 (Tabla suplementaria 30), mientras que en el CCR de READ-1 se encontraron dos variantes *stopgain* (p.(Arg283\*) y p.(Gln1406\*)) (Tabla suplementaria 45). El paciente 304 (CMS4) porta dos variantes a nivel somático en *APC*, una localizada en la región canónica de *splicing* (c.1409-1G>A) y una variante *frameshift* (p.(Thr1556Asnfs\*3)) (Tabla suplementaria 26). El tumor del paciente 310, clasificado como CMS2, presenta a nivel somático una variante de *splicing* en el gen *APC* (c.1548+1G>A) junto a la variante *frameshift* p.(Glu1309Aspfs\*4) (Tabla suplementaria 29). En la paciente READ-

6, CMS3, se identificaron a nivel somático dos variantes *stopgain* en el gen *APC* (p.(Arg1114\*) y p.(Glu1309\*)) (Tabla suplementaria 48) , y en el paciente 501 (no clasificado) las variantes *frameshift* p.(Lys975Glu\*5) y *stopgain* p.(Glu1286\*) también en *APC* (Tabla suplementaria 32). Excepto la variante c.1409-1G>A encontrada en el CCR del paciente 304 y clasificada como *passenger*, el resto son variantes *driver* de acuerdo a Cancer Genome Interpreter<sup>101</sup>. Destacó la variante p.(Gln1406\*) encontrada en el CCR del paciente READ-1 y recogida como *hotspot* en la base de datos Cancer Hotspots<sup>100</sup>.

Los CCR de los pacientes 109 (no clasificado dentro de ningún CMS) y COAD-5 (CMS4) presentaban un doble evento somático en el gen *TP53*. En el primero de ellos, 109, se observó la duplicación *inframe* p.(Ile232\_Tyr234dup) acompañada del *hotspot*<sup>100</sup> p.(Arg196\*), ambas clasificadas como variantes *driver*<sup>101</sup> (Tabla suplementaria 22). En el segundo caso (COAD-5) se encontró la variante *driver* localizada en la región canónica de *splicing* c.559+1G>A y una variante *missense* (p.Ala79Val) catalogada como *passenger*<sup>101</sup> (Tabla suplementaria 36).

También se observó más de un evento somático en otros genes *driver*. En el paciente 109, además de las variantes somáticas localizadas en *TP53*, también se identificaron las variantes *driver stopgain* p.(Lys164\*) y p.(Arg83\*) en el gen *FBXW7* (Tabla suplementaria 22). El paciente 401, además de las dos alteraciones en *APC*, mostró también dos variantes *driver* p.(His304Leu) y p.(Glu305del) en el gen *ACVR2A*, con función TSG y de fusión (Tabla suplementaria 30). En el tumor de la paciente COAD-6 se identificaron dos variantes *driver* (p.(Leu135Phe) y p.(Tyr451Phefs\*128)) en el gen *SOX9* (Tabla suplementaria 37).

También destacó *TGIF1* con dos variantes truncantes (p.(Gln107\*) y p.(Gly249Aspfs\*5) en el CCR del paciente 209, y *BCL11B*, gen con función ambigua y de fusión, con dos variantes *missense*, p.(Arg841His) y p.(Thr260Met) en el CCR de la paciente COAD-11. En ambos pacientes las variantes presentes fueron clasificadas como variantes *passenger* por Cancer Genome Interpreter<sup>101</sup> (Tablas suplementarias 23 y 41).

### 6.3 INTEGRACIÓN DE ÓMICAS: FILTRADO DE VARIANTES Y APROXIMACIÓN *ALL VS ONE*

Los datos de expresión diferencial de cada uno de los individuos procedentes de la aproximación *all vs one* se compararon con la correspondiente lista de variantes germinales priorizadas (procedentes del filtrado de exoma) y el listado de variantes exclusivas de tumor (Filter = PASS) para así estudiar la existencia de segundos eventos mutacionales en genes alterados a nivel germinal. Para el filtrado de variantes germinales en exoma se seleccionaron, tal como se describe en el apartado 5.5.1, variantes con gnomAD v2.1.1\_genome\_NFE  $\leq 0,1\%$  para variantes heterocigotas y gnomAD v2.1.1\_genome\_NFE  $\leq 1\%$  para variantes homocigotas y posibles heterocigotas compuestas localizadas en las regiones exónicas (tipo: *startloss*, *stopgain*, *stoploss*, *frameshift*, *inframe*, *missense* y sinónimas) y en posiciones intrónicas hasta  $\pm 20$ . Se llevó a cabo también una comparación entre los distintos individuos pertenecientes al mismo Subgrupo Molecular Consenso para detectar genes que presentasen alteraciones en más de un individuo y descartar aquellos genes cuyos patrones de expresión eran constantes a lo largo de la cohorte independientemente de la existencia, o no, de alteraciones a nivel germinal o tumoral.

#### 6.3.1 Subgrupo Molecular Consenso 1 (CMS1)

Dos tumores de la cohorte TOGETHER (105 y 401) y uno de la cohorte TCGA (READ-1) fueron agrupados dentro del subgrupo CMS1 (Tabla suplementaria 7).

En ninguno de estos individuos se detectaron variantes patogénicas, probablemente patogénicas o de significado incierto a nivel germinal en los genes asociados previamente a susceptibilidad hereditaria a CCR (Tabla 1).

Ninguna variante fue seleccionada tras el proceso de integración de los resultados del análisis *all vs one* con la lista de variantes germinales resultante del filtrado de exomas (Tablas suplementarias 50-52). Los motivos fueron debidos a inconsistencias entre el patrón de expresión y el tipo de variante, localización génica de la alteración, o ausencia de implicación en rutas biológicas que pudiesen señalar la relación del gen con la tumorigénesis colorrectal del individuo.

### 6.3.2 Subgrupo Molecular Consenso 2 (CMS2)

Ocho tumores (303, 310, COAD-1, COAD-2, COAD-12, COAD-13, COAD-17 y READ-5) procedentes del grupo de individuos con diagnóstico de CCR a una edad no superior a los 50 años fueron agrupados como CMS2 (Tablas suplementarias 7, 53-60). En ninguna de estas muestras se detectaron variantes germinales patogénicas, probablemente patogénicas o de significado incierto en genes relacionados previamente con susceptibilidad hereditaria a cáncer colorrectal (Tabla 1).

No se detectaron genes relacionados con las rutas propias del subgrupo molecular consenso con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que mostrasen alteraciones en más de un paciente a nivel germinal. Se detectaron dos genes en dos de los individuos pertenecientes a este subgrupo cuyo patrón de expresión y la localización de la alteración germinal a nivel de gen, así como su relación con rutas implicadas en la aparición de cáncer, sugieren una posible implicación en la aparición de la tumorigénesis colorrectal (Tabla 2).

En la paciente 303 se observa a nivel germinal la variante *missense* p.(Val145Leu) en heterocigosis en el gen implicado en la ruta de señalización Wnt (ruta KEGG hsa04310) *LEF1* ( $\log_2 \text{fold change} = 2,17$ ,  $\text{FDR} = 6,90 \times 10^{-3}$ ) (Tabla suplementaria 53). La variante se encuentra situada en el dominio N-terminal de unión a CTNNB1. *LEF1* también está asociado a cáncer y más concretamente a cáncer colorrectal (rutas KEGG hsa05200 y hsa05210) (Tabla 2). No se encontraron variantes a nivel somático en *LEF1* en la paciente 303 (Tabla suplementaria 25). Cuatro pacientes más, pertenecientes a CMS2, mostraron un aumento significativo de la expresión de *LEF1* a nivel tumoral (310, COAD-13, COAD-17 y READ-5) (Tabla suplementaria 61); sin embargo, en ninguno de estos pacientes se encontró una variante en el gen *LEF1* a nivel germinal o somático (Tablas suplementarias 29, 43, 44 y 47). Además, los pacientes 310, COAD-13, COAD-17 y READ-5 mostraron un aumento significativo de la expresión génica tumoral de *CTNNB1*, gen que regula la expresión de *LEF1*, aumento que no se observó en la paciente 303 (Tabla suplementaria 61). No se detectaron variantes somáticas en el gen



*CTNNB1* en los pacientes que mostraron un aumento significativo de la expresión de este gen (Tablas suplementarias 29, 43, 44 y 47).

El paciente READ-5 muestra a nivel germinal una variante heterocigota en el gen *CDH26* ( $\log_2$  fold change= -2,57 y FDR=  $1,67 \times 10^{-3}$ ) localizada en la región canónica de *splicing* (c.394-2A>G) (Tabla 2 y tabla suplementaria 60). No se detectaron segundos eventos mutacionales a nivel somático en la región codificante de este gen.

**Tabla 2. Listado de genes candidatos expresados diferencialmente en tumores CMS2 de pacientes portadores de variantes germinales seleccionadas tras aplicar los criterios de priorización de variantes.** KEGG hsa04310: Ruta de señalización Wnt, KEGG hsa05200: Rutas en cáncer, KEGG hsa05210: Cáncer colorrectal (<https://www.genome.jp/kegg/>). Chr: cromosoma; Cig: cigosidad; Het: heterocigosis.

| Gen          | Tránsito<br>(MANE<br>GRCh37) | Chr | Variante germinal    |                        |                        | gnomAD<br>v2.1.1_<br>genome_NFE<br>(%) | Cig | Paciente | Relación<br>con CCR   |
|--------------|------------------------------|-----|----------------------|------------------------|------------------------|--|-----|----------|---|
|              |                              |     | Variante<br>(GRCh37) | Cambio<br>nucleotídico | Cambio<br>aminoacídico |  |     |          |   |
| <i>LEF1</i>  | NM_016269.5                  | 4   | g. 109010395C>A      | c.433G>T               | p.(Val145Leu)          | 0,04                                   | Het | 303      | hsa04310<br>hsa05200<br>hsa05210  |
| <i>CDH26</i> | NM_177980.4                  | 20  | g.58557976A>G        | c.394-2A>G             |                        | 0,01                                   | Het | READ-5   | Ruta Wnt e<br>interacción<br>con $\beta$ -<br>catenina <sup>125-</sup><br>127 |

### 6.3.3 Subgrupo Molecular Consenso 3 (CMS3)

Cinco de los tumores se clasificaron dentro del subgrupo CMS3 (104, 107, 108, COAD-6 y READ-6) (Tablas suplementarias 7, 62-66). En estos pacientes tampoco se detectaron variantes germinales (probablemente) patogénicas o de significado incierto en genes de predisposición hereditaria a CCR (Tabla 1).

No se detectaron genes comunes a nivel germinal en el grupo de pacientes relacionados con las rutas de señalización relacionadas con CMS3. Sin embargo, en tres de ellos se encontraron variantes germinales cuya alteración de la expresión de ARN, así como su función y/o pertenencia a rutas génicas es compatible con la tumorigénesis colorrectal.

El paciente 108 porta en línea germinal una variante *startloss* en heterocigosis en el gen *TCEA3* (p.Met1?) que codifica un factor de transcripción (Tabla 3). El análisis *all vs one* reveló que en esta muestra la expresión del gen se encuentra disminuida con respecto a la mucosa colónica normal ( $\log_2$  fold change= -1,71 y FDR=  $4,94 \times 10^{-3}$ ) (Tabla suplementaria 64). El efecto de esta variante fue visualizado a nivel de ARN de mucosa colónica sana y de CCR del paciente, observándose un descenso del porcentaje del alelo variante a nivel germinal y la pérdida casi total de este a nivel somático. No se detectó un segundo evento mutacional en el exoma tumoral.

En la paciente COAD-6 se encontró una variante de pérdida de función (p.His44Glnfs\*8) a nivel germinal en heterocigosis en el gen relacionado con procesos metabólicos del ARN, *ZC3H12C* (Tabla 3). No se encontraron variantes somáticas en la región codificante de este gen en los datos de exoma tumoral. La expresión ( $\log_2$  fold change) de este gen a nivel tumoral en la muestra es de -2,79 y un FDR de  $1,58 \times 10^{-5}$  (Tabla suplementaria 65).

En la paciente READ-6 se detectó a nivel germinal la variante *stopgain* p.(Arg577\*) en heterocigosis en el gen *ADCY4* implicado en rutas de señalización metabólicas y de cáncer (Rutas KEGG hsa01100 y hsa05200) (Tabla 3). La expresión en tejido tumoral de *ADCY4* se encontraba disminuida de manera significativa,  $\log_2$  fold change de -2,51 y FDR=  $3,83 \times 10^{-5}$  (Tabla suplementaria 66). No se encontró un segundo evento mutacional el *ADCY4*.

**Tabla 3. Listado de genes candidatos expresados diferencialmente en tumores CMS3 de pacientes portadores de variantes germinales seleccionadas tras aplicar los criterios de priorización de variantes.** KEGG hsa01100: rutas metabólicas, KEGG hsa05200: rutas en cáncer (<https://www.genome.jp/kegg/>). Chr: cromosoma; Cig: cigosidad; Het: heterocigosis.

| Gen            | Tránsito<br>(MANE GRCh37) | Chr | Variante germinal    |                        |                        | gnomAD<br>v2.1.1_<br>genome_NFE<br>(%) | Cig | Paciente | Relación con<br>CCR               |
|----------------|---------------------------|-----|----------------------|------------------------|------------------------|--|-----|----------|-----------------------------------|
|                |                           |     | Variante<br>(GRCh37) | Cambio<br>nucleotídico | Cambio<br>aminoacídico |  |     |          |                                   |
| <i>TCEA3</i>   | NM_003196.3               | 1   | g.23751126T>A        | c.1A>T                 | p.Met1?                | 0,05                                   | Het | 108      | Factor de<br>transcripción        |
| <i>ZC3H12C</i> | NM_033390.2               | 11  | g.110007497dup       | c.131dup               | p.(His44Glnfs*8)       | 0                                      | Het | COAD-6   | Metabolismo<br>ARN <sup>128</sup> |
| <i>ADCY4</i>   | NM_001198568.2            | 14  | g.24794677G>A        | c.1729C>T              | p.(Arg577*)            | 0,01                                   | Het | READ-6   | hsa01100<br>hsa05200              |

### 6.3.4 Subgrupo Molecular Consenso 4 (CMS4)

Trece de los tumores pertenecientes a individuos de la cohorte de descubrimiento fueron clasificados dentro del subgrupo CMS4 (102, 209, 302, 304, 305, 308, COAD-3, COAD-5, COAD-7, COAD-8, COAD-9, COAD-11, READ-8) (Tablas suplementarias 7, 67-79).

El análisis de variantes germinales en genes relacionados con susceptibilidad hereditaria a CCR (Tabla 1) mostró la presencia de la variante heterocigota de significado incierto *MUTYH* (NM\_001048174.2): c.724G>A, p.(Val242Met) en el paciente COAD-9.

En los pacientes cuyos tumores se clasificaron dentro de este subgrupo no se detectaron variantes comunes a nivel germinal que mostrasen una diferencia de expresión significativa en el ARN y estuviesen implicados en procesos tumorogénicos. Se detectaron cuatro genes candidatos a nivel germinal cuya función, junto con la alteración del patrón de expresión podrían ser indicativos de su implicación en la susceptibilidad a CCR (Tabla 4).

En el paciente 304 se identificó en línea germinal la variante *stopgain* p.(Arg49\*) heterocigota en el gen *TRABD2A* asociado a la ruta Wnt<sup>129</sup> (Tabla 4), cuya expresión a nivel de ARN en tumor se encuentra disminuida ( $\log_2$  fold change= -2,20 y FDR=  $2,38 \times 10^{-3}$ ). La disponibilidad de datos de ARN de mucosa colónica normal permitió observar en IGV que esta variante provoca la degradación de ARN mensajero mediada por mutaciones terminadoras.

El paciente COAD-3 presenta a nivel germinal la variante *missense* p.(Arg1183Gln) en heterocigosis en el gen *CDHR2* implicado en la morfología intestinal<sup>130</sup> (Tabla 4). Este gen presenta una disminución de la expresión en este paciente a nivel tumoral ( $\log_2$  fold change= -2,84 y FDR=  $4,18 \times 10^{-2}$ ), mientras que en el resto de tumores del subgrupo no se apreció alteración significativa de los niveles de expresión. También se detectó en el paciente COAD-5 la variante heterocigota *missense* p.(Arg491Cys) en el gen *ARHGAP10* situada en el dominio proteico de activación de la GTPasa Rho (Tabla 4). La expresión de este gen en tumor se encuentra disminuida ( $\log_2$  fold change= -2,46 y FDR=  $4,98 \times 10^{-2}$ ).

Se encontró una variante *stopgain* p.(Arg677\*) en el gen *PLEKHG6* en la paciente COAD-11 (Tabla 4). Este gen muestra una disminución significativa de su expresión génica a nivel tumoral en la paciente ( $\log_2$  fold change = -2,29 y FDR=  $7,70 \times 10^{-3}$ ) (Tabla suplementaria 78), lo que concuerda con la naturaleza de la variante encontrada.

En ninguno de los genes mencionados se detectó un segundo evento somático en la región codificante analizada.

**Tabla 4. Listado de genes candidatos expresados diferencialmente en tumores CMS4 de pacientes portadores de variantes germinales seleccionadas tras aplicar los criterios de priorización de variantes.** Chr: cromosoma; Cig: cigosidad; Het: heterocigosis; TSG: gen supresor tumoral.

| Gen             | Tránsito (MANE GRCh37) | Chr | Variante germinal |                     |                   | gnomAD v2.1.1_genome_NFE (%) | Cig | Paciente | Relación con CCR                         |
|-----------------|------------------------|-----|-------------------|---------------------|-------------------|------------------------------|-----|----------|--|
|                 |                        |     | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoácido |                              |     |          |  |
| <i>TRABD2A</i>  | NM_001277053.2         | 2   | g.85097873G>A     | c.145C>T            | p.(Arg49*)        | 0,00648                      | Het | 304      | Ruta Wnt <sup>129</sup>                  |
| <i>CDHR2</i>    | NM_017675.6            | 5   | g.176018219G>A    | c.3548G>A           | p.(Arg1183Gln)    | 0,00648                      | Het | COAD-3   | Morfología intestinal <sup>130</sup>     |
| <i>ARHGAP10</i> | NM_024605.4            | 4   | g.148886195C>T    | c.1471C>T           | p.(Arg491Cys)     | 0                            | Het | COAD-5   | Ruta señalización Rho/AKT <sup>131</sup> |
| <i>PLEKHG6</i>  | NM_001384598.1         | 12  | g.6436778C>T      | c.2029C>T           | p.(Arg677*)       | 0                            | Het | COAD-11  | Morfología intestinal <sup>132</sup>     |

### 6.3.5 Tumores no incluidos en Subgrupos Moleculares

#### Consenso

Los tumores de los pacientes 103, 109, 402, 501 y READ-4 no se agruparon en ninguno de los CMS (Tablas suplementarias 7, 80-84). El filtrado en línea germinal de variantes en genes de predisposición a cáncer colorrectal (Tabla 1) no mostró la presencia de variantes causales o de significado incierto en ninguna de las muestras.

No se detectaron variantes germinales comunes implicadas en el proceso tumorigénico en estos cinco pacientes. Sin embargo, dos de los pacientes presentaron variantes en línea germinal cuya función podría relacionarse con el inicio de la tumorigénesis colorrectal (Tabla 5).

En los pacientes 402 y 501 se detectaron variantes germinales en genes que mostraron alteración significativa de la expresión a nivel tumoral (Tablas suplementarias 82 y 83). Se encontró una variante *missense* p.(Val942Met) en heterocigosis situada en los dominios proteicos *Cadherin* y *Cadherin-like* en el gen *CDHR2* en el paciente 402. La expresión génica en este tumor fue de  $\log_2$  fold change = -2,92 y  $FDR = 4,12 \times 10^{-2}$ . No se detectaron segundos eventos mutacionales en el tumor (Tabla 5 y tabla suplementaria 82).

El paciente 501 presenta la variante *stopgain* p.(Trp99\*) monoalélica en el gen *EEF2K* (Tabla 5), cuya expresión génica en tumor fue de  $\log_2$  fold change = -1,53 y  $FDR = 3,20 \times 10^{-2}$  (Tabla suplementaria 83). La actuación del mecanismo de degradación del ARN mensajero mediada por mutaciones terminadoras se observó en los datos de ARN tanto a nivel germinal como tumoral empleando el visor IGV. No se observaron segundos eventos mutacionales somáticos en la región codificante de este gen.



**Tabla 5. Listado de genes candidatos expresados diferencialmente en tumores no incluidos en ninguno de los Subgrupos Moleculares Consenso de pacientes portadores de variantes germinales seleccionadas tras aplicar los criterios de priorización de variantes. Chr: cromosoma; Cig: cigosidad; Het: heterocigosis.**

| Gen          | Tránsito<br>(MANE<br>GRCh37) | Chr | Variante germinal    |                        |                        | gnomAD<br>v2.1.1_<br>genome_NFE<br>(%) | Cig | Paciente | Relación<br>con CCR                        |
|--------------|------------------------------|-----|----------------------|------------------------|------------------------|--|-----|----------|--|
|              |                              |     | Variante<br>(GRCh37) | Cambio<br>nucleotídico | Cambio<br>aminoacídico |  |     |          |  |
| <i>CDHR2</i> | NM_017675.6                  | 5   | g.176013811G>A       | c.2824G>A              | p.(Val942Met)          | 0                                      | Het | 402      | Morfología<br>intestinal <sup>130</sup>    |
| <i>EEF2K</i> | NM_013302.5                  | 16  | g.22256001G>A        | c.297G>A               | p.(Trp99*)             | 0                                      | Het | 501      | Kinasa del<br>factor de<br>elongación<br>2 |

#### 6.4 COHORTE DE RÉPLICA

Se seleccionó una cohorte de 39 pacientes con diagnóstico de cáncer colorrectal a una edad igual o inferior a los 50 años pertenecientes a 38 familias no relacionadas (Tabla suplementaria 5).

Se analizó la firma mutacional tumoral y la carga mutacional tumoral en los CCR de los pacientes procedentes de la base de datos TCGA (cohortes COAD y READ) para comprobar la concordancia entre los datos fenotípicos y moleculares de los pacientes, y conocer las características moleculares de los tumores de los pacientes sin información fenotípica. El análisis reveló que el tumor del paciente COAD-60, sin datos fenotípicos, presentaba un patrón de firma y carga mutacional tumoral acorde con defectos en el sistema de reparación MMR por lo que fue descartado. En el caso de los pacientes COAD-61, COAD-63, COAD-64 y COAD-69, en los cuales el fenotipo tumoral aportado era sugestivo de defectos en el sistema de reparación anteriormente mencionado, los análisis de firma mutacional tumoral y/o carga mutacional tumoral no pudieron confirmar dichas características molecularmente y por lo tanto no fueron eliminados de la cohorte de réplica (Figura suplementaria 1 y tabla suplementaria 85).

En los 38 pacientes resultantes pertenecientes a 37 familias no relacionadas se estudió la presencia de variantes germinales en los 10 genes candidatos seleccionados en la fase de descubrimiento y también se analizó la existencia de variantes germinales (probablemente) patogénicas o de significado incierto en genes relacionados con la tumorigénesis colorrectal (Tabla 1).

En el paciente READ-31 se encontró una variante *stopgain* en heterocigosis a nivel germinal en un gen relacionado con susceptibilidad hereditaria a CCR, *NTHL1* (NM\_002528.7): c.835C>T, p.(Gln279\*). No se detectaron otras variantes patogénicas o probablemente patogénicas en los genes relacionados con predisposición a cáncer colorrectal en el resto de los pacientes.

Se identificaron a nivel germinal un total de tres variantes *missense* en los genes candidatos en tres pacientes no relacionados. Dos de los pacientes portaron variantes germinales en el gen *EEF2K* y el tercero una variante germinal en *ADCY4*.

En el paciente COAD-65, diagnosticado de cáncer colorrectal a los 49 años, se detectó en heterocigosis la variante p.(Pro686Leu) en el gen *EEF2K*. La paciente COAD-66, diagnosticada de CCR a los 50 años, presentó otra variante *missense* p.(Ala164Thr) germinal en heterocigosis en este mismo gen. En ambos casos, de acuerdo a las predicciones *in silico* realizadas, estas variantes podrían afectar a la funcionalidad de la proteína (Tablas 6 y 7).

En el paciente 1046, con diagnóstico de CCR con expresión conservada de las proteínas MMR a los 42 años, se encontró en línea germinal la variante *missense* p.(Arg982Cys) en heterocigosis en el gen *ADCY4* (Tabla 6), que también podría afectar a la función de la proteína en base a los predictores *in silico* consultados (Tabla 7).

**Tabla 6. Variantes germinales raras identificadas tras aplicar criterios de priorización de variantes en los 10 genes candidatos procedentes de la fase de descubrimiento.** Chr: cromosoma; Cig: cigosidad; Het: heterocigosis

| Gen          | Tránsito<br>(MANE GRCh37) | Chr | Variante germinal |                        |                        | gnomAD<br>v2.1.1_genome_NFE<br>(%) | Cig | Paciente |
|--------------|---------------------------|-----|-------------------|------------------------|------------------------|------------------------------------|-----|----------|
|              |                           |     | Variante (GRCh37) | Cambio<br>nucleotídico | Cambio<br>aminoacídico |                                    |     |          |
| <i>EEF2K</i> | NM_013302.5               | 16  | g.22291686C>T     | c.2057C>T              | p.(Pro686Leu)          | 0                                  | Het | COAD-65  |
| <i>EEF2K</i> | NM_013302.5               | 16  | g.22262515G>A     | c.490G>A               | p.(Ala164Thr)          | 0                                  | Het | COAD-66  |
| <i>ADCY4</i> | NM_001198568.2            | 14  | g.24788316G>A     | c.2944C>T              | p.(Arg982Cys)          | 0,08%                              | Het | 1046     |

**Tabla 7. Predicciones *in silico* de las variantes germinales priorizadas e identificadas en los genes candidatos *EEF2K* y *ADCY4* en la fase de réplica.** SIFT, T: *Tolerated*, D: *Deleterious*; PolyPhen-2, D: *Probably Damaging*, P: *Possibly Damaging*; Mutation Taster, D: *Disease Causing*.

| Gen          | Tránsito<br>(MANE GRCh37) | Variante germinal      |                        | Predictores <i>in silico</i> |            |                |         |      |
|--------------|---------------------------|------------------------|------------------------|------------------------------|------------|----------------|---------|------|
|              |                           | Cambio<br>nucleotídico | Cambio<br>aminoacídico | SIFT                         | PolyPhen-2 | MutationTaster | PROVEAN | CADD |
| <i>EEF2K</i> | NM_013302.5               | c.2057C>T              | p.(Pro686Leu)          | T                            | P          | D              | -4.32   | 23.4 |
| <i>EEF2K</i> | NM_013302.5               | c.490G>A               | p.(Ala164Thr)          | D                            | D          | D              | -2.69   | 24.4 |
| <i>ADCY4</i> | NM_001198568.2            | c.2944C>T              | p.(Arg982Cys)          | D                            | D          | D              | -5.33   | 31   |



# DISCUSIÓN



## 7 DISCUSIÓN

El CCR es una enfermedad común que surge de la interacción de factores genéticos y ambientales<sup>37</sup>. En los últimos años se ha identificado un aumento de la incidencia de un 1,4% anual en individuos con edades no superiores a los 50 años<sup>133</sup>, y se estima que para el año 2030 el 11% de los tumores de colon y el 23% de los de recto serán diagnosticados en individuos por debajo de esta edad<sup>134</sup>. Generalmente, se denomina CCR de inicio temprano a aquellos tumores colorrectales diagnosticados en individuos con edades no superiores a los 50 años.

Aunque actualmente no existen grandes estudios poblacionales que definan los agentes responsables del CCR de inicio temprano, se postula que entre el 14 y 35% de estos CCR están ocasionados por variantes germinales patogénicas implicadas en síndromes Mendelianos de predisposición a cáncer conocidos o genes de susceptibilidad moderada a CCR<sup>70,71,135,136</sup>. También intervienen otras causas como la preexistencia de patologías intestinales que predisponen a la aparición de la enfermedad, presencia de variantes comunes de baja penetrancia o el llamado efecto de cohorte que hace referencia a la exposición a determinados agentes ambientales en función del año de nacimiento<sup>137,138</sup>. Sin embargo, en muchos casos, debido a las características de los pacientes y en ocasiones también a su historia familiar de cáncer cabe pensar que el componente hereditario es superior al indicado por los estudios hasta ahora existentes.

Se ha postulado que el CCR de inicio temprano constituye una entidad heterogénea y distinta de aquellos tumores de inicio tardío<sup>44,133</sup>. Se recogieron muestras de mucosa colónica normal y tumoral de 20 pacientes procedentes de los hospitales participantes en este proyecto de investigación (cohorte TOGETHER) a las que se añadieron datos de 26 pacientes pertenecientes a la base de datos TCGA (cohortes COAD y READ) para conocer las características moleculares de los tumores en comparación con sus características fenotípicas.



Una de las características principales de los CCR de inicio temprano es la localización anatómica preferencial de estos tumores, ya que a pesar de que en la mayoría de la literatura se señala que un porcentaje más elevado se localiza en la porción distal del colon y el recto<sup>138</sup>, los estudios en población europea discrepan en este sentido<sup>137</sup>. En nuestra cohorte observamos que el 71,74% de los tumores se encontraban localizados en sigma y recto (80% de la cohorte TOGETHER y 65,40% de la cohorte TCGA).

Otro aspecto importante a tener en cuenta es que este tipo de tumores se diagnostican en estadios más avanzados. Se ha especulado que esto podría ser debido a un diagnóstico tardío de la enfermedad ocasionado en su mayor parte porque la aparición de los síntomas (en el caso de haberlos) suele relacionarse con otras causas debido a la juventud del paciente y en algunos casos también a la ausencia de historia familiar (el 21,7% de los individuos de esta cohorte no presentan antecedentes familiares oncológicos)<sup>70,71,139</sup>. Sin embargo, esta podría no ser la razón de que la mayoría de estos tumores se diagnostiquen en estadios III y IV, sino que los CCR de inicio temprano presenten una histopatología más agresiva<sup>133</sup>. En la cohorte de estudio el 63% de los tumores se diagnosticaron en uno de estos dos estadios, siendo el estadio III el más común.

Con el fin de caracterizar molecularmente los CCR de inicio temprano e identificar nuevos genes candidatos de susceptibilidad hereditaria a esta enfermedad, se emplearon las aproximaciones de exoma completo y RNA-seq tanto en tejido colónico normal como tumoral y el análisis de metilación tumoral en pacientes no relacionados diagnosticados a edades iguales o inferiores a los 50 años sin alteración de la vía MMR y con o sin historia familiar de cáncer.

Se tomaron los datos de RNA-seq de la cohorte TOGETHER, tanto de tejido normal como tumoral, a los que se unieron datos tumorales de pacientes del TCGA y de mucosa colónica normal de la base de datos GTEx. El procesamiento de datos se hizo empleando el mismo procedimiento bioinformático y se eliminaron los efectos de *batch* correspondientes. Primeramente, se evaluó la pureza de las muestras de CCR comparándolas con tejido sano, lo que mostró que dos muestras de CCR de la cohorte TOGETHER tenían un patrón de expresión

génica concordante con tejido colónico sano, y por tanto fueron eliminadas del estudio.

*Endofenotipado tumoral:*

Sobre las muestras de tejido tumoral se realizó la clasificación en Subgrupos Moleculares Consenso. El 9% de los tumores de la cohorte de descubrimiento se agruparon como fenotipo MSI inmune (CMS1), un 23% seguían la ruta canónica (CMS2), un 15% eran metabólicos (CMS3), los mesenquimales o CMS4 representaban el 38%, y el 15% restante no se clasificó dentro de ninguno de los subgrupos anteriores. Esta distribución difiere de la propuesta por Guinney y colaboradores en el año 2015<sup>33</sup>, donde el 14% correspondían a CMS1, un 37% a CMS2, el 13% CMS3, un 23% a CMS4 y el 13% restante presentaban un fenotipo de transición entre los distintos subgrupos CMS<sup>33</sup>. El estudio de Guinney y colaboradores<sup>33</sup> se realizó sobre tumores colorrectales de pacientes sin un rango de edad definido. Debido a que algunos estudios apuntan a que los CCR de inicio temprano podrían constituir una entidad distinta al presentar rasgos histológicos más agresivos<sup>133</sup>, el enriquecimiento de este tipo de pacientes en nuestra cohorte podría ser la causa de que la distribución de los CCR de nuestra cohorte difiera de la propuesta por Guinney y colaboradores<sup>33</sup>.

En el año 2019 Willauer y colaboradores<sup>140</sup> llevaron a cabo un estudio retrospectivo con más de 36.000 pacientes con CCR para estudiar si existían diferencias entre individuos diagnosticados a edad temprana y los de inicio tardío. Observaron que los CCR CMS1 son más comunes en individuos con inicio temprano, lo que contradice lo visto en nuestro estudio, donde únicamente el 9% de los CCR se clasificaron como CMS1<sup>140</sup>. Esto podría explicarse por el hecho de que en nuestra cohorte no se incluyen individuos con CCR que presentan alteración de la vía MMR, los cuales se clasifican dentro de este CMS<sup>33</sup>. Además, los CCR con alteraciones en la vía MMR son muy comunes en individuos menores de 50 años, debido a la predisposición hereditaria a este tipo de tumores que presentan los individuos con diagnóstico de Síndrome de Lynch o CMMRD, y a los CCR esporádicos que muestran metilación del promotor de *MLH1* o dobles eventos somáticos en alguno de estos genes reparadores<sup>141</sup>. Los análisis

GSA de los CCR CMS1 mostraron sobreexpresión de grupos de genes asociados a inestabilidad de microsatélites y reparación del ADN, condición característica de este grupo, así como la alteración del estado de metilación en distintas posiciones del genoma que afecta a diferentes rutas implicadas en cáncer. Además, el análisis de genes *driver* reveló que dos de los tumores pertenecientes a este subgrupo (105 y 401) portan la mutación p.Val600Glu de *BRAF*. Esta variante activadora está enriquecida en CMS1 y en la mayoría de los casos se asocia a la ruta serrada de aparición de CCR, cuya progresión a malignidad se ve ligada a la alteración de las proteínas reparadoras MMR<sup>33</sup>. Sin embargo, en estos dos tumores destaca la presencia de variantes truncantes en el gen *APC*, y en el caso del paciente 105 también la existencia de un *hotspot* en *TP53*. La co-ocurrencia de alteraciones somáticas en los genes *BRAF* y *APC*, aunque pertenecientes a rutas de inicio de la tumorigénesis distintas, ha sido reportada en la literatura y asociada a tumores colorrectales preferencialmente MSS<sup>142,143</sup>. El CCR del paciente READ-1 presenta además dos variantes truncantes a nivel somático en el gen *APC*. La combinación de alteraciones oncogénicas en *BRAF* con mutaciones en genes de las rutas de señalización Wnt y p53, así como la presencia de un doble evento mutacional en *APC*, podrían indicar la implicación de la progresión adenoma-carcinoma en la tumorigénesis colorrectal o un fenotipo intermedio entre esta y la ruta serrada en los tumores CMS1.

El 23% de los tumores de la cohorte de descubrimiento se clasificaron como CMS2. El análisis de enriquecimiento de genes mostró una sobreexpresión relacionada con estabilidad de microsatélites, algo acorde a las características propias de este grupo. Sin embargo, no se observó un aumento de expresión significativo de los genes de las rutas de señalización Wnt y *MYC* (FDR 0,686 y 0,549, respectivamente). Esto podría ser debido a la selección del conjunto de genes que componen estas rutas en el programa *CMSCaller* v0.99.2<sup>107</sup> empleado para la clasificación, así como al hecho de que la desregulación de la ruta Wnt- $\beta$ -catenina es un evento común en la tumorigénesis colorrectal<sup>34,107</sup>.

En la literatura existente los subgrupos CMS3 y CMS4 fueron observados de manera menos frecuente en individuos con diagnóstico

temprano de la enfermedad en comparación con aquellos de inicio tardío, representando un 4% y un 13% respectivamente<sup>140</sup>. Sin embargo, en nuestra cohorte los porcentajes observados fueron del 15% para CMS3 y del 38% para CMS4. Destaca la elevada presencia de tumores clasificados como CMS4 en nuestra cohorte en comparación con los datos obtenidos por Willauer y colaboradores (13%)<sup>140</sup>, quienes analizaron CCRs metastásicos que se esperaba que se agruparan dentro de CMS4 al presentar un fenotipo mesenquimal<sup>33,34</sup>.

A nivel molecular, el análisis GSA de expresión génica mostró que los tumores CMS3 presentan sobreexpresión de genes implicados en rutas metabólicas, mientras que los CMS4 muestran un marcado aumento de expresión en las rutas TGF- $\beta$  y EMT. Se observó la presencia de mutaciones somáticas en *KRAS* en CCRs pertenecientes tanto a CMS3 como CMS4, y también en tumores CMS2. En el tumor del paciente 304, clasificado como CMS4, se detectó a nivel somático la variante p.(Val600Glu) de *BRAF*, junto a dos variantes en *APC* y un *hotspot* en *TP53*. En conjunto, a partir del perfil de mutaciones somáticas *driver* se observa, como ya era de esperar, que este no es exclusivo de un subtipo concreto<sup>33</sup>. También se aprecia la coexistencia de alteraciones en *BRAF* con genes de las rutas de señalización Wnt y p53, al igual que ocurre en CMS1, lo que podría señalar una menor compartimentalización entre las dos rutas principales de la tumorigénesis colorrectal existentes hasta ahora.

Los análisis de metilación mostraron que los tumores de los subgrupos CMS1 y CMS3 presentan un mayor número de islas CpG metiladas a lo largo del genoma, siendo CMS1 el que muestra valores más altos de metilación. El estado de hipermetilación observado en CMS2 parece concordar con lo descrito previamente en la literatura que señala que la hipermetilación no es característica de este subgrupo<sup>33</sup>. Los estudios GSA llevados a cabo sobre estos datos, muestran que el subgrupo CMS1 presenta alteraciones en el epigenoma que afectan a un mayor número de rutas relacionadas con cáncer, en comparación con el resto de CMSs estudiados. El análisis de rutas KEGG sobre posiciones diferencialmente metiladas en CMS4 también muestra alteraciones en rutas relacionadas con la tumorigénesis. El perfil epigenético de CMS4 parece no ser muy confiable, según lo descrito en

la literatura, debido a la baja densidad celular que presenta el estroma<sup>144</sup>. De acuerdo con lo publicado, los perfiles de metilación de ADN permitirían diferenciar los subgrupos CMS1, CMS2 y CMS3, pero no CMS4<sup>144</sup>.

Por último, un 15% de los CCR no se adscribieron a ninguno de los CMS anteriormente citados. Se ha propuesto que estos casos podrían presentar patrones de expresión transicionales entre distintos subgrupos, y no constituir un CMS distinto<sup>33</sup>.

La clasificación CMS dejó patente la existencia de heterogeneidad molecular en una cohorte con fenotipo clínico homogéneo. Esta aproximación se centra principalmente en la biología y comportamiento tumoral y por ello en los distintos genes y rutas que se ven alterados una vez comienza el proceso tumorogénico. Para intentar precisar qué eventos eran responsables de la aparición de la neoplasia se estudió la firma y carga mutacional de los distintos CCR.

El empleo de la aproximación matemática propuesta por Alexandrov y colaboradores<sup>16,17</sup> y el uso de la base de datos COSMIC<sup>19</sup> permitió extraer de manera individualizada los distintos procesos implicados en el desarrollo de la neoplasia colorrectal de los pacientes. Se observaron patrones mutacionales relacionados tanto con factores exógenos como endógenos, así como alteraciones de la maquinaria celular.

En la cohorte de estudio destacó la presencia en todos los tumores de al menos una de las denominadas firmas planas (SBS1, SBS5 y SBS40) en porcentajes variables, y cinco de los tumores únicamente presentaron perfiles mutacionales asociados a estas. La firma SBS1 se asocia al proceso endógeno de desaminación de 5-metilcitosinas, ligado en la mayoría de los casos a la edad del individuo y las divisiones mitóticas que tienen lugar<sup>145</sup>. Sin embargo, investigaciones recientes también han asociado el patrón mutacional SBS1 a alteraciones bialélicas en el gen *MBD4* en adenomas colorrectales<sup>53</sup>. Este patrón es muy similar a la firma SBS96 propuesta por Degasperi y colaboradores y asociada a variantes patogénicas bialélicas en dicho gen en mixofibrosarcomas y melanomas uveales<sup>32</sup>. En ninguna de las muestras que componen nuestra cohorte se detectaron variantes germinales en *MBD4* en la región estudiada. Las firmas SBS5 y SBS40, aunque su

etiología permanece desconocida, también han sido relacionadas en algunos tumores con la edad del individuo<sup>17,18</sup>. Este tipo de firmas planas suponen un reto a nivel matemático, ya que son difíciles de deconvolucionar<sup>18</sup>. Es posible que, a medida que el número de datos de secuenciación de tumores aumente, estos patrones puedan sufrir variaciones y descomponerse en una o varias firmas con etiologías muy diversas.

Múltiples individuos mostraron en sus tumores firmas asociadas a carcinógenos de origen exógeno. Destaca la presencia de la firma mutacional compatible con la exposición a haloalcanos SBS42 en seis tumores de la cohorte de estudio TOGETHER (pacientes 102, 104, 107, 109, 305 y 401), en su mayoría con porcentajes entre 17,78-20% excepto en el caso del CCR del paciente 401 que es del 9,16%. Este patrón mutacional fue descrito por primera vez en trabajadores de una fábrica de impresión que presentaban un tipo de tumor de vías biliares con una incidencia rara en la población (colangiocarcinoma)<sup>146</sup>. El estudio de los tumores reveló este patrón de firma característico y la presencia de un elevado número de mutaciones somáticas (44,8 mutaciones/Mb de media)<sup>146</sup>, sin embargo, ninguno de los CCR de nuestra cohorte supera las 10 mutaciones/Mb. Dados los datos obtenidos con respecto a este patrón mutacional, podría ser que la presencia de esta firma en la cohorte TOGETHER pudiese estar relacionada con algún tipo de exposición ambiental común. Sin embargo, no se disponen de datos epidemiológicos para poder hacer esta asociación, ni tampoco se puede descartar que se deba a una exposición a haloalcanos, aunque la carga mutacional no sea coincidente con la descrita previamente.

También se detectaron espectros mutacionales compatibles con daños en el ADN ocasionados por exposición a la luz ultravioleta (SBS7a, SBS7b y SBS7d) en las muestras 402, COAD-3 y READ-6. En el caso de COAD-3 este patrón representó un 32,97% del total, mientras que en los otros dos CCR su porcentaje fue próximo al 10%. Este tipo de patrón es propio de melanomas, tumores con carga mutacional elevada<sup>147,148</sup>. Por el contrario, ninguno de estos CCR era hipermutado, ya que sus cargas mutacionales fueron inferiores a 10 mutaciones/Mb, además de ser un tejido no expuesto a radiación

solar<sup>48,149</sup>. Un estudio reciente sobre firmas mutacionales específicas de tejido recoge la similitud existente entre la firma de COSMIC<sup>19</sup> SBS7a y tres firmas presentes en su cohorte<sup>32</sup>. En este estudio se propone que el parecido en los patrones mutacionales podría dar lugar a un error en la asociación de la etiología del daño<sup>32</sup>, algo que podría ser plausible en este caso dado el contexto anatómico y molecular.

Uno de los tumores (CCR del paciente 302) mostró entre los patrones mutacionales el hábito tabáquico como uno de los responsables de la aparición neoplásica (22,66%). La firma SBS92 ha sido reportada en tumores de vejiga de individuos fumadores<sup>21</sup>. Aunque en este caso el tejido afectado es distinto, no extraña la aparición de esta huella en la neoplasia colorrectal en la cual el tabaquismo ha sido descrito como un factor de riesgo<sup>28</sup>.

Para terminar con los factores exógenos que dejan su marca en el ADN, seis CCR de la cohorte de estudio (pacientes 104, 304, 310, 501, COAD-1 y COAD-3) mostraron firmas compatibles con posibles errores del proceso de secuenciación asociados a baja calidad de las muestras<sup>19</sup> (rango del 11,61-29,93%). Esto contradice los resultados obtenidos en los controles de calidad llevados a cabo en distintas etapas del proceso de análisis de los datos.

En los tumores de los pacientes 104 y COAD-12 se detectó la firma mutacional SBS88 asociada a la colibactina, un compuesto endógeno asociado a las bacterias *Escherichia coli* pks<sup>+</sup> que forman parte de la microbiota intestinal, en porcentajes del 21,35% y 26,67% respectivamente<sup>31</sup>. La colibactina es un compuesto genotóxico que se ha asociado a riesgo de padecer cáncer colorrectal, así como a otros tumores en tejidos mucosos, al provocar alquilación de las adeninas en el ADN e introducir roturas de doble cadena<sup>29,31</sup>.

Diversas firmas también señalaron la participación de procesos implicados en el mantenimiento de la viabilidad celular dentro de los CCR de la cohorte de estudio, como el sistema de reparación MMR, diversas polimerasas y el sistema BER.

Las firmas mutacionales asociadas a defectos en el sistema de reparación MMR<sup>17,22</sup> están presentes en 10 de los CCR (302, 304, 308, 310, 401, 402, COAD-5, COAD-13, COAD-17 y READ-5) en los que oscila entre el 5,76% y el 24,86%, siendo la SBS15 la más común entre

las muestras. Defectos en este sistema de reparación se asocian con MSI y tumores hipermutados ( $\geq 10$  mutaciones/Mb)<sup>149,150</sup>. Del conjunto de tumores que muestran este patrón, solo el CCR del paciente 310 tiene una carga mutacional compatible con hipermutación, 11,32 mutaciones/Mb. Sin embargo, el porcentaje de firmas en este tumor asociado a defectos en el sistema MMR representa menos del 6% del total, pero co-ocurre con casi un 11% de firmas asociadas a la actividad enzimática de APOBEC (SBS2 y SBS13) que podrían explicar el elevado número de mutaciones en este caso<sup>17,145</sup>. En el resto de los tumores la presencia de las firmas mutacionales asociadas a defectos en el sistema MMR muestra valores de carga mutacional por debajo de los que cabría esperar si existiesen defectos en dicho sistema de reparación. En cambio, valores altos de carga mutacional sí han sido observados en otros CCR descartados de este estudio por no cumplir los criterios de inclusión al presentar alteración del sistema MMR. Todo esto podría indicar que el patrón mutacional observado no se relacionaría con defectos en estas proteínas reparadoras.

Lo mismo ocurre en los tumores que contienen espectros mutacionales compatibles con efectos en el ED de las polimerasas  $\epsilon$  (SBS10a y SBS10b) y  $\delta 1$  (SBS10d) (CCR de pacientes 304, 308, 501, COAD-12, COAD-13 y READ-4), los cuales en la mayoría de los casos representaban menos del 10% del total<sup>18,25</sup>. Defectos en estas regiones exonucleasa se asocian a tumores ultramutados ( $>100$  mutaciones/Mb), condición que no se cumplió en ninguno de los casos<sup>51,151</sup>. Tampoco se observó una elevada carga mutacional en el CCR del paciente COAD-3 que presentaba la firma SBS14 asociada a defectos combinados en el sistema de reparación MMR y el ED de la polimerasa  $\epsilon$  en un porcentaje del 21,98%<sup>17</sup>.

La firma SBS30<sup>23,24</sup> de la ruta BER asociada a variantes patogénicas bialélicas en el gen *NTHL1* se encontró en tres CCR (pacientes COAD-9, COAD-17 y READ-4) junto a la firma SBS18 asociada a daños ocasionados por especies reactivas del oxígeno<sup>17</sup>, y cuyos porcentajes conjuntos en todos los casos superan el 20%. El espectro mutacional asociado a SBS18 es muy similar al de la firma SBS36 relacionado con mutaciones bialélicas en *MUTYH*<sup>26,27</sup>. La firma SBS18 ha sido descrita en pacientes portadores de alteraciones



bialélicas en *MUTYH* y relacionada por similitud con otras firmas también asociadas a defectos en la ruta BER<sup>32,152</sup>. La firma SBS30 estaba presente también en el CCR del paciente COAD-5 en un 10,86%, mientras que la SBS18 se encontró en los tumores de los pacientes 305, COAD-13 y READ-8 en porcentajes mayores del 18%. En ninguno de los casos se encontraron variantes germinales patogénicas o probablemente patogénicas en la región codificante de los genes *NTHL1* ni *MUTYH*. Sin embargo, no podemos descartar la presencia de variantes en regiones reguladoras o intrónicas profundas en estos genes al no disponer de datos de genoma completo ni de RNA-seq germinal de la cohorte TCGA. Aun así, estos hallazgos podrían señalar una posible relación con la alteración de la ruta BER en los tumores en los que está presente.

Por último, en la cohorte de descubrimiento se observan 8 firmas mutacionales (SBS8, SBS16, SBS17b, SBS37, SBS41, SBS89, SBS93 y SBS94) con un rango porcentual que va del 16,44% al 43,17% y cuya etiología hasta el momento permanece desconocida<sup>17,18,30,96,153</sup>.

El análisis de firma mutacional en los CCR de la cohorte de estudio parece indicar la heterogeneidad y diversidad existente en la tumorigénesis colorrectal, al igual que la observada en la clasificación CMS. Sin embargo, aunque muchas de las firmas señalan procesos asociados a la aparición de cáncer colorrectal, las características de estos tumores contradicen la posible relación biológica de estas con la aparición de CCR. En este punto es importante destacar que el estudio de la firma mutacional es una aproximación matemática que tiene en cuenta un número de factores determinado, además de que asume que cada proceso mutacional está asociado a una única etiología<sup>18</sup>. Esto puede ser una manera un tanto simplista de interpretación que puede verse afectada en el caso de que varias firmas coincidan en un mismo tumor. Tampoco hay que olvidar que en este caso los datos disponibles para el análisis proceden de WES, por lo que se analizaron únicamente las regiones codificantes del genoma humano y las posiciones intrónicas hasta  $\pm 20$ . Las firmas mutacionales más aceptadas por la comunidad científica hasta el momento proceden de análisis matemáticos realizados sobre muestras de tumores con localizaciones muy diversas. Sin embargo, en estos últimos años se ha apuntado que

una aproximación con genomas y la extracción de firmas específicas de órgano sería más apropiada para obtener unos resultados más fiables<sup>32,154</sup>.

#### *Integración de ómicas:*

Los estudios anteriores no permitieron obtener evidencias de la implicación de genes concretos en la aparición del CCR en estos pacientes. Por ello, se llevó a cabo una estrategia empleando los datos de RNA-seq que denominamos *all vs one*. Esta estrategia consiste en un análisis de expresión diferencial, a través del cual cada uno de los tumores de manera individualizada fue enfrentado al conjunto de datos de expresión de mucosa colónica normal. Se consideró que un gen presentaba alteración en su expresión cuando sus valores eran  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 0,05$ . La lista de genes con expresión diferencial obtenida se comparó con los datos germinales priorizados de exoma de cada uno de los pacientes, datos previamente filtrados por calidad y frecuencia poblacional, obteniéndose un total de 10 genes candidatos. Estos candidatos están relacionados con distintas rutas de señalización previamente asociadas a CCR y procesos biológicos relacionados con la aparición de neoplasias como son Wnt/ $\beta$ -catenina (*LEF1* y *TRABD2A*), moléculas implicadas en la adhesión celular proteínas de la superfamilia de las cadherinas (genes *CDH26* y *CDHR2*) y el gen *PLEKHG6*, PI3K/Akt (*ARHGAP10*), genes relacionados con el proceso de elongación de la transcripción (*EEF2K* y *TCEA3*) y procesos metabólicos que ocurren a nivel celular (*ZC3H12C* y *ADCY4*).

En el gen *LEF1* (*Lymphoid enhancer-binding factor 1*, OMIM \*153245) se detectó la variante *missense* p.(Val145Leu) localizada en el dominio proteico de unión a la  $\beta$ -catenina en la paciente 303, diagnosticada de cáncer de recto a los 47 años y sin antecedentes familiares oncológicos conocidos. Este gen localizado en el cromosoma 4, además de ser un componente de la ruta canónica de Wnt (KEGG hsa04310) también ha sido asociado a las rutas implicadas en cáncer (KEGG hsa05200) entre las que se encuentra la de cáncer colorrectal (KEGG hsa05210). *LEF1* codifica un factor de transcripción perteneciente a la familia TCF/LEF asociados a la ruta de señalización Wnt<sup>155,156</sup>. La familia TCF/LEF es un grupo de factores de transcripción

que constan de un dominio de unión al ADN localizado en su extremo C-terminal y un dominio de unión a la  $\beta$ -catenina en el extremo N-terminal<sup>155</sup>. Normalmente *LEF1* no se expresa en tejidos colónicos sanos, sin embargo, se encuentra sobreexpresado en CCR donde presenta efectos oncogénicos, lo que concuerda con el aumento de expresión génica detectado a nivel tumoral en nuestro estudio<sup>155-158</sup>. Esta sobreexpresión puede ser debida a la desregulación de la ruta Wnt<sup>157</sup>. La desregulación de esta ruta ocasiona la estabilización de la  $\beta$ -catenina, que no puede ser degradada, y su translocación al núcleo donde interacciona físicamente con LEF1<sup>159,160</sup>. De esta forma, se activa la transcripción de genes diana de Wnt relacionados con la proliferación celular al unirse LEF1 a las regiones diana a través de su dominio C-terminal<sup>142,155,157,159,161-163</sup>. Sin embargo, en la paciente 303, no se observó un aumento significativo de la expresión génica de *CTNBI* (gen que codifica para la  $\beta$ -catenina) como sí se observó en otros pacientes del CMS2 (310, COAD-13, COAD-17 y READ-5) que mostraron sobreexpresión de *LEF1* a nivel tumoral. En estos últimos, el aumento de expresión de *CTNBI*, que regula la expresión de *LEF1*, podría indicar que la sobreexpresión del factor de transcripción se debe a una alteración de la ruta Wnt previa a este punto. Sin embargo, la ausencia de sobreexpresión significativa del gen de la  $\beta$ -catenina en la paciente 303 hace pensar que la variante *missense* de *LEF1*, localizada en línea germinal, es la responsable del aumento de expresión del factor de transcripción, punto a partir del cual se produciría la desregulación de la ruta.

Además de a través de la desregulación de la ruta de señalización Wnt, la actividad transcripcional de *LEF1* puede ser promovida a través de *MYC* en el epitelio colónico<sup>163</sup>. *MYC* es una diana de la ruta canónica Wnt/ $\beta$ -catenina, la cual podría promover la transcripción de LEF1 potenciando más la activación de la ruta Wnt<sup>163</sup>. No se observaron variantes de pequeño tamaño a nivel tumoral en *MYC* en el CCR de la paciente 303, aunque no podemos descartar la existencia de duplicaciones o translocaciones a nivel somático. Por todo lo expuesto anteriormente, hipotetizamos que alteraciones en el dominio de unión a la  $\beta$ -catenina en *LEF1* podrían favorecer la actividad de este factor de transcripción.

El otro gen que presenta relación con la ruta Wnt es **TRABD2A** (*TRAB domain-containing protein 2A*, OMIM \*614912). En el paciente 304, diagnosticado de cáncer de recto a los 30 años y con antecedentes familiares de CCR por vía materna, se detectó en heterocigosis la variante germinal p.(Arg49\*). Esta variante provoca la degradación del ARN mensajero mediada por mutaciones terminadoras, según se observó en las secuencias de ARN germinal y tumoral. La expresión de *TRABD2A* se encuentra disminuida a nivel tumoral, hecho que concuerda con la naturaleza de la variante encontrada. Hasta la fecha no existe literatura que relacione la función de esta proteína con procesos neoplásicos en humanos. Sin embargo, ha sido reportada como antagonista de la función de Wnt al presentar actividad proteasa en embriones y células humanas, donde parece mediar la eliminación de 8 residuos N-terminales de un conjunto de proteínas Wnt<sup>129</sup>. Además, *TRABD2A* presenta una expresión enriquecida en el intestino<sup>164</sup>. La desregulación de la ruta Wnt provoca la aparición, crecimiento y metástasis en cáncer colorrectal, por tanto, la ausencia de la proteína *TRABD2A* podría favorecer la progresión de estos procesos<sup>165</sup>. Además, en el CCR de este paciente se detectaron eventos somáticos en el gen *APC*, donde una de las variantes (p.Thr1556Asnfs\*3) fue clasificada como variante *driver* en Cancer Genome Interpreter<sup>101</sup>.

Otros tres genes propuestos de susceptibilidad a CCR se encuentran asociados a procesos de adhesión celular. En el paciente READ-5, diagnosticado de CCR a los 43 años, se encontró una variante localizada en el sitio canónico de *splicing* del intrón 4 del gen **CDH26** (*Cadherin 26*, OMIM \*617685). Este gen codifica una proteína con el mismo nombre que se encuentra presente en las células epiteliales gastrointestinales y que está implicada en el proceso de adhesión celular al interactuar con distintas cateninas ( $\alpha$ ,  $\beta$  y p120)<sup>127</sup>. Se trata de una cadherina atípica perteneciente a la superfamilia de las cadherinas que se expresa principalmente en tejidos epiteliales y media procesos críticos implicados en inflamación, adhesión celular, migración y diferenciación<sup>127,166</sup>. Se ha reportado que en el epitelio colónico que presenta condiciones de homeostasis la expresión de *CDH26* no es elevada, y que el aumento de expresión se produce como respuesta a

procesos alérgicos<sup>127,164</sup>. También se ha postulado que este aumento de expresión puede ser debido en parte a un mecanismo por el cual se trate de resolver el proceso inflamatorio asociado a la alergia, y con ello poder reestablecer las condiciones homeostáticas del epitelio colónico<sup>127</sup>. De acuerdo a esto último, la disminución significativa de la expresión génica en este tumor con respecto a tejido colónico normal podría señalar una imposibilidad de restaurar las condiciones normales del tejido colónico en el paciente, una vez iniciado un proceso inflamatorio asociado a la neoplasia colorrectal. Sin embargo, en este punto, cabe señalar que este gen presenta una elevada variedad de variantes de *splicing* lo que da lugar a múltiples transcritos, varios de ellos expresados en el tracto gastrointestinal, entre los que se encuentra el transcrito canónico (NM\_177980.4)<sup>164</sup>. Sería necesario llevar a cabo un análisis de expresión génica de las distintas isoformas presentes, para conocer el efecto de esta variante, así como la inspección visual de los patrones de *splicing* en la mucosa colónica normal.

Los pacientes COAD-3 y 402 de la cohorte de descubrimiento con diagnóstico de cáncer de recto a los 37 y 47 años, respectivamente, y cuyos tumores fueron clasificados en distintos subgrupos (COAD-3 como CMS4 y 402 no clasificado en ninguno de los CMS), mostraron variantes *missense* germinales en heterocigosis (p.(Arg1183Gln) y p.(Val942Met)) en otro gen perteneciente a la superfamilia de las cadherinas, **CDHR2**. En ambos casos el gen **CDHR2** presenta una expresión génica significativamente disminuida a nivel tumoral. **CDHR2** (*Cadherin-related family, member 2*, OMIM \*619713) es una protocadherina que se expresa principalmente en intestino delgado, intestino grueso, hígado y riñón<sup>164</sup>. **CDHR2** desempeña un papel en el mecanismo de adhesión célula-célula, induciendo la inhibición por contacto en la superficie lateral de las células epiteliales al favorecer la retención de la  $\beta$ -catenina en una localización submembrana y disminuyendo así la señalización de las dianas de la ruta Wnt<sup>167-172</sup>. De este modo, el gen **CDHR2**, posiblemente pueda desempeñar un papel de supresor tumoral al inhibir la proliferación celular. Su ausencia a nivel transcripcional y proteico ha sido demostrada en estudios *in vitro* e *in vivo*<sup>170,172,173</sup>. Esta molécula también participa en la formación de las microvellosidades intestinales, implicadas en adhesión celular pero

también en el procesamiento y toma de nutrientes, por lo que desempeña un papel muy importante en los enterocitos y la estructura y funcionalidad intestinal<sup>130,173</sup>. CDHR2 se localiza en el extremo apical de las células epiteliales intestinales y junto con CDHR5 y sus proteínas citoplasmáticas de unión favorecen la formación y reclutamiento de las protusiones adyacentes para la formación del borde en cepillo<sup>130</sup>. Se ha observado que la ausencia de CDHR2 en ratones desnudos provoca perturbaciones en la morfología del borde en cepillo pudiendo ocasionar una reducción en la capacidad funcional del intestino<sup>130</sup>.

Otro gen implicado en la formación de microvellosidades y pliegues de membrana en las células epiteliales y que está implicado en la macropinocitosis<sup>132</sup>, es **PLEKHG6** (*Pleckstrin homology domain. And RhoGEF domain-containing protein G6*, OMIM \*611743), que presentó a nivel germinal una variante *stopgain* p.(Arg677\*) en heterocigosis y disminución de la expresión génica de manera significativa a nivel tumoral en la paciente COAD-11. Esta paciente presentaba un diagnóstico de cáncer de colon ascendente a los 48 años y antecedentes familiares de la enfermedad. PLEKHG6 interactúa con proteínas GTPasas pequeñas como RhoA, RhoC y Rho G<sup>132,174</sup>. En cáncer de mama el silenciamiento de *PLEKHG6* suprime la activación de RhoA y RhoC reduciendo la invasión celular y metástasis<sup>174-177</sup>. En cáncer colorrectal se ha descrito que la activación de RhoA también contribuye a la progresión y metástasis<sup>131,132,176</sup>. Sin embargo, existen evidencias de que la inactivación de RhoA también parece promover la progresión tumoral y metástasis en CCR al aumentar la actividad de las rutas de señalización Wnt y *MYC*, y se ha propuesto un papel antioncogénico de *PLEKHG6* en este tipo de neoplasia, aunque esto último todavía debe confirmarse con estudios funcionales<sup>131,132,176</sup>.

Otra molécula que puede interactuar con las funciones de RhoA es **ARRGAP10** (*Rho GTPase-activating protein 10*, OMIM \*609746), un miembro de la familia de proteínas activadoras Rho GTPasas que activa la transformación de las formas activas guanosin trifosfato (GTP) a guanosín difosfato (GDP)<sup>131,178</sup>. El paciente COAD-5, diagnosticado de cáncer de sigma a los 43 años y sin historia familiar oncológica, presentó disminución significativa de la expresión génica a nivel

tumoral de *ARHGAP10* y la presencia de la variante heterocigota p.(Arg491Cys) a nivel germinal. La variante germinal se encuentra localizada en los dominios proteicos *Rho GTPase-activating protein domain* y *Rho GTPase activation protein*. En el cáncer de sigma de este paciente se identificaron dos variantes en *TP53* no recogidas en la base de datos *Cancer Hotspots*, una de ellas fue clasificada como evento *driver* (c.559+1G>A) por *Cancer Genome Interpreter*<sup>100,101,179</sup>. La disminución de la expresión a nivel transcripcional ha sido detectada en varios tipos de tumores como próstata, pulmón, ovario, y glioblastoma, en los que se ha propuesto como un posible supresor tumoral al observarse que inhibe procesos de proliferación celular y metástasis<sup>131,180</sup>. En CCR también se ha observado este descenso de la expresión génica, indicando que también podría desempeñar una función antitumorogénica al inhibir la ruta PI3K/Akt, regulando de manera negativa la fosforilación de Akt y la actividad de RhoA en las células tumorales, e impidiendo la activación de dianas de la ruta de señalización Wnt<sup>131,181</sup>. En cáncer de ovario se ha observado, además, que esta molécula podría inhibir rutas implicadas en el ciclo celular, replicación del ADN y la ruta BER al interactuar de manera directa con Cdc42 e inactivarla<sup>180</sup>.

También se detectaron variantes germinales que podrían estar implicadas en la tumorogénesis colorrectal en los genes *EEF2K* y *TCEA3* relacionados con el proceso de elongación de la transcripción.

En el gen *EEF2K* se identificaron variantes germinales heterocigotas en tres pacientes, el paciente 501 perteneciente a la cohorte de descubrimiento y los pacientes COAD-65 y COAD-66 de la cohorte de réplica.

El paciente 501, diagnosticado de cáncer de sigma a los 43 años y con antecedentes familiares oncológicos, presenta la variante germinal *stopgain* (p.Trp99\*) en el gen *EEF2K* (*Eukaryotic elongation factor 2 kinase*, OMIM \*606968). Esta variante *stopgain* activa el mecanismo de degradación del ARN mensajero mediado por mutaciones terminadoras, como se observó en datos de ARN de mucosa colónica normal y tumoral. En el tumor de este paciente se identificaron dos eventos mutacionales somáticos en *APC*.



El paciente COAD-65, con diagnóstico de cáncer colorrectal a los 49 años, es portador de la variante p.(Pro686Leu) en *EEF2K*. Esta variante se localiza en los dominios *Sell-like* y *Elongation factor 2 kinase*, implicados en la interacción de esta molécula con su único sustrato conocido, eEF2. En la paciente COAD-66, diagnosticada a los 50 años de cáncer colorrectal, también se identificó en *EEF2K* la variante p.(Ala164Thr) que afecta a los dominios proteicos *MHCK/EF2 kinase*, *Protein kinase-like domain* y *Elongation factor 2 kinase* implicados en la actividad funcional de esta. La ausencia de datos de RNA-seq en estos dos pacientes de la fase de réplica hace necesaria la validación funcional de estas variantes para conocer el efecto que tienen en la función proteica.

*EEF2K* codifica la molécula eEF2K, una proteína kinasa inusual perteneciente al grupo de las  $\alpha$ -quinasas, cuya función es regular la elongación de la síntesis proteica al fosforilar al factor de elongación 2 (eEF2)<sup>182-184</sup>. De esta manera, eEF2K impide que eEF2 pueda unirse a los ribosomas, evitando así la traducción del ARN mensajero<sup>182-184</sup>. La actividad de eEF2K es impulsada por calmodulina, hidroxilación, autofosforilación y fosforilación mediada por otras proteínas efectoras (algunas de ellas implicada en rutas oncogénicas frecuentemente alteradas en cáncer como PI3K/Akt y Ras/Raf/MEK/ERK a través de mTORC1) o de forma directa en algunos casos<sup>185-188</sup>. Se ha visto que eEF2K es regulado por rutas implicadas en procesos tumorigénicos, manteniendo así su actividad a bajos niveles y permitiendo que el proceso de elongación mediado por eEF2 continúe<sup>186</sup>. eEF2K presenta un efecto protumorigénico en determinados tipos de tumores como mama, páncreas y tumores cerebrales al ejercer una función citoprotectora frente al estrés en situaciones de hipoxia, ausencia de nutrientes o acidosis, al fosforilar a su sustrato e impedir la síntesis proteica<sup>182,186,189</sup>. Por el contrario, en tumores intestinales y hepatomas se ha descrito que esta kinasa podría inhibir el desarrollo y aparición de cáncer<sup>190,191</sup>. Faller y colaboradores observaron que en ratones que presentaban *APC* inactivo, mTORC1 muestra mayor actividad durante la regeneración de las criptas colónicas en comparación con ratones con fenotipos salvaje, efecto que se revierte al suministrar rapamicina que inhibe la señalización de mTORC1<sup>192</sup>. El efecto de la rapamicina se ve



anulado tras la delección de eEF2K, lo que apunta al papel supresor tumoral de esta proteína en la neoplasia colorrectal<sup>192</sup>. En el año 2014, Xie y colaboradores mostraron que la sobreexpresión de eEF2K en células de CCR (tanto deficientes en APC como sin alteración de este) suprimía la autofagia<sup>192</sup>. La autofagia es un mecanismo que emplean las células tumorales en situaciones de estrés para garantizar su supervivencia y que ha sido descrito en otros tipos de tumores bajo las mismas circunstancias de actividad de eEF2K, apoyando la idea de que esta kinasa presenta funciones supresoras tumorales<sup>193-195</sup>. Así, los resultados de estudios realizados en cáncer colorrectal discrepan de los obtenidos en otros tumores donde eEF2K presenta una función oncogénica, lo que parece señalar que esta molécula desempeña una función dependiente de tejido.

En *TCEA3* (*Transcription elongation factor A, 3*, OMIM \*604128) se encontró la variante *startloss* (p.Met1?) en el paciente 108 diagnosticado de un tumor en sigma a los 47 años y con antecedentes familiares de la enfermedad a edad avanzada en su abuela materna. El efecto de la variante se observó en la secuencia de ARN, a nivel tanto germinal como tumoral. En humanos *TCEA3* codifica la subunidad h del factor de elongación TFIIS (TFIIS.h), el cual es requerido para el correcto funcionamiento de la ARN polimerasa II durante la transcripción<sup>196,197</sup>. Esta subunidad puede unirse directamente con la ARN polimerasa II permitiéndole a esta escapar del arresto al que se ve sometida por determinadas secuencias de ADN, y de este modo continuar con la síntesis del ARN mensajero<sup>196,198,199</sup>. La expresión de *TCEA3* se ha encontrado disminuida en comparación con tejido normal en varios contextos neoplásicos, como es el caso de rhabdomyosarcoma, cáncer de ovario o cáncer gástrico, en los que se sugiere que desenvuelve un papel en la atenuación de la proliferación celular y la inducción de la apoptosis<sup>198,200,201</sup>. En rhabdomyosarcoma se ha observado que TFIIS.h inicia la apoptosis a través de la ruta extrínseca que implica a los factores de muerte celular (TNF) y/o la ruta intrínseca promoviendo la permeabilización de la membrana externa mitocondrial, ambas vías resultan en la activación de las caspasas<sup>200</sup>. En el caso de las células tumorales de ovario la apoptosis mediada por caspasas es inducida a través de la ruta intrínseca, TFIIS.h se une a los

receptores I de TGF- $\beta$  e induce la activación de la apoptosis mediada por JNK e independiente de SMAD<sup>198,199,201</sup>. Los miembros de la familia TGF- $\beta$  presentan múltiples funciones asociadas que regulan la proliferación y diferenciación celular, modificación de la matriz extracelular, angiogénesis y apoptosis<sup>201</sup>. En el caso del epitelio colónico estas proteínas desempeñan un papel dual en la oncogénesis, pudiendo en algunos casos inhibir y en otros favorecer el proceso neoplásico, por lo que la actividad de los miembros de la familia TGF- $\beta$  es altamente dependiente de un contexto y ambiente celular específico<sup>201,202</sup>. Además de las rutas de apoptosis previamente señaladas, TFIIS.h ha sido identificado como una diana transcripcional de p53<sup>196</sup>. Estudios realizados sobre líneas celulares de cáncer colorrectal mostraron que p53 promueve de forma selectiva la elongación de la transcripción de los genes bax asociados a apoptosis a través de TFIIS.h, pero no la p21 relacionado con el ciclo celular<sup>196</sup>. Todo lo expuesto anteriormente asociado al descenso de expresión génica en el paciente 108 parece indicar que *TCEA3* desempeña un papel de TSG.

Cabe destacar que dos de los pacientes pertenecientes a CMS3 presentaron a nivel germinal alteraciones en genes relacionados con procesos metabólicos celulares. La paciente COAD-6 es portadora a nivel germinal una variante en el gen *ZC3H12C* relacionado con procesos metabólicos del ARN<sup>128</sup>. Por otro lado, se detectó una variante germinal en el gen *ADCY4*, relacionado con distintos procesos metabólicos (ruta KEGG hsa001100) en la paciente READ-6.

La paciente COAD-6, diagnosticada a los 46 años de cáncer de sigma, presenta en heterocigosis la variante germinal p.(His44Glnfs\*8) en el gen *ZC3H12C* (*Zinc Finger CCCH domain-containing protein 12C*, OMIM \*615001). A nivel tumoral se observó una disminución significativa de los niveles de transcripción de este gen en esta paciente en comparación con los niveles de la mucosa colónica normal. *ZC3H12C* es uno de los cuatro miembros de la familia de proteínas inducidas por la proteína quimioatrayente de monocitos, caracterizadas por poseer dominios *zinc finger CCCH* y un dominio N-terminal PIN (dominio PiIT-N-terminal) implicado en procesos inflamatorios<sup>203,204</sup>. La mayoría de los miembros de esta familia actúan como reguladores

de procesos metabólicos del ARN entre los que se encuentra el *splicing* del ARN mensajero, la poliadenilación, exportación, traducción y la degradación del ARN mensajero mediada por mutación terminadora<sup>128</sup>. La proteína ZC3H12C hasta el momento no ha sido relacionada con cáncer. En células endoteliales se ha observado que su sobreexpresión atenúa de manera significativa la expresión de quimoquinas y moléculas adhesivas inducidas por TNF- $\alpha$ , desempeñando una función antiinflamatoria, posiblemente al suprimir la ruta de NF- $\kappa$ B, que se encuentra sobreexpresada en CCR<sup>205</sup>. ZC3H12C presenta una elevada homología con ZC3H12A, la cual parece tener funciones proapoptóticas, antiproliferativas de regulación de la angiogénesis en procesos inflamatorios y de inhibición de la EMT en cáncer de mama y carcinoma renal de células claras<sup>206,207</sup>. Con todo lo expuesto anteriormente, ZC3H12C podría presentar unas funciones parecidas a ZC3H12A, que serían acordes a los niveles de expresión observados a nivel tumoral en el caso de la paciente COAD-6 y a la alteración encontrada a nivel germinal; sin embargo, para poder confirmarlo sería necesario realizar estudios funcionales sobre células de CCR. También es importante mencionar que, en el tumor de la paciente se identificaron dos eventos somáticos en el TSG *SOX9*, la variante *missense* p.(Leu135Phe) y la variante p.(Tyr451Phefs\*128) ambas clasificadas como *drivers*<sup>101</sup>.

Por otro lado, en la paciente READ-6, diagnosticada de cáncer de recto a los 48 años y con antecedentes familiares de la enfermedad, se encontró la variante germinal p.(Arg577\*) en heterocigosis a nivel germinal en el gen *ADCY4* (*Adenylate cyclase 4*, OMIM \*600292), y además se observó una reducción significativa de la expresión génica de este gen en el tumor. *ADCY4* está implicado en las rutas KEGG de señalización metabólica y las relacionadas con cáncer (rutas KEGG hasa01100 y hasa05200, respectivamente). Las variantes de pérdida de función han sido descritas en este gen como un mecanismo de patogenicidad<sup>208</sup>. En el paciente 1046, diagnosticado de cáncer de colon a los 42 años y que forma parte de la cohorte de réplica, se detectó en este gen a nivel germinal la variante p.(Arg982Cys) en heterocigosis, la cual se localiza en el dominio *Adenylyl cyclase class-3/4/guanylyl cyclase*. La ausencia de datos de RNA-seq en este caso hace necesaria

una validación funcional para poder determinar el efecto de esta variante.

La familia de adenilato ciclasas es un grupo de 18 glicoproteínas que están implicadas en varios procesos celulares como proliferación celular, apoptosis, migración celular, metabolismo celular, angiogénesis y escape inmune, al encargarse de la producción de adenosín monofosfato cíclico (AMPC) a partir de adenosín trifosfato (ATP), lo que desencadena el aumento del primero a nivel intracelular<sup>209-212</sup>. Los distintos subtipos de adenilato ciclasas presentan patrones distintos de expresión según el tipo de tumor, así algunos de los miembros de esta familia favorecen el desarrollo y progresión de neoplasias desempeñando un papel oncogénico, mientras que otros muestran efectos antitumorales<sup>208</sup>. La expresión de *ADCY4* se encuentra significativamente silenciada en diversos tipos de tumores y en algunos casos ha sido asociada a niveles de metilación en la neoplasia, concretamente en cáncer de mama este gen se asocia a funciones antitumorogénicas<sup>208,213</sup>. El descenso de los niveles de expresión de este gen y la presencia de una variante truncante a nivel germinal en la paciente READ-6 podrían ser indicativos de que *ADCY4* muestra también un papel supresor tumoral en la tumorogénesis colorrectal. Cabe destacar la presencia de un doble evento mutacional somático en *APC* en el CCR de esta paciente (variantes p.(Arg1114\*) y p.(Glu1309\*)), ambas clasificadas como variantes *driver*<sup>101</sup>.

Por último, es importante mencionar dos casos en los que se detectaron a nivel germinal variantes en genes previamente relacionados con síndromes hereditarios que predisponen a la susceptibilidad a cáncer colorrectal y que se encuentran relacionados con la ruta BER. En ambos casos los síndromes a los que dan lugar alteraciones patogénicas en estos genes presentan una herencia autosómica recesiva, sin embargo, las variantes halladas fueron monoalélicas.

La variante de *MUTYH* p.(Val242Met) descrita en el paciente COAD-9 y clasificada como de significado incierto de acuerdo a las guías del *American College of Medical Genetics and Genomics and the Association for Molecular Pathology*<sup>214</sup>, se encuentra localizada en los dominios proteicos *HhH-GDP domain* y *DNA glycosylase*. Las

variantes bialélicas (homocigotas y heterocigotas compuestas) patogénicas son responsables de la poliposis asociada a *MUTYH*<sup>215</sup>. También se ha descrito que las variantes patogénicas monoalélicas (heterocigotas) en *MUTYH* confieren cierto riesgo a padecer cáncer colorrectal<sup>216,217</sup>. Es importante mencionar que en el tumor del paciente COAD-9 se encontró en un elevado porcentaje la firma SBS18 (34,51%), cuya etiología se asocia a daños en el ADN causados por especies reactivas del oxígeno<sup>17</sup>. El patrón mutacional de SBS18 es muy similar al de la firma SBS36, cuya etiología se asocia a defectos en la ruta BER ocasionados por variantes patogénicas bialélicas en *MUTYH*<sup>26,27</sup>. Algunos autores han señalado que la firma SBS18 también se relaciona con alteraciones bialélicas en *MUTYH*<sup>32,152</sup>. Además, se encontró en un 10,98% la firma SBS30, principalmente relacionada con defectos en el sistema de reparación BER debido a alteraciones en *NTHL1*<sup>23,24</sup>. En conjunto, ambos patrones mutacionales podrían indicar una alteración de la ruta BER en este paciente.

Por otro lado, en la paciente READ-31 diagnosticada de cáncer colorrectal a los 31 años y que presenta antecedentes familiares de cáncer, se encontró a nivel germinal la variante p.(Gln279\*) en heterocigosis en el gen *NTHL1*. Las variantes bialélicas patogénicas en este gen son responsables de la poliposis asociada a *NTHL1*<sup>218</sup>, mientras que las variantes monoalélicas no aumentarían el riesgo a desarrollar CCR<sup>219</sup>. De acuerdo con el estudio publicado, el tumor de la paciente no mostró la firma SBS30 asociada a alteraciones bialélicas en este gen<sup>23,24</sup>.

En ninguno de los genes candidatos propuestos que podrían desempeñar un papel supresor de la tumorigénesis se encontró un segundo evento mutacional a nivel somático. Sin embargo, se debe tener en cuenta que el análisis se centró en la búsqueda de variantes de pequeño tamaño presentes en la región codificante del genoma y las regiones flanqueantes de los exones solapantes entre los distintos kits de secuenciación de exoma empleados. No se estudió la existencia de variantes estructurales u otro tipo de alteraciones, como por ejemplo interacciones con ARNs no codificantes, o alteraciones localizadas fuera del intercepto de ambos kits de secuenciación.

En este punto es importante mencionar que en los últimos años han surgido modelos alternativos a la hipótesis de los dos eventos mutacionales en TSG como inicio de la tumorigénesis, propuesta por Alfred Knudson en 1971<sup>220</sup>. Actualmente se sabe que existen TSG cuya alteración de un único alelo en determinados contextos puede ser suficiente para promover la aparición de la neoplasia<sup>221</sup>. Esto puede darse en casos de haploinsuficiencia, como se ha visto en algunos pacientes con diagnóstico de síndrome de Li-Fraumeni, donde un único alelo alterado es suficiente para que se inicie la tumorigénesis<sup>222</sup>. También se han reportado casos donde un TSG que presenta un único alelo alterado interacciona con otros genes y cuyo efecto combinado favorece el desarrollo neoplásico, como se ha observado en algunos casos con pérdida monoalélica de *PTEN*, *BRCA1* y *BRCA2*<sup>223-225</sup>.

En 23 de los 34 pacientes de la cohorte de descubrimiento no se encontraron variantes germinales candidatas de susceptibilidad a CCR. La literatura existente apunta a un origen esporádico en la mayoría de los casos de CCR de inicio temprano asociado a un fuerte componente ambiental relacionado con hábitos de vida poco saludables, como sedentarismo, poca actividad física, o consumo excesivo de carnes rojas y procesadas y bebidas azucaradas, entre otros<sup>44,133</sup>. A pesar de ello, en estos casos, no se puede descartar la existencia de variantes genéticas cuyo efecto, bien por sí solo o en combinación con otros, favorezca la aparición de la enfermedad a edades tempranas.

En este estudio ha quedado patente la complejidad subyacente al CCR. El endofenotipado tumoral realizado a través de distintas ómicas mostró la heterogeneidad molecular existente entre tumores colorrectales en una cohorte fenotípicamente homogénea (CCR de aparición temprana  $\leq 50$  años sin alteración de la vía MMR). Además, el análisis de los distintos pacientes de manera individualizada a través de los resultados de WES (germinal y tumoral) junto a los niveles de expresión génica en mucosa colónica normal y tumoral permitió, no solo reducir el número de genes candidatos en comparación con los estudios clásicos de WES realizados únicamente a nivel de ADN, sino que la incorporación de la transcriptómica posibilitó observar el impacto ocasionado por dichas alteraciones. Como resultado de ello, se proponen 10 genes candidatos de susceptibilidad hereditaria a CCR.

Estos genes candidatos deben ser validados en una nueva cohorte, de la que se disponga tanto de datos a nivel genómico como de ARN. Actualmente este proceso se está llevando a cabo. Posteriormente, los genes/variantes que estén sobrerrepresentados en las dos fases se caracterizarán en estudios funcionales del tipo CRISPR/Cas9.

En un futuro próximo, la incorporación de datos ómicos procedentes de tejido normal y tumoral pareado de pacientes con cáncer colorrectal en bases de datos públicas, facilitará el descubrimiento de nuevos genes relacionados con la susceptibilidad hereditaria a este cáncer. Este trabajo de Tesis constituye una primera aproximación a la integración de datos transcriptómicos y datos WES, tanto germinales como tumorales, para la identificación de genes/variantes de riesgo.

Esta aproximación supone un paso más en el campo de la medicina personalizada, pues con el descubrimiento de nuevos genes de susceptibilidad hereditaria a cáncer se favorecerá no solo el tratamiento del propio paciente, sino también el manejo de sus familiares en riesgo.

# CONCLUSIONES





## 8 CONCLUSIONES

A partir de los resultados obtenidos en el presente trabajo se extraen las siguientes conclusiones:

PRIMERA. El endofenotipado tumoral, a través de las distintas ómicas, reveló la existencia de heterogeneidad molecular en una cohorte de pacientes con el mismo fenotipo clínico de cáncer colorrectal sin alteración de la vía MMR.

SEGUNDA. La integración de datos germinales y tumorales de RNA-seq y exoma facilita la priorización de variantes hereditarias de riesgo, aumentando con respecto a los estudios de WES clásicos, las probabilidades de identificar nuevos genes de susceptibilidad a cáncer colorrectal.

TERCERA. *LEF1*, *TRABD2A*, *CDH26*, *CDHR2*, *PLEKHG6*, *ARHGAP10*, *TCEA3*, *EEF2K*, *ZC3H12C* y *ADCY4*, se proponen como posibles genes candidatos de susceptibilidad hereditaria a cáncer colorrectal. Estos genes están relacionados con rutas implicadas previamente en la tumorigénesis colorrectal como Wnt/ $\beta$ -catenina, PI3K/Akt o TGF- $\beta$ , así como otros procesos celulares.

CUARTA. La caracterización molecular de los tumores junto con la integración de distintas ómicas para identificar variantes hereditarias de riesgo a CCR, permitirán establecer medidas de prevención y tratamiento específicas, constituyendo la base de la medicina personalizada.



# BIBLIOGRAFÍA



## 9 BIBLIOGRAFÍA

1. Global Cancer Observatory. <https://gco.iarc.fr/>.
2. 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).
3. Sociedad Española de Oncología Médica (SEOM): Las cifras del cáncer en España. (2023).
4. Dickinson, B. T., Kisiel, J., Ahlquist, D. A. & Grady, W. M. Molecular markers for colorectal cancer screening. *Gut* **64**, 1485–1494 (2015).
5. Pancione, M., Remo, A. & Colantuoni, V. Genetic and Epigenetic Events Generate Multiple Pathways in Colorectal Cancer Progression. *Patholog Res Int* **2012**, 11 (2012).
6. Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M. & Warusavitarne, J. Molecular pathways in colorectal cancer. *J Gastroenterol Hepatol* **27**, 1423–1431 (2012).
7. Pino, M. S. & Chung, D. C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059 (2010).
8. Fearon, E. F. & Vogelstein, B. A Genetic Model for Colorectal Tumorigenesis. *Cell* **61**, 759–767 (1990).
9. Rajagopalan, H. & Lengauer, C. Aneuploidy and cancer. *Nature* **432**, 338–341 (2004).
10. Grady, W. M. Genomic instability and colon cancer. *Cancer Metastasis Rev* **23**, 11–27 (2004).
11. Colussi, D., Brandi, G., Bazzoli, F. & Ricciardiello, L. Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int J Mol Sci* **14**, 16365–16385 (2013).
12. Mojarad, E. N., Kuppen, P. J. K., Aghdaei, H. A. & Zali, M. R. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench* **6**, 120 (2013).

13. Lao, V. V. & Grady, W. M. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8**, 686–700 (2011).
14. Toyota, M. & Issa, J. P. J. CpG island methylator phenotypes in aging and cancer. *Semin Cancer Biol* **9**, 349–357 (1999).
15. Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331**, 1553–1558 (2011).
16. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Rep* **3**, 246–259 (2013).
17. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
18. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
19. COSMIC | Mutational Signatures. <https://cancer.sanger.ac.uk/signatures/>.
20. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).
21. Islam, S. M. A. *et al.* Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell genomics* **2**, (2022).
22. Drost, J. *et al.* Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* (1979) **358**, 234–238 (2017).
23. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
24. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nature Genetics* **47**, 1402–1407 (2015).
25. Robinson, P. S. *et al.* Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. *Nature Genetics* **53**, 1434–1442 (2021).
26. Viel, A. *et al.* A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine* **20**, 39–49 (2017).

27. Pilati, C. *et al.* Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J Pathol* **242**, 10–15 (2017).
28. Tsoi, K. K. F. *et al.* Cigarette Smoking and the Risk of Colorectal Cancer: A Meta-analysis of Prospective Cohort Studies. *Clinical Gastroenterology and Hepatology* **7**, 682–688.e5 (2009).
29. Boot, A. *et al.* Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. *Genome Res* **30**, 803–813 (2020).
30. Lee-Six, H. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**:7779 **574**, 532–537 (2019).
31. Pleguezuelos-Manzano, C. *et al.* Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. *Nature* **580**:7802 **580**, 269–273 (2020).
32. Degasperi, A. *et al.* Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* (1979) **376**, (2022).
33. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nature Medicine* **21**:11 **21**, 1350–1356 (2015).
34. Dienstmann, R. *et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer* **17**:2 **17**, 79–92 (2017).
35. Yun, J. *et al.* Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells. *Science* (1979) **325**, 1555–1559 (2009).
36. Tomlinson, I. *et al.* COGENT (COlorectal cancer GENeTics) revisited. *Mutagenesis* **27**, 143–151 (2012).
37. Peters, U., Bien, S. & Zubair, N. Genetic architecture of colorectal cancer. *Gut* **64**, 1623–1636 (2015).
38. Hahn, M. M. *et al.* The genetic heterogeneity of colorectal cancer predisposition - guidelines for gene discovery. *Cellular Oncology* **39**:6 **39**, 491–510 (2016).



39. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**, 78–85 (2000).
40. Rattray, N. J. W. *et al.* Environmental influences in the etiology of colorectal cancer: the premise of metabolomics. *Curr Pharmacol Rep* **3**, 114–125 (2017).
41. Song, M. & Chan, A. T. Environmental Factors, Gut Microbiota, and Colorectal Cancer Prevention. *Clinical Gastroenterology and Hepatology* **17**, 275–289 (2019).
42. Feng, Y. L. *et al.* Dietary patterns and colorectal cancer risk: a meta-analysis. *Eur J Cancer Prev* **26**, 201–211 (2017).
43. Go, Y. M. *et al.* Reference Standardization for Mass Spectrometry and High-resolution Metabolomics Applications to Exposome Research. *Toxicol Sci* **148**, 531–543 (2015).
44. Patel, S. G., Karlitz, J. J., Yen, T., Lieu, C. H. & Boland, C. R. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. *Lancet Gastroenterol Hepatol* **7**, 262–274 (2022).
45. Low, E. E. *et al.* Risk Factors for Early-Onset Colorectal Cancer. *Gastroenterology* **159**, 492-501.e7 (2020).
46. Valle, L., Vilar, E., Tavtigian, S. V. & Stoffel, E. M. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. *J Pathol* **247**, 574–588 (2019).
47. Marley, A. R. & Nan, H. Epidemiology of colorectal cancer. *Int J Mol Epidemiol Genet* **7**, 105 (2016).
48. Schubert, S. A., Morreau, H., de Miranda, N. F. C. C. & van Wezel, T. The missing heritability of familial colorectal cancer. *Mutagenesis* **35**, 221–231 (2020).
49. Ahmad, O. & Försti, A. The complementary roles of genome-wide approaches in identifying genes linked to an inherited risk of colorectal cancer. *Hered Cancer Clin Pract* **21**, (2023).
50. Valle, L. *et al.* Update on genetic predisposition to colorectal cancer and polyposis. *Mol Aspects Med* **69**, 10–26 (2019).

51. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics* 2012 45:2 **45**, 136–144 (2012).
52. Gammon, A., Kohlmann, W. & Burt, R. Can we identify the high-risk patients to be screened? A genetic approach. *Digestion* **76**, 7–19 (2007).
53. Palles, C. *et al.* Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. *Am J Hum Genet* **109**, 953–960 (2022).
54. ClinGen. <https://clinicalgenome.org/>.
55. Lammi, L. *et al.* Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* **74**, 1043–1050 (2004).
56. Gammon, A., Jasperson, K., Kohlmann, W. & Burt, R. W. Hamartomatous polyposis syndromes. *Best Pract Res Clin Gastroenterol* **23**, 219–231 (2009).
57. Blumenthal, G. M. & Dennis, P. A. PTEN hamartoma tumor syndromes. *European Journal of Human Genetics* 2008 16:11 **16**, 1289–1300 (2008).
58. Vivanco, I. & Sawyers, C. L. The phosphatidylinositol 3-Kinase–AKT pathway in human cancer. *Nature Reviews Cancer* 2002 2:7 **2**, 489–501 (2002).
59. Lieberman, S. *et al.* Features of Patients With Hereditary Mixed Polyposis Syndrome Caused by Duplication of GREM1 and Implications for Screening and Surveillance. *Gastroenterology* **152**, 1876-1880.e1 (2017).
60. 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. *Nature Genetics* 2012 44:6 **44**, 699–703 (2012).
61. Hao, H. X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195–202 (2012).
62. Gala, M. K. *et al.* Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. *Gastroenterology* **146**, (2014).

63. Carballal, S., Balaguer, F. & IJspeert, J. E. G. Serrated polyposis syndrome; epidemiology and management. *Best Pract Res Clin Gastroenterol* **58–59**, (2022).
64. Liang, J. *et al.* APC polymorphisms and the risk of colorectal neoplasia: a HuGE review and meta-analysis. *Am J Epidemiol* **177**, 1169–1179 (2013).
65. 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**, 1208-1211.e5 (2014).
66. Olkinuora, A. *et al.* Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. *Genet Med* **21**, 1868–1873 (2019).
67. Garre, P. *et al.* BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. *Clin Genet* **87**, 582–587 (2015).
68. Phelan, C. M. *et al.* Incidence of colorectal cancer in BRCA1 and BRCA2 mutation carriers: results from a follow-up study. *Br J Cancer* **110**, 530–534 (2014).
69. Oh, M. *et al.* BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. *JNCI: Journal of the National Cancer Institute* **110**, 1178–1189 (2018).
70. Yurgelun, M. B. *et al.* Cancer susceptibility gene mutations in individuals with colorectal cancer. *Journal of Clinical Oncology* **35**, 1086–1095 (2017).
71. Pearlman, R. *et al.* Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA Oncol* **3**, 464–471 (2017).
72. Huyghe, J. R. *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nature Genetics* **2018 51:1** **51**, 76–87 (2018).
73. Fletcher, O. & Houlston, R. S. Architecture of inherited susceptibility to common cancer. *Nature Reviews Cancer* **2010 10:5** **10**, 353–361 (2010).
74. Pharoah, P. D. P., Dunning, A. M., Ponder, B. A. J. & Easton, D. F. Association studies for finding cancer-susceptibility

- genetic variants. *Nature Reviews Cancer* 2004 4:11 **4**, 850–860 (2004).
75. Tenesa, A. & Dunlop, M. G. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nature Reviews Genetics* 2009 10:6 **10**, 353–358 (2009).
76. Frampton, M. J. E. *et al.* Implications of polygenic risk for personalised colorectal cancer screening. *Annals of Oncology* **27**, 429–434 (2016).
77. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical instruments. *Genome Med* **12**, 1–11 (2020).
78. Fernandez-Rozadilla, C. *et al.* Telomere length and genetics are independent colorectal tumour risk factors in an evaluation of biomarkers in normal bowel. *British Journal of Cancer* 2018 118:5 **118**, 727–732 (2018).
79. Bonjoch, L. *et al.* BMPR2 as a novel predisposition gene for hereditary colorectal polyposis. *Gastroenterology* **165**, 162-172.e5 (2023).
80. GATK. <https://gatk.broadinstitute.org/hc/en-us>.
81. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Preprint at (2010).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. Picard Tools - By Broad Institute. <http://broadinstitute.github.io/picard/>.
84. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987 (2011).
85. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
86. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80 (2012).

87. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, (2010).
88. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15 (2013).
89. Deluca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
90. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat Biotechnol* **29**, 24 (2011).
91. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192 (2013).
92. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review with the integrative genomics viewer. *Cancer Res* **77**, e31–e34 (2017).
93. Robinson, J. T., Thorvaldsdottir, H., Turner, D. & Mesirov, J. P. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* **39**, (2023).
94. Bergstrom, E. N. *et al.* SigProfilerMatrixGenerator: A tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* **20**, 1–12 (2019).
95. Bergstrom, E. N., Barnes, M., Martincorena, I. & Alexandrov, L. B. Generating realistic null hypothesis of cancer mutational landscapes using SigProfilerSimulator. *BMC Bioinformatics* **21**, 1–10 (2020).
96. Islam, S. M. A. *et al.* Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *bioRxiv* 2020.12.13.422570 (2022) doi:10.1101/2020.12.13.422570.
97. GitHub - bioinfo-pf-curie/TMB: Tumor Mutational Burden. <https://github.com/bioinfo-pf-curie/TMB>.
98. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nature Reviews Cancer* 20:10 **20**, 555–572 (2020).

99. GitHub - xbello/dffiltering.  
<https://github.com/xbello/dffiltering>.
100. Cancer Hotspots. <https://www.cancerhotspots.org/#/home>.
101. Cancer Genome Interpreter - Identification of therapeutically actionable genomic alterations in tumors.  
<https://www.cancergenomeinterpreter.org/home>.
102. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
103. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288–4297 (2012).
104. Chen, Y. *et al.* From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* **2016 5:1438** **5**, 1438 (2016).
105. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biology* **2016 17:1** **17**, 1–19 (2016).
106. Blighe, K. PCAtools: PCAtools: Everything Principal Components Analysis. Preprint at (2023).
107. Eide, P. W., Bruun, J., Lothe, R. A. & Sveen, A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep* **7**, (2017).
108. Tian, Y. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **33**, 3982–3984 (2017).
109. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430 (2014).
110. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22–e22 (2017).
111. Nordlund, J. *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol* **14**, 1–15 (2013).

112. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
113. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 1–17 (2016).
114. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
115. KEGG: Kyoto Encyclopedia of Genes and Genomes. <https://www.genome.jp/kegg/>.
116. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* **35**, W193 (2007).
117. Neph, S. *et al.* BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
118. Ghosh, R., Oak, N. & Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol* **18**, (2017).
119. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812–3814 (2003).
120. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet* **76**, 7.20.1–7.20.41 (2013).
121. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **2014 11:4** **11**, 361–362 (2014).
122. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One* **7**, e46688 (2012).
123. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **2014 46:3** **46**, 310–315 (2014).

124. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).
125. PathCards: Pathways integration. <https://pathcards.genecards.org/>.
126. STRING: functional protein association networks. <https://string-db.org/>.
127. Caldwell, J. M. *et al.* Cadherin 26 is an alpha integrin-binding epithelial receptor regulated during allergic inflammation. *Mucosal Immunol* **10**, 1190–1201 (2017).
128. Fu, M. & Blakeshear, P. J. RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. *Nat Rev Immunol* **17**, 130–143 (2017).
129. Zhang, X. *et al.* Tikil is required for head formation via wnt cleavage-oxidation and inactivation. *Cell* **149**, 1565–1577 (2012).
130. Pinette, J. A. *et al.* Brush border protocadherin CDHR2 promotes the elongation and maximized packing of microvilli in vivo. *Mol Biol Cell* **30**, 108–118 (2019).
131. Liu, L. *et al.* ARHGAP10 Inhibits the Proliferation and Metastasis of CRC Cells via Blocking the Activity of RhoA/AKT Signaling Pathway. *Onco Targets Ther* **12**, 11507–11516 (2019).
132. D’Angelo, R. *et al.* Interaction of ezrin with the novel guanine nucleotide exchange factor PLEKHG6 promotes RhoG-dependent apical cytoskeleton rearrangements in epithelial cells. *Mol Biol Cell* **18**, 4780–4793 (2007).
133. Saraiva, M. R., Rosa, I. & Claro, I. Early-onset colorectal cancer: A review of current knowledge. <http://www.wjgnet.com/> **29**, 1289–1303 (2023).
134. Eng, C. *et al.* A comprehensive framework for early-onset colorectal cancer research. *Lancet Oncol* **23**, e116–e128 (2022).
135. Stoffel, E. M. *et al.* Germline Genetic Features of Young Individuals With Colorectal Cancer. *Gastroenterology* **154**, 897-905.e1 (2018).



136. Mork, M. E. *et al.* High prevalence of hereditary cancer syndromes in adolescents and young adults with colorectal cancer. *Journal of Clinical Oncology* **33**, 3544–3549 (2015).
137. Vuik, F. E. R. *et al.* Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut* **68**, 1820–1826 (2019).
138. Mauri, G. *et al.* Early-onset colorectal cancer in young individuals. *Mol Oncol* **13**, 109–131 (2019).
139. Chen, F. W., Sundaram, V., Chew, T. A. & Ladabaum, U. Advanced-Stage Colorectal Cancer in Persons Younger Than 50 Years Not Associated With Longer Duration of Symptoms or Time to Diagnosis. *Clinical Gastroenterology and Hepatology* **15**, 728-737.e3 (2017).
140. Willauer, A. N. *et al.* Clinical and molecular characterization of early-onset colorectal cancer. *Cancer* **125**, 2002–2010 (2019).
141. Taieb, J. *et al.* Deficient mismatch repair/microsatellite unstable colorectal cancer: Diagnosis, prognosis and treatment. *Eur J Cancer* **175**, 136–157 (2022).
142. Fennell, L. J. *et al.* APC Mutation Marks an Aggressive Subtype of BRAF Mutant Colorectal Cancers. *Cancers* 2020, Vol. 12, Page 1171 **12**, 1171 (2020).
143. Reischmann, N. *et al.* BRAFV600E drives dedifferentiation in small intestinal and colonic organoids and cooperates with mutant p53 and Apc loss in transformation. *Oncogene* 2020 39:38 **39**, 6053–6070 (2020).
144. van den Berg, I. *et al.* A panel of DNA methylation markers for the classification of consensus molecular subtypes 2 and 3 in patients with colorectal cancer. *Mol Oncol* **15**, 3348–3362 (2021).
145. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
146. Mimaki, S. *et al.* Hypermutation and unique mutational signatures of occupational cholangiocarcinoma in printing workers exposed to haloalkanes. *Carcinogenesis* **37**, 817–826 (2016).

147. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* 2017 545:7653 **545**, 175–180 (2017).
148. Saini, N. *et al.* The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. *PLoS Genet* **12**, e1006385 (2016).
149. Campbell, B. B. *et al.* Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042-1056.e10 (2017).
150. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012 487:7407 **487**, 330–337 (2012).
151. Mur, P. *et al.* Role of POLE and POLD1 in familial cancer. *Genetics in Medicine* **22**, 2089–2100 (2020).
152. Georgeson, P. *et al.* Evaluating the utility of tumour mutational signatures for identifying hereditary colorectal cancer and polyposis syndrome carriers. *Gut* **70**, 2138–2149 (2021).
153. Secrier, M. *et al.* Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nature Genetics* 2016 48:10 **48**, 1131–1141 (2016).
154. Tomlinson, I. *et al.* Whole genome sequencing of 2,023 colorectal cancers reveals mutational landscapes, new driver genes and immune interactions. (2022) doi:10.21203/RS.3.RS-2273265/V1.
155. Li, T. W.-H. *et al.* Wnt Activation and Alternative Promoter Repression of LEF1 in Colon Cancer. <https://doi.org/10.1128/MCB.00105-06> **26**, 5284–5299 (2023).
156. Xiao, L. *et al.* Lef1 enhances the progression of colonic adenocarcinoma via remodeling the cell motility associated structures. *Int J Mol Sci* **22**, 10870 (2021).
157. Lau, W. de & Clevers, H. LEF1 turns over a new leaf. *Nature Genetics* 2001 28:1 **28**, 3–4 (2001).
158. Wang, W. J. *et al.* Knockdown of Lymphoid Enhancer factor 1 Inhibits Colon Cancer Progression In Vitro and In Vivo. *PLoS One* **8**, e76596 (2013).

159. Hovanes, K. *et al.*  $\beta$ -catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nature Genetics* 2001 28:1 **28**, 53–57 (2001).
160. Shtutman, M. *et al.* The cyclin D1 gene is a target of the  $\beta$ -catenin/LEF-1 pathway. *Proc Natl Acad Sci U S A* **96**, 5522–5527 (1999).
161. Najdi, R., Holcombe, R. & Waterman, M. Wnt signaling and colon carcinogenesis: Beyond APC. *J Carcinog* **10**, (2011).
162. Stamos, J. L. & Weis, W. I. The  $\beta$ -Catenin Destruction Complex. *Cold Spring Harb Perspect Biol* **5**, a007898 (2013).
163. Hao, Y. H. *et al.* Induction of LEF1 by MYC activates the WNT pathway and maintains cell proliferation. *Cell Communication and Signaling* **17**, 1–16 (2019).
164. GTEx Portal. <https://gtexportal.org/home/>.
165. Schatoff, E. M., Leach, B. I. & Dow, L. E. Wnt Signaling and Colorectal Cancer. *Curr Colorectal Cancer Rep* **13**, 101 (2017).
166. Lachowicz-Scroggins, M. E. *et al.* Cadherin-26 (CDH26) regulates airway epithelial cell cytoskeletal structure and polarity. *Cell Discovery* 2018 4:1 **4**, 1–15 (2018).
167. Sano, K. *et al.* Protocadherins: a large family of cadherin-related molecules in central nervous system. *EMBO J* **12**, 2249 (1993).
168. Bradley, R. S., Espeseth, A. & Kintner, C. NF-protocadherin, a novel member of the cadherin superfamily, is required for *Xenopus* ectodermal differentiation. *Current Biology* **8**, 325–334 (1998).
169. Yamagata, K. *et al.* Arcadlin is a neural activity-regulated cadherin involved in long term potentiation. *Journal of Biological Chemistry* **274**, 19473–19479 (1999).
170. Ose, R., Yanagawa, T., Ikeda, S., Ohara, O. & Koga, H. PCDH24-induced contact inhibition involves downregulation of  $\beta$ -catenin signaling. *Mol Oncol* **3**, 54–66 (2009).
171. Katoh, M. & Katoh, M. Cross-talk of WNT and FGF signaling pathways at GSK3 $\beta$  to regulate  $\beta$ -catenin and SNAIL signaling cascades. <http://dx.doi.org/10.4161/cbt.5.9.3151> **5**, 1059–1064 (2006).

172. Ose, R., Oharaa, O. & Nagase, N. Galectin-1 and Galectin-3 Mediate Protocadherin-24-Dependent Membrane Localization of  $\beta$ -catenin in Colon Cancer Cell Line HCT116. *Curr Chem Genomics* **6**, 18–26 (2012).
173. Okazaki, N., Takahashi, N., Kojima, S. I., Masuho, Y. & Koga, H. Protocadherin LKC, a new candidate for a tumor suppressor of colon and liver cancers, its association with contact inhibition of cell proliferation. *Carcinogenesis* **23**, 1139–1148 (2002).
174. Wu, D., Asiedu, M. & Wei, Q. Myosin-interacting guanine exchange factor (MyoGEF) regulates the invasion activity of MDA-MB-231 breast cancer cells through activation of RhoA and RhoC. *Oncogene* 2009 28:22 **28**, 2219–2230 (2009).
175. Wu, D., Asiedu, M., Adelstein, R. S. & Wei, Q. A Novel Guanine Nucleotide Exchange Factor, MYOGEF, is Required for Cytokinesis. <http://dx.doi.org/10.4161/cc.5.11.2815> **5**, 1234–1239 (2006).
176. Ke, J. *et al.* Genetic Predisposition to Colon and Rectal Adenocarcinoma Is Mediated by a Super-enhancer Polymorphism Coactivating CD9 and PLEKHG6. *Cancer Epidemiology Biomarkers and Prevention* **29**, 850–859 (2020).
177. Wu, D., Haruta, A. & Wei, Q. GIPC1 interacts with MyoGEF and promotes MDA-MB-231 breast cancer cell invasion. *Journal of Biological Chemistry* **285**, 28643–28650 (2010).
178. Bigarella, C. L. *et al.* Post-translational modification of the RhoGTPase activating protein 21, ARHGAP21, by SUMO2/3. *FEBS Lett* **586**, 3522–3528 (2012).
179. Chang, M. T. *et al.* Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov* **8**, 174–183 (2018).
180. Luo, N. *et al.* ARHGAP10, downregulated in ovarian cancer, suppresses tumorigenicity of ovarian cancer cells. *Cell Death & Disease* 2016 7:3 **7**, e2157–e2157 (2016).
181. Lin, L. L. *et al.* ARHGAP10 inhibits the epithelial–mesenchymal transition of non-small cell lung cancer by inactivating PI3K/Akt/GSK3 $\beta$  signaling pathway. *Cancer Cell Int* **21**, 1–13 (2021).

182. Kenney, J. W., Moore, C. E., Wang, X. & Proud, C. G. Eukaryotic elongation factor 2 kinase, an unusual enzyme with multiple roles. *Adv Biol Regul* **55**, 15–27 (2014).
183. Ryazanov, A. G., Pavur, K. S. & Dorovkov, M. V. Alpha-kinases: a new class of protein kinases with a novel catalytic domain. *Curr Biol* **9**, (1999).
184. Middelbeek, J., Clark, K., Venselaar, H., Huynen, M. A. & Van Leeuwen, F. N. The alpha-kinase family: An exceptional branch on the protein kinase tree. *Cellular and Molecular Life Sciences* **67**, 875–890 (2010).
185. Chok, A. Y., Koh, Y. X., Ow, M. Y. L., Allen, J. C. & Goh, B. K. P. A systematic review and meta-analysis comparing pancreaticoduodenectomy versus limited resection for duodenal gastrointestinal stromal tumors. *Ann Surg Oncol* **21**, 3429–3438 (2014).
186. Liu, R. & Proud, C. G. Eukaryotic elongation factor 2 kinase as a drug target in cancer, and in cardiovascular and neurodegenerative diseases. *Acta Pharmacol Sin* **37**, 285–294 (2016).
187. Pyr Dit Ruys, S. *et al.* Identification of autophosphorylation sites in eukaryotic elongation factor-2 kinase. *Biochem J* **442**, 681–692 (2012).
188. Tavares, C. D. J. *et al.* Calcium/calmodulin stimulates the autophosphorylation of elongation factor 2 kinase on Thr-348 and Ser-500 to regulate its activity and calcium dependence. *Biochemistry* **51**, 2232–2245 (2012).
189. Ng, T. H. *et al.* Eukaryotic elongation factor-2 kinase expression is an independent prognostic factor in colorectal cancer. *BMC Cancer* **19**, 1–9 (2019).
190. Wang, N. *et al.* Inhibition of eukaryotic elongation factor-2 confers to tumor suppression by a herbal formulation Huanglian-Jiedu decoction in human hepatocellular carcinoma. *J Ethnopharmacol* **164**, 309–318 (2015).
191. Wang, X., Xie, J. & Proud, C. G. Eukaryotic Elongation Factor 2 Kinase (eEF2K) in Cancer. *Cancers (Basel)* **9**, (2017).

192. Faller, W. J. *et al.* mTORC1 mediated translational elongation limits intestinal tumour initiation and growth. *Nature* **517**, 497 (2015).
193. Xie, C. M., Liu, X. Y., Sham, K. W. Y., Lai, J. M. Y. & Cheng, C. H. K. Silencing of EEF2K (eukaryotic elongation factor-2 kinase) reveals AMPK-ULK1-dependent autophagy in colon cancer cells. *Autophagy* **10**, 1495–1508 (2014).
194. Wu, H., Yang, J. M., Jin, S., Zhang, H. & Hait, W. N. Elongation factor-2 kinase regulates autophagy in human glioblastoma cells. *Cancer Res* **66**, 3015–3023 (2006).
195. Delaidelli, A. *et al.* MYCN amplified neuroblastoma requires the mRNA translation regulator eEF2 kinase to adapt to nutrient deprivation. *Cell Death Differ* **24**, 1564–1576 (2017).
196. Liao, J. M. *et al.* TFIIS.h, a new target of p53, regulates transcription efficiency of pro-apoptotic bax gene. *Scientific Reports 2016 6:1* **6**, 1–10 (2016).
197. Labhart, P. & Morgan, G. T. Identification of novel genes encoding transcription elongation factor TFIIS (TCEA) in vertebrates: conservation of three distinct TFIIS isoforms in frog, mouse, and human. *Genomics* **52**, 278–288 (1998).
198. Li, J. *et al.* TCEA3 Attenuates Gastric Cancer Growth by Apoptosis Induction. *Med Sci Monit* **21**, 3241 (2015).
199. Ge, Y. *et al.* TCEA3 promotes differentiation of C2C12 cells via an Annexin A1-mediated transforming growth factor- $\beta$  signaling pathway. *J Cell Physiol* **234**, 10554–10565 (2019).
200. Kazim, N., Adhikari, A., Oh, T. J. & Davie, J. The transcription elongation factor TCEA3 induces apoptosis in rhabdomyosarcoma. *Cell Death & Disease 2020 11:1* **11**, 1–16 (2020).
201. Cha, Y., Kim, D. K., Hyun, J., Kim, S. J. & Park, K. S. TCEA3 binds to TGF- $\beta$  receptor I and induces Smad-independent, JNK-dependent apoptosis in ovarian cancer cells. *Cell Signal* **25**, 1245–1251 (2013).
202. Itatani, Y., Kawada, K. & Sakai, Y. Transforming Growth Factor- $\beta$  Signaling Pathway in Colorectal Cancer and Its Tumor Microenvironment. *Int J Mol Sci* **20**, (2019).

203. Miekus, K., Kotlinowski, J., Lichawska-Cieslar, A., Rys, J. & Jura, J. Activity of MCP1P1 RNase in tumor associated processes. *Journal of Experimental and Clinical Cancer Research* **38**, 1–11 (2019).
204. Liang, J., Song, W., Tromp, G., Kolattukudy, P. E. & Fu, M. Genome-wide survey and expression profiling of CCCH-zinc finger family reveals a functional module in macrophage activation. *PLoS One* **3**, (2008).
205. Liu, L. *et al.* Zc3h12c inhibits vascular inflammation by repressing NF- $\kappa$ B activation and pro-inflammatory gene expression in endothelial cells. *Biochem J* **451**, 55–60 (2013).
206. Marona, P. *et al.* MCP1P1 Downregulation in Clear Cell Renal Cell Carcinoma Promotes Vascularization and Metastatic Progression. *Cancer Res* **77**, 4905–4920 (2017).
207. Lu, W. *et al.* MCP1P1 Selectively Destabilizes Transcripts Associated with an Antiapoptotic Gene Expression Program in Breast Cancer Cells That Can Elicit Complete Tumor Regression. *Cancer Res* **76**, 1429–1440 (2016).
208. Guo, R. *et al.* Targeting Adenylate Cyclase Family: New Concept of Targeted Cancer Therapy. *Front Oncol* **12**, (2022).
209. Seamon, K. B., Padgett, W. & Daly, J. W. Forskolin: unique diterpene activator of adenylate cyclase in membranes and in intact cells. *Proc Natl Acad Sci U S A* **78**, 3363–3367 (1981).
210. Ma, M. *et al.* MicroRNA-23a-3p Inhibits Mucosal Melanoma Growth and Progression through Targeting Adenylate Cyclase 1 and Attenuating cAMP and MAPK Pathways. *Theranostics* **9**, 945–960 (2019).
211. Sanchez-Collado, J., Lopez, J. J., Jardin, I., Salido, G. M. & Rosado, J. A. Cross-Talk Between the Adenylate Cyclase/cAMP Pathway and Ca<sup>2+</sup> Homeostasis. *Rev Physiol Biochem Pharmacol* **179**, 73–116 (2021).
212. Pierre, S., Eschenhagen, T., Geisslinger, G. & Scholich, K. Capturing adenylate cyclases as potential drug targets. *Nat Rev Drug Discov* **8**, 321–335 (2009).

213. Fan, Y. *et al.* Epigenetic identification of ADCY4 as a biomarker for breast cancer: an integrated analysis of adenylate cyclases. *Epigenomics* **11**, 1561–1579 (2019).
214. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405–424 (2015).
215. Aretz, S. *et al.* MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int J Cancer* **119**, 807–814 (2006).
216. Win, A. K. *et al.* Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. *Int J Cancer* **129**, 2256–2262 (2011).
217. Barreiro, R. A. S. *et al.* Monoallelic deleterious MUTYH germline variants as a driver for tumorigenesis. *J Pathol* **256**, 214–222 (2022).
218. 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).
219. Elsayed, F. A. *et al.* Monoallelic NTHL1 Loss-of-Function Variants and Risk of Polyposis and Colorectal Cancer. *Gastroenterology* **159**, 2241–2243.e6 (2020).
220. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820–823 (1971).
221. Berger, A. H., Knudson, A. G. & Pandolfi, P. P. A continuum model for tumour suppression. *Nature* **476**, 163–169 (2011).
222. Varley, J. M., Evans, D. G. R. & Birch, J. M. Li-Fraumeni syndrome--a molecular and clinical review. *Br J Cancer* **76**, 1–14 (1997).
223. Park, S., Supek, F. & Lehner, B. Higher order genetic interactions switch cancer genes from two-hit to one-hit drivers. *Nat Commun* **12**, (2021).
224. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).



225. Bellacosa, A. *et al.* Altered gene expression in morphologically normal epithelial cells from heterozygous carriers of BRCA1 or BRCA2 mutations. *Cancer Prev Res (Phila)* **3**, 48–61 (2010).

# **MATERIAL SUPLEMENTARIO**



## 10 MATERIAL SUPLEMENTARIO

**Tabla suplementaria 1. Datos clínicos de los pacientes incluidos en la cohorte de descubrimiento TOGETHER. 20** pacientes diagnosticados de CCR ( $\leq 50$  años) con expresión conservada de las proteínas reparadoras MMR y que no han recibido tratamiento previo de la enfermedad.

| Paciente ID | Edad de diagnóstico CCR (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer   | Procedencia de la muestra |
|-------------|--------------------------------|--------|----------------------|---------|---|---------------------------|
| 102         | 46                             | Mujer  | Sigma                | IV      | Hermano (melanoma, 40)  | Endoscopia                |
| 103         | 49                             | Hombre | Recto                | IV      | No  | Endoscopia                |
| 104         | 40                             | Hombre | Recto                | IV      | No  | Endoscopia                |
| 105         | 47                             | Hombre | Sigma                | III     | Tía materna (ovario, 60; mama, 65)  | Endoscopia                |
| 107         | 49                             | Mujer  | Colon transverso     | III     | Padre (próstata, 72; CCR, 78)   | Endoscopia                |
| 108         | 47                             | Hombre | Sigma                | IV      | Abuela materna (CCR, 80)  | Endoscopia                |
| 109         | 47                             | Hombre | Recto                | III     | Madre (mama, 74), padre (mieloma, 71), tío paterno (pulmón, 64), tío paterno (pulmón, 67)                                       | Endoscopia                |
| 202         | 48                             | Hombre | Colon ascendente     | II      | Madre (endometrio, 42), hermana (endometrio, 37; tracto urinario, 46), abuela materna (gástrico, 72), tío paterno (cerebro, 66) | Endoscopia                |
| 204         | 41                             | Hombre | Colon transverso     | II      | Abuelo paterno (cáncer desconocido, 55)   | Endoscopia                |
| 209         | 50                             | Hombre | Sigma                | III     | Hermano (testículos, 30), tío paterno (CCR, 72)   | Cirugía                   |
| 302         | 49                             | Hombre | Sigma                | III     | No  | Endoscopia                |
| 303         | 47                             | Mujer  | Recto                | II      | No  | Endoscopia                |

Continuación tabla suplementaria 1.

| Paciente (ID) | Edad de diagnóstico CCR (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer  | Procedencia de la muestra |
|---------------|--------------------------------|--------|----------------------|---------|--|---------------------------|
| 304           | 30                             | Hombre | Recto                | IV      | Primo segundo materno (CCR, 50; CCR, 60)   | Endoscopia                |
| 305           | 49                             | Mujer  | Sigma                | III     | Madre (pulmón, 64; mama, edad desconocida), padre (pulmón, 56), tío paterno (pulmón, 55), prima paterna (leucemia, 56) | Endoscopia                |
| 307           | 48                             | Mujer  | Unión Recto-Sigma    | IV      | No   | Endoscopia                |
| 308           | 50                             | Mujer  | Unión Recto-Sigma    | II      | Padre (gástrico, 72)   | Endoscopia                |
| 310           | 46                             | Hombre | Sigma                | I       | No   | Endoscopia                |
| 401           | 42                             | Hombre | Ciego                | III     | Padre (CCR, 72), tío paterno (pulmón, 65)  | Cirugía                   |
| 402           | 47                             | Hombre | Recto                | III     | Tío paterno (cáncer desconocido, 42)   | Cirugía                   |
| 501           | 43                             | Hombre | Sigma                | II      | Madre (mama, 63), padre (pulmón, 50), abuelo materno (próstata, 72)  | Cirugía                   |

**Tabla suplementaria 2. Datos clínicos de la cohorte de descubrimiento TCGA.** 26 pacientes con diagnóstico de CCR  $\leq$  50 años procedentes de los grupos de pacientes de las cohortes COAD y READ. IHQ: datos inmunohistoquímicos de las proteínas reparadoras MMR; -: datos no disponibles; MSS: estabilidad de microsatélites (nº de *loci* microsatélite alterados/ nº de *loci* microsatélites testados, en el caso de haber sido reportados); pMMR: expresión conservada por IHQ de las proteínas reparadoras MMR; d: indica deficiencia de expresión por IHQ de las proteínas reparadoras a las que precede.

| Paciente ID | Edad de diagnóstico (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer | Fenotipo tumoral |                                 |
|-------------|----------------------------|--------|----------------------|---------|-----------------------------|------------------|---------------------------------|
|             |                            |        |                      |         |                             | IHQ              | Inestabilidad de microsatélites |
| COAD-1      | 46                         | Hombre | Ángulo esplénico     | IV      | No                          | pMMR             | MSS (0/5)                       |
| COAD-2      | 40                         | Mujer  | Sigma                | II      | No                          | pMMR             | MSS (0/4)                       |
| COAD-3      | 37                         | Hombre | Ciego                | IV      | -                           | -                | -                               |
| COAD-4      | 49                         | Hombre | -                    | IV      | -                           | -                | -                               |
| COAD-5      | 43                         | Hombre | Sigma                | II      | No                          | -                | -                               |
| COAD-6      | 46                         | Mujer  | Sigma                | III     | -                           | -                | -                               |
| COAD-7      | 45                         | Mujer  | Sigma                | III     | -                           | -                | MSS (0/10)                      |
| COAD-8      | 45                         | Mujer  | Sigma                | IV      | -                           | -                | MSS (0/10)                      |
| COAD-9      | 44                         | Hombre | Colon ascendente     | II      | No                          | -                | -                               |
| COAD-10     | 36                         | Mujer  | Sigma                | I       | No                          | -                | MSS (0/5)                       |
| COAD-11     | 48                         | Mujer  | Colon ascendente     | III     | Sí                          | dMLH1/PMS2       | -                               |
| COAD-12     | 46                         | Mujer  | Sigma                | II      | Sí                          | -                | -                               |
| COAD-13     | 48                         | Mujer  | Colon ascendente     | I       | No                          | -                | -                               |
| COAD-14     | 39                         | Hombre | Ángulo hepático      | II      | Sí                          | dMSH2/MSH6       | -                               |
| COAD-15     | 45                         | Mujer  | Colon transverso     | -       | -                           | -                | -                               |
| COAD-16     | 34                         | Hombre | Colon transverso     | II      | Sí                          | -                | MSS                             |
| COAD-17     | 49                         | Mujer  | Sigma                | IV      | Sí                          | -                | -                               |
| COAD-18     | 45                         | Hombre | Ciego                | I       | No                          | -                | -                               |
| READ-1      | 50                         | Hombre | Unión Recto-Sigma    | III     | No                          | pMMR             | -                               |
| READ-2      | 48                         | Hombre | Unión Recto-Sigma    | I       | No                          | -                | MSS                             |
| READ-3      | 41                         | Mujer  | Recto                | IV      | -                           | -                | -                               |

**Continuación tabla suplementaria 2.**

| Paciente ID | Edad de diagnóstico (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer | Fenotipo tumoral |                                 |
|-------------|----------------------------|--------|----------------------|---------|-----------------------------|------------------|---------------------------------|
|             |                            |        |                      |         |                             | IHQ              | Inestabilidad de microsatélites |
| READ-4      | 49                         | Mujer  | Unión Recto-Sigma    | III     | Sí                          | -                | -                               |
| READ-5      | 43                         | Hombre | Unión Recto-Sigma    | III     | No                          | -                | -                               |
| READ-6      | 48                         | Mujer  | Recto                | III     | Sí                          | -                | -                               |
| READ-7      | 33                         | Hombre | Recto                | III     | -                           | -                | -                               |
| READ-8      | 37                         | Mujer  | Unión Recto-Sigma    | -       | No                          | -                | MSS                             |

**Tabla suplementaria 3. Listado de los individuos procedentes de la base de datos GTEEx.** Los datos de expresión génica se emplearon como referencia de tejido colónico normal junto a los datos de mucosa colónica sana de los individuos de la cohorte TOGETHER. Datos de edad, sexo y localización de la muestra de mucosa colónica normal de 47 individuos sanos.

| Individuo ID | Edad (años) | Sexo   | Localización     |
|--------------|-------------|--------|------------------|
| GTEEx-1      | 33          | Mujer  | Colon transverso |
| GTEEx-2      | 36          | Hombre | Colon transverso |
| GTEEx-3      | 50          | Hombre | Colon transverso |
| GTEEx-4      | 43          | Mujer  | Colon transverso |
| GTEEx-5      | 22          | Mujer  | Colon transverso |
| GTEEx-6      | 48          | Mujer  | Colon transverso |
| GTEEx-7      | 44          | Mujer  | Colon transverso |
| GTEEx-8      | 44          | Hombre | Colon transverso |
| GTEEx-9      | 27          | Mujer  | Colon transverso |
| GTEEx-10     | 41          | Mujer  | Colon transverso |
| GTEEx-11     | 22          | Hombre | Colon transverso |
| GTEEx-12     | 22          | Mujer  | Colon transverso |
| GTEEx-13     | 50          | Hombre | Colon transverso |
| GTEEx-14     | 44          | Hombre | Colon transverso |
| GTEEx-15     | 31          | Mujer  | Colon transverso |
| GTEEx-16     | 45          | Mujer  | Colon transverso |
| GTEEx-17     | 27          | Mujer  | Colon transverso |
| GTEEx-18     | 50          | Hombre | Colon transverso |
| GTEEx-19     | 33          | Mujer  | Colon transverso |
| GTEEx-20     | 47          | Hombre | Colon transverso |
| GTEEx-21     | 25          | Hombre | Colon transverso |
| GTEEx-22     | 49          | Hombre | Colon transverso |
| GTEEx-23     | 43          | Hombre | Colon transverso |
| GTEEx-24     | 50          | Hombre | Colon transverso |
| GTEEx-25     | 46          | Mujer  | Colon transverso |
| GTEEx-26     | 24          | Hombre | Colon transverso |
| GTEEx-27     | 32          | Hombre | Colon transverso |
| GTEEx-28     | 42          | Mujer  | Colon transverso |
| GTEEx-29     | 29          | Mujer  | Colon transverso |
| GTEEx-30     | 46          | Hombre | Colon transverso |
| GTEEx-31     | 48          | Mujer  | Colon transverso |
| GTEEx-32     | 47          | Hombre | Colon transverso |



Continuación tabla suplementaria 3.

| Paciente ID | Edad (años) | Sexo   | Localización     |
|-------------|-------------|--------|------------------|
| GTEEx-33    | 27          | Hombre | Colon transverso |
| GTEEx-34    | 21          | Mujer  | Colon transverso |
| GTEEx-35    | 21          | Hombre | Colon transverso |
| GTEEx-36    | 23          | Hombre | Colon transverso |
| GTEEx-37    | 47          | Mujer  | Colon transverso |
| GTEEx-38    | 26          | Mujer  | Colon transverso |
| GTEEx-39    | 50          | Hombre | Colon transverso |
| GTEEx-40    | 49          | Hombre | Colon transverso |
| GTEEx-41    | 44          | Hombre | Colon transverso |
| GTEEx-42    | 33          | Mujer  | Colon transverso |
| GTEEx-43    | 22          | Hombre | Colon transverso |
| GTEEx-44    | 44          | Hombre | Colon transverso |
| GTEEx-45    | 37          | Hombre | Colon transverso |
| GTEEx-46    | 44          | Hombre | Colon transverso |
| GTEEx-47    | 50          | Hombre | Colon transverso |

**Tabla suplementaria 4. Datos clínicos de los pacientes obtenidos de la base de datos TCGA cuyos datos de RNA-seq se emplearon para la clasificación en CMS.** 55 pacientes con diagnóstico de CCR a edades comprendidas entre los 51 y 65 años procedentes de los grupos de pacientes de las cohortes COAD y READ. IHQ: datos inmunohistoquímicos de las proteínas reparadoras MMR; -: datos no disponibles; MSI: inestabilidad de microsatélites (nº de *loci* microsatélite alterados/ nº de *loci* microsatélites testados, en el caso de haber sido reportados); MSS: estabilidad de microsatélites (nº de *loci* microsatélite alterados/ nº de *loci* microsatélites testados, en el caso de haber sido reportados); pMMR: expresión conservada por IHQ de las proteínas reparadoras MMR; d: indica deficiencia de expresión por IHQ de las proteínas reparadoras a las que precede.

| Paciente ID | Edad de diagnóstico (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer | Fenotipo tumoral     |                                 |
|-------------|----------------------------|--------|----------------------|---------|-----------------------------|----------------------|---------------------------------|
|             |                            |        |                      |         |                             | IHQ                  | Inestabilidad de microsatélites |
| COAD-19     | 55                         | Hombre | Ángulo hepático      | III     | No                          | -                    | -                               |
| COAD-20     | 63                         | Mujer  | Sigma                | IV      | Sí                          | -                    | -                               |
| COAD-21     | 55                         | Mujer  | Colon transverso     | III     | No                          | -                    | -                               |
| COAD-22     | 58                         | Mujer  | Ciego                | I       | -                           | dMLH1/PMS2           | MSI (9/10)                      |
| COAD-23     | 63                         | Hombre | Sigma                | III     | No                          | -                    | -                               |
| COAD-24     | 58                         | Hombre | Sigma                | II      | No                          | dMLH1/MSH2/MSH6/PMS2 | -                               |
| COAD-25     | 62                         | Mujer  | Colon ascendente     | II      | -                           | -                    | -                               |
| COAD-26     | 61                         | Hombre | Sigma                | II      | -                           | -                    | -                               |
| COAD-27     | 64                         | Hombre | Colon transverso     | III     | No                          | -                    | -                               |
| COAD-28     | 56                         | Mujer  | Sigma                | IV      | Sí                          | pMMR                 | MSS (0/5)                       |
| COAD-29     | 58                         | Hombre | Ciego                | -       | Sí                          | dMLH1/PMS2           | MSS                             |
| COAD-30     | 65                         | Mujer  | Colon descendente    | III     | No                          | pMMR                 | MSS (0/5)                       |
| COAD-31     | 61                         | Mujer  | Sigma                | III     | -                           | -                    | -                               |
| COAD-32     | 56                         | Hombre | -                    | IV      | Sí                          | pMMR                 | MSS (0/5)                       |
| COAD-33     | 55                         | Mujer  | Ciego                | IV      | No                          | -                    | -                               |
| COAD-34     | 58                         | Mujer  | Ciego                | II      | No                          | dPMS2                | MSI                             |
| COAD-35     | 61                         | Hombre | Ciego                | I       | No                          | -                    | -                               |
| COAD-36     | 62                         | Hombre | Ciego                | III     | Sí                          | -                    | -                               |

Continuación tabla suplementaria 4.

| Paciente ID | Edad de diagnóstico (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer | Fenotipo tumoral     |                                 |
|-------------|----------------------------|--------|----------------------|---------|-----------------------------|----------------------|---------------------------------|
|             |                            |        |                      |         |                             | IHQ                  | Inestabilidad de microsatélites |
| COAD-37     | 60                         | Hombre | Ciego                | I       | Sí                          | dMLH1/MSH2/MSH6/PMS2 |                                 |
| COAD-38     | 52                         | Mujer  | Sigma                | IV      | No                          | -                    | -                               |
| COAD-39     | 62                         | Hombre | Colon descendente    | II      | No                          | -                    | -                               |
| COAD-40     | 59                         | Mujer  | Sigma                | I       | -                           | -                    | MSS                             |
| COAD-41     | 51                         | Mujer  | Sigma                | III     | No                          | pMMR                 | MSS                             |
| COAD-42     | 59                         | Mujer  | Sigma                | IV      | No                          | -                    | -                               |
| COAD-43     | 63                         | Mujer  | Ciego                | II      | No                          | -                    | MSS                             |
| COAD-44     | 51                         | Mujer  | Ciego                | I       | No                          | -                    | MSS                             |
| COAD-45     | 58                         | Mujer  | Colon descendente    | II      | -                           | -                    | -                               |
| COAD-46     | 61                         | Hombre | Sigma                | IV      | -                           | -                    | -                               |
| COAD-47     | 51                         | Mujer  | Sigma                | I       | No                          | -                    | MSS                             |
| COAD-48     | 51                         | Mujer  | Colon descendente    | I       | No                          | -                    | -                               |
| COAD-49     | 51                         | Mujer  | Ángulo hepático      | III     | -                           | -                    | MSS                             |
| COAD-50     | 62                         | Hombre | Ciego                | III     | No                          | -                    | -                               |
| COAD-51     | 60                         | Hombre | Ciego                | II      | No                          | dMSH2/MSH6           | MSI (4/5)                       |
| COAD-52     | 51                         | Hombre | Colon transverso     | II      | No                          | -                    | -                               |
| COAD-53     | 61                         | Hombre | Ciego                | II      | No                          | -                    | -                               |
| COAD-54     | 60                         | Hombre | Ciego                | III     | No                          | -                    | -                               |
| COAD-55     | 54                         | Mujer  | Sigma                | IV      | No                          | -                    | -                               |
| COAD-56     | 57                         | Hombre | Sigma                | I       | No                          | -                    | MSS                             |
| COAD-57     | 65                         | Hombre | Colon descendente    | III     | No                          | -                    | -                               |
| COAD-58     | 60                         | Hombre | Colon ascendente     | III     | -                           | -                    | MSS                             |
| READ-9      | 59                         | Hombre | Unión Recto-Sigma    | III     | No                          | -                    | -                               |

Continuación tabla suplementaria 4.

| Paciente ID | Edad de diagnóstico (años) | Sexo   | Localización tumoral | Estadio | Historia familiar de cáncer | Fenotipo tumoral |                                 |
|-------------|----------------------------|--------|----------------------|---------|-----------------------------|------------------|---------------------------------|
|             |                            |        |                      |         |                             | IHQ              | Inestabilidad de microsatélites |
| READ-10     | 57                         | Hombre | Unión Recto-Sigma    | I       | No                          | -                | -                               |
| READ-11     | 57                         | Hombre | Recto                | II      | No                          | -                | -                               |
| READ-12     | 62                         | Hombre | Unión Recto-Sigma    | IV      | Sí                          | -                | -                               |
| READ-13     | 57                         | Mujer  | Unión Recto-Sigma    | IV      | No                          | -                | -                               |
| READ-14     | 64                         | Mujer  | Unión Recto-Sigma    | II      | No                          | -                | -                               |
| READ-15     | 60                         | Hombre | Recto                | II      | No                          | -                | -                               |
| READ-16     | 57                         | Hombre | Unión Recto-Sigma    | II      | No                          | -                | -                               |
| READ-17     | 54                         | Hombre | Recto                | II      | Sí                          | -                | MSS                             |
| READ-18     | 64                         | Mujer  | Unión Recto-Sigma    | III     | No                          | -                | -                               |
| READ-19     | 58                         | Hombre | Recto                | IV      | No                          | -                | -                               |
| READ-20     | 59                         | Hombre | Unión Recto-Sigma    | II      | No                          | -                | MSS                             |
| READ-21     | 57                         | Mujer  | Unión Recto-Sigma    | IV      | No                          | -                | -                               |
| READ-22     | 60                         | Mujer  | Recto                | II      | -                           | -                | -                               |
| READ-23     | 63                         | Hombre | Unión Recto-Sigma    | II      | No                          | -                | MSS                             |

**Tabla suplementaria 5. Características clínicas de la cohorte de réplica.** 39 pacientes procedentes de 38 familias no relacionadas con diagnóstico de CCR ( $\leq 50$  años) obtenidos de la base de datos TCGA (cohortes COAD y READ) y de una base de datos interna. IHQ: datos inmunohistoquímicos de las proteínas reparadoras MMR; -: datos no disponibles; MSS: estabilidad de microsatélites (nº de loci microsatélite alterados/ nº de loci microsatélites testados, en el caso de haber sido reportados); pMMR: expresión conservada por IHQ de las proteínas reparadoras MMR; d: indica deficiencia de expresión por IHQ de las proteínas reparadoras a las que precede.

| Paciente ID | Familia ID | Edad de diagnóstico (años) | Sexo   | Historia familiar de cáncer | Fenotipo tumoral     |                                 |
|-------------|------------|----------------------------|--------|-----------------------------|----------------------|---------------------------------|
|             |            |                            |        |                             | IHQ                  | Inestabilidad de microsatélites |
| COAD-59     | Fam-1      | 43                         | Mujer  | Sí                          | pMMR                 | MSS (0/4)                       |
| COAD-60     | Fam-2      | 43                         | Hombre | -                           | -                    | -                               |
| COAD-61     | Fam-3      | 47                         | Hombre | Sí                          | dMLH1/MSH2/MSH6/PMS2 | MSS                             |
| COAD-62     | Fam-4      | 48                         | Mujer  | -                           | -                    | MSS (0/10)                      |
| COAD-63     | Fam-5      | 47                         | Hombre | No                          | dMLH1/MSH2/MSH6/PMS2 | -                               |
| COAD-64     | Fam-6      | 34                         | Mujer  | No                          | dMLH1/MSH2/MSH6/PMS2 | -                               |
| COAD-65     | Fam-7      | 49                         | Hombre | -                           | -                    | -                               |
| COAD-66     | Fam-8      | 50                         | Mujer  | No                          | -                    | MSS (0/5)                       |
| COAD-67     | Fam-9      | 50                         | Hombre | No                          | -                    | MSS (1/6)                       |
| COAD-68     | Fam-10     | 41                         | Mujer  | -                           | -                    | -                               |
| COAD-69     | Fam-11     | 49                         | Mujer  | Sí                          | dMSH6/PMS2           | -                               |
| COAD-70     | Fam-12     | 37                         | Mujer  | No                          | pMMR                 | MSS                             |
| READ-24     | Fam-13     | 41                         | Mujer  | -                           | -                    | -                               |
| READ-25     | Fam-14     | 48                         | Mujer  | No                          | -                    | -                               |
| READ-26     | Fam-15     | 48                         | Hombre | No                          | -                    | MSS (0/5)                       |
| READ-27     | Fam-16     | 43                         | Hombre | No                          | -                    | MSS                             |
| READ-28     | Fam-17     | 46                         | Hombre | No                          | -                    | -                               |
| READ-29     | Fam-18     | 41                         | Hombre | -                           | -                    | -                               |
| READ-30     | Fam-19     | 44                         | Hombre | -                           | -                    | -                               |
| READ-31     | Fam-20     | 31                         | Mujer  | Sí                          | -                    | -                               |
| 1001        | Fam-21     | 43                         | Mujer  | No                          | pMMR                 | -                               |
| 1046        | Fam-22     | 42                         | Hombre | No                          | pMMR                 | -                               |
| 1055        | Fam-23     | 50                         | Hombre | No                          | pMMR                 | -                               |
| 3024        | Fam-24     | 39                         | Mujer  | No                          | pMMR                 | -                               |
| 3131        | Fam-25     | 38                         | Hombre | No                          | pMMR                 | -                               |

Continuación tabla suplementaria 5.

| Paciente ID | Familia ID | Edad de diagnóstico (años) | Sexo   | Historia familiar de cáncer     | Fenotipo tumoral |                                 |
|-------------|------------|----------------------------|--------|---------------------------------|------------------|---------------------------------|
|             |            |                            |        |                                 | IHQ              | Inestabilidad de microsatélites |
| 3143        | Fam-26     | 47                         | Hombre | Hermana (CCR), padre (leucemia) | pMMR             | -                               |
| 3146        | Fam-26     | 44                         | Mujer  |                                 | pMMR             | -                               |
| 3210        | Fam-27     | 48                         | Hombre | No                              | pMMR             | -                               |
| 5027        | Fam-28     | 44                         | Mujer  | No                              | pMMR             | -                               |
| 5033        | Fam-29     | 36                         | Hombre | No                              | pMMR             | -                               |
| 16014       | Fam-30     | 46                         | Mujer  | No                              | pMMR             | -                               |
| 16044       | Fam-31     | 50                         | Hombre | Abuelo (CCR), hermana (mama)    | pMMR             | -                               |
| 16067       | Fam-32     | 34                         | Mujer  | No                              | pMMR             | -                               |
| 16097       | Fam-33     | 47                         | Hombre | Abuelo (CCR)                    | pMMR             | -                               |
| 16112       | Fam-34     | 40                         | Hombre | No                              | pMMR             | -                               |
| 16138       | Fam-35     | 49                         | Hombre | No                              | pMMR             | -                               |
| 16151       | Fam-36     | 43                         | Mujer  | No                              | pMMR             | -                               |
| CC27        | Fam-37     | 47                         | Hombre | No                              | pMMR             | -                               |
| CC93        | Fam-38     | 39                         | Mujer  | No                              | pMMR             | -                               |

**Tabla suplementaria 6. Firma mutacional tumoral.** Patrones de firma mutacional de cada uno de los individuos pertenecientes a la cohorte de descubrimiento obtenidas empleando SigProfilerAssignment<sup>94-96</sup> y tomando como referencia la base de datos COSMIC v3.3<sup>19</sup>.

|   | 102    | 103    | 104    | 105    | 107    |
|---|--------|--------|--------|--------|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 29,44% | 18,28% | 16,48% | 40,66% | 38,95% |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        |        |        |        | 6,32%  |
| SBS5: Origen desconocido (firma plana)  | 16,11% | 26,88% | 20,22% | 42,86% | 34,74% |
| SBS6: Sistema de reparación MMR defectuoso  |        |        |        |        |        |
| SBS7a: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7b: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7d: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS8: Origen desconocido  |        |        |        |        |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |        |        |        |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |        |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |        |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |        |        |        |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        |        | 11,99% |        |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |        |        |        |        |
| SBS15: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS16: Origen desconocido   |        |        |        | 16,48% |        |
| SBS17b: Origen desconocido  |        |        |        |        |        |
| SBS18: Daño por especies reactivas del oxígeno  |        |        |        |        |        |
| SBS26: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS28: Origen desconocido   |        |        |        |        |        |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        |        |        |        |        |
| SBS37: Origen desconocido   |        |        |        |        |        |
| SBS40: Origen desconocido (firma plana)   |        |        |        |        |        |
| SBS41: Origen desconocido   |        | 36,20% |        |        |        |
| SBS42: Exposición a haloalcanos   | 17,78% |        | 18,35% |        | 20,00% |
| SBS44: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |        | 21,35% |        |        |
| SBS89: Origen desconocido   |        |        |        |        |        |
| SBS92: Tabaco   |        |        |        |        |        |
| SBS93: Origen desconocido   | 36,67% |        |        |        |        |
| SBS94: Origen desconocido   |        | 18,64% |        |        |        |
| Posible artefacto de secuenciación  |        |        | 11,61% |        |        |
| Tratamiento quimioterápico  |        |        |        |        |        |

Continuación tabla suplementaria 6.

|   | 108    | 109    | 202    | 204    | 209    |
|---|--------|--------|--------|--------|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 32,91% | 44,39% | 23,91% | 40,82% | 38,46% |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        |        |        |        |        |
| SBS5: Origen desconocido (firma plana)  | 67,09% | 36,45% | 11,19% | 11,22% | 61,54% |
| SBS6: Sistema de reparación MMR defectuoso  |        |        |        |        |        |
| SBS7a: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7b: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7d: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS8: Origen desconocido  |        |        |        |        |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |        |        |        |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |        |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |        |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |        |        |        |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        |        |        |        |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |        |        |        |        |
| SBS15: Sistema de reparación MMR defectuoso   |        |        |        | 19,39% |        |
| SBS16: Origen desconocido   |        |        |        |        |        |
| SBS17b: Origen desconocido  |        |        |        |        |        |
| SBS18: Daño por especies reactivas del oxígeno  |        |        |        |        |        |
| SBS26: Sistema de reparación MMR defectuoso   |        |        | 31,28% |        |        |
| SBS28: Origen desconocido   |        |        |        |        |        |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        |        |        |        |        |
| SBS37: Origen desconocido   |        |        |        |        |        |
| SBS40: Origen desconocido (firma plana)   |        |        |        |        |        |
| SBS41: Origen desconocido   |        |        |        |        |        |
| SBS42: Exposición a haloalcanos   |        | 19,16% |        |        |        |
| SBS44: Sistema de reparación MMR defectuoso   |        |        | 33,62% |        |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |        |        |        |        |
| SBS89: Origen desconocido   |        |        |        |        |        |
| SBS92: Tabaco   |        |        |        |        |        |
| SBS93: Origen desconocido   |        |        |        |        |        |
| SBS94: Origen desconocido   |        |        |        |        |        |
| Posible artefacto de secuenciación  |        |        |        | 28,57% |        |
| Tratamiento quimioterápico  |        |        |        |        |        |



Continuación tabla suplementaria 6.

|   | 302    | 303    | 304    | 305    | 307    |
|---|--------|--------|--------|--------|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 5,76%  | 31,78% | 26,09% | 32,47% | 16,67% |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        |        |        |        |        |
| SBS5: Origen desconocido (firma plana)  | 22,66% | 30,62% | 36,96% | 29,89% | 21,21% |
| SBS6: Sistema de reparación MMR defectuoso  |        |        | 15,76% |        |        |
| SBS7a: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7b: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7d: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS8: Origen desconocido  |        |        |        |        |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |        |        |        |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        | 7,61%  |        |        |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |        |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |        |        |        |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        |        |        |        |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |        |        |        |        |
| SBS15: Sistema de reparación MMR defectuoso   | 5,76%  |        |        |        | 21,21% |
| SBS16: Origen desconocido   |        |        |        |        |        |
| SBS17b: Origen desconocido  |        |        |        |        |        |
| SBS18: Daño por especies reactivas del oxígeno  |        |        |        | 18,82% |        |
| SBS26: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS28: Origen desconocido   |        |        |        |        |        |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        |        |        |        |        |
| SBS37: Origen desconocido   |        |        |        |        |        |
| SBS40: Origen desconocido (firma plana)   |        |        |        |        |        |
| SBS41: Origen desconocido   |        |        |        |        |        |
| SBS42: Exposición a haloalcanos   |        |        |        | 18,82% | 18,18% |
| SBS44: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |        |        |        |        |
| SBS89: Origen desconocido   |        | 37,60% |        |        |        |
| SBS92: Tabaco   | 22,66% |        |        |        |        |
| SBS93: Origen desconocido   | 43,17% |        |        |        |        |
| SBS94: Origen desconocido   |        |        |        |        |        |
| Posible artefacto de secuenciación  |        |        | 13,59% |        | 22,73% |
| Tratamiento quimioterápico  |        |        |        |        |        |

Continuación tabla suplementaria 6.

|   | 308    | 310    | 401    | 402    | 501    |
|---|--------|--------|--------|--------|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 15,77% | 5,22%  | 28,29% | 32,37% | 23,16% |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        | 3,73%  |        |        |        |
| SBS5: Origen desconocido (firma plana)  | 54,50% | 25,12% | 42,23% | 31,21% | 12,13% |
| SBS6: Sistema de reparación MMR defectuoso  |        |        |        |        |        |
| SBS7a: Exposición a luz ultravioleta  |        |        |        | 11,56% |        |
| SBS7b: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7d: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS8: Origen desconocido  |        |        |        |        |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |        |        |        |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          | 7,66%  |        |        |        | 6,62%  |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |        |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |        |        |        |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        | 7,21%  |        |        |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |        |        |        |        |
| SBS15: Sistema de reparación MMR defectuoso   | 22,07% | 5,97%  | 20,32% | 24,86% |        |
| SBS16: Origen desconocido   |        |        |        |        |        |
| SBS17b: Origen desconocido  |        |        |        |        |        |
| SBS18: Daño por especies reactivas del oxígeno  |        |        |        |        |        |
| SBS26: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS28: Origen desconocido   |        |        |        |        |        |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        |        |        |        |        |
| SBS37: Origen desconocido   |        |        |        |        |        |
| SBS40: Origen desconocido (firma plana)   |        |        |        |        |        |
| SBS41: Origen desconocido   |        |        |        |        |        |
| SBS42: Exposición a haloalcanos   |        |        | 9,16%  |        |        |
| SBS44: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |        |        |        |        |
| SBS89: Origen desconocido   |        |        |        |        |        |
| SBS92: Tabaco   |        |        |        |        |        |
| SBS93: Origen desconocido   |        | 38,81% |        |        | 36,76% |
| SBS94: Origen desconocido   |        |        |        |        |        |
| Posible artefacto de secuenciación  |        | 13,93% |        |        | 21,32% |
| Tratamiento quimioterápico  |        |        |        |        |        |

Continuación tabla suplementaria 6.

|   | COAD-<br>1 | COAD-<br>2 | COAD-<br>3 | COAD-<br>4 | COAD-<br>5 |
|---|------------|------------|------------|------------|------------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 27,74%     | 21,53%     | 21,98%     | 25,79%     | 8,57%      |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |            |            |            |            |            |
| SBS5: Origen desconocido (firma plana)  | 42,34%     | 78,47%     |            | 32,63%     | 45,14%     |
| SBS6: Sistema de reparación MMR defectuoso  |            |            |            |            |            |
| SBS7a: Exposición a luz ultravioleta  |            |            | 9,89%      |            |            |
| SBS7b: Exposición a luz ultravioleta  |            |            |            |            |            |
| SBS7d: Exposición a luz ultravioleta  |            |            | 23,08%     |            |            |
| SBS8: Origen desconocido  |            |            |            |            |            |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |            |            |            |            |            |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |            |            |            |            |            |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |            |            |            |            |            |
| SBS10d: Defectos en la lectura de prueba de POLD1   |            |            |            |            |            |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |            |            |            | 8,95%      |            |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |            |            | 21,98%     |            |            |
| SBS15: Sistema de reparación MMR defectuoso   |            |            |            |            | 6,86%      |
| SBS16: Origen desconocido   |            |            |            |            |            |
| SBS17b: Origen desconocido  |            |            |            |            |            |
| SBS18: Daño por especies reactivas del oxígeno  |            |            |            |            |            |
| SBS26: Sistema de reparación MMR defectuoso   |            |            |            |            |            |
| SBS28: Origen desconocido   |            |            |            |            |            |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |            |            |            |            | 10,86%     |
| SBS37: Origen desconocido   |            |            |            |            |            |
| SBS40: Origen desconocido (firma plana)   |            |            |            |            |            |
| SBS41: Origen desconocido   |            |            |            |            |            |
| SBS42: Exposición a haloalcanos   |            |            |            |            |            |
| SBS44: Sistema de reparación MMR defectuoso   |            |            |            |            |            |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |            |            |            | 15,79%     |            |
| SBS89: Origen desconocido   |            |            |            |            |            |
| SBS92: Tabaco   |            |            |            |            |            |
| SBS93: Origen desconocido   |            |            |            |            | 28,57%     |
| SBS94: Origen desconocido   |            |            |            |            |            |
| Posible artefacto de secuenciación  | 29,93%     |            | 23,08%     |            |            |
| Tratamiento quimioterápico  |            |            |            | 16,84%     |            |

Continuación tabla suplementaria 6.

|   | COAD-6 | COAD-7 | COAD-8 | COAD-9 | COAD-10 |
|---|--------|--------|--------|--------|---------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 29,00% | 34,18% | 29,69% | 28,63% | 18,48%  |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        |        |        |        |         |
| SBS5: Origen desconocido (firma plana)  | 47,50% |        | 47,66% | 25,88% | 19,43%  |
| SBS6: Sistema de reparación MMR defectuoso  |        |        |        |        |         |
| SBS7a: Exposición a luz ultravioleta  |        |        |        |        |         |
| SBS7b: Exposición a luz ultravioleta  |        |        |        |        |         |
| SBS7d: Exposición a luz ultravioleta  |        |        |        |        |         |
| SBS8: Origen desconocido  |        |        | 22,66% |        |         |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |        |        |        |         |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |         |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        |         |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |        |        |        |         |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        |        |        |        |         |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |        |        |        |         |
| SBS15: Sistema de reparación MMR defectuoso   |        |        |        |        |         |
| SBS16: Origen desconocido   |        |        |        |        |         |
| SBS17b: Origen desconocido  |        |        |        |        |         |
| SBS18: Daño por especies reactivas del oxígeno  |        |        |        | 34,51% | 20,85%  |
| SBS26: Sistema de reparación MMR defectuoso   |        |        |        |        |         |
| SBS28: Origen desconocido   |        |        |        |        |         |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        |        |        | 10,98% | 11,37%  |
| SBS37: Origen desconocido   | 23,50% |        |        |        |         |
| SBS40: Origen desconocido (firma plana)   |        | 65,82% |        |        |         |
| SBS41: Origen desconocido   |        |        |        |        |         |
| SBS42: Exposición a haloalcanos   |        |        |        |        |         |
| SBS44: Sistema de reparación MMR defectuoso   |        |        |        |        |         |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |        |        |        | 14,69%  |
| SBS89: Origen desconocido   |        |        |        |        |         |
| SBS92: Tabaco   |        |        |        |        |         |
| SBS93: Origen desconocido   |        |        |        |        |         |
| SBS94: Origen desconocido   |        |        |        |        |         |
| Posible artefacto de secuenciación  |        |        |        |        |         |
| Tratamiento quimioterápico  |        |        |        |        | 15,17%  |

Continuación tabla suplementaria 6.

|   | COAD-11 | COAD-12 | COAD-13 | COAD-14 | COAD-15 |
|---|---------|---------|---------|---------|---------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 70,55%  | 18,89%  | 21,38%  | 37,62%  | 19,01%  |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |         |         |         |         |         |
| SBS5: Origen desconocido (firma plana)  | 29,45%  | 43,89%  | 24,14%  | 29,07%  | 41,44%  |
| SBS6: Sistema de reparación MMR defectuoso  |         |         |         |         |         |
| SBS7a: Exposición a luz ultravioleta  |         |         |         |         |         |
| SBS7b: Exposición a luz ultravioleta  |         |         |         |         |         |
| SBS7d: Exposición a luz ultravioleta  |         |         |         |         |         |
| SBS8: Origen desconocido  |         |         |         |         |         |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |         |         |         |         | 28,52%  |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |         |         | 4,14%   |         |         |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |         |         | 8,97%   |         |         |
| SBS10d: Defectos en la lectura de prueba de POLD1   |         | 10,56%  |         |         |         |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |         |         |         |         |         |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |         |         |         |         |         |
| SBS15: Sistema de reparación MMR defectuoso   |         |         | 17,93%  |         |         |
| SBS16: Origen desconocido   |         |         |         |         |         |
| SBS17b: Origen desconocido  |         |         |         |         |         |
| SBS18: Daño por especies reactivas del oxígeno  |         |         | 23,45%  |         |         |
| SBS26: Sistema de reparación MMR defectuoso   |         |         |         |         |         |
| SBS28: Origen desconocido   |         |         |         |         |         |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |         |         |         |         |         |
| SBS37: Origen desconocido   |         |         |         |         |         |
| SBS40: Origen desconocido (firma plana)   |         |         |         |         |         |
| SBS41: Origen desconocido   |         |         |         |         |         |
| SBS42: Exposición a haloalcanos   |         |         |         |         |         |
| SBS44: Sistema de reparación MMR defectuoso   |         |         |         | 33,31%  |         |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |         | 26,67%  |         |         |         |
| SBS89: Origen desconocido   |         |         |         |         |         |
| SBS92: Tabaco   |         |         |         |         |         |
| SBS93: Origen desconocido   |         |         |         |         |         |
| SBS94: Origen desconocido   |         |         |         |         |         |
| Posible artefacto de secuenciación  |         |         |         |         |         |
| Tratamiento quimioterápico  |         |         |         |         | 11,03%  |

Continuación tabla suplementaria 6.

|   | COAD-16 | COAD-17 | COAD-18 | READ-1 | READ-2 |
|---|---------|---------|---------|--------|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 18,68%  | 0,38%   | 11,12%  | 27,11% | 34,05% |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |         | 3,41%   |         |        |        |
| SBS5: Origen desconocido (firma plana)  | 17,20%  | 17,42%  | 31,58%  | 56,44% | 28,45% |
| SBS6: Sistema de reparación MMR defectuoso  |         | 17,05%  |         |        |        |
| SBS7a: Exposición a luz ultravioleta  |         |         |         |        |        |
| SBS7b: Exposición a luz ultravioleta  |         |         |         |        |        |
| SBS7d: Exposición a luz ultravioleta  |         |         |         |        |        |
| SBS8: Origen desconocido  |         |         |         |        |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |         |         |         |        |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |         |         |         |        |        |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |         |         |         |        |        |
| SBS10d: Defectos en la lectura de prueba de POLD1   |         |         |         |        |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |         | 7,58%   |         |        |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |         |         |         |        |        |
| SBS15: Sistema de reparación MMR defectuoso   |         |         | 42,59%  |        |        |
| SBS16: Origen desconocido   |         |         |         |        |        |
| SBS17b: Origen desconocido  |         |         |         | 16,44% |        |
| SBS18: Daño por especies reactivas del oxígeno  |         | 17,05%  |         |        | 18,53% |
| SBS26: Sistema de reparación MMR defectuoso   | 21,87%  |         |         |        |        |
| SBS28: Origen desconocido   |         |         |         |        |        |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |         | 7,58%   |         |        |        |
| SBS37: Origen desconocido   |         |         |         |        |        |
| SBS40: Origen desconocido (firma plana)   |         |         |         |        |        |
| SBS41: Origen desconocido   |         |         |         |        |        |
| SBS42: Exposición a haloalcanos   |         |         | 14,72%  |        |        |
| SBS44: Sistema de reparación MMR defectuoso   | 42,25%  |         |         |        |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |         |         |         |        |        |
| SBS89: Origen desconocido   |         |         |         |        |        |
| SBS92: Tabaco   |         |         |         |        |        |
| SBS93: Origen desconocido   |         | 29,55%  |         |        |        |
| SBS94: Origen desconocido   |         |         |         |        |        |
| Posible artefacto de secuenciación  |         |         |         |        |        |
| Tratamiento quimioterápico  |         |         |         |        | 18,97% |

Continuación tabla suplementaria 6.

|   | READ-3 | READ-4 | READ-5 | READ-6 | READ-7 |
|---|--------|--------|--------|--------|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 36,22% | 30,97% | 24,83% | 20,63% |        |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        |        |        |        |        |
| SBS5: Origen desconocido (firma plana)  | 40,16% | 23,87% | 58,39% | 32,29% | 8,46%  |
| SBS6: Sistema de reparación MMR defectuoso  |        |        |        |        |        |
| SBS7a: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS7b: Exposición a luz ultravioleta  |        |        |        | 8,52%  |        |
| SBS7d: Exposición a luz ultravioleta  |        |        |        |        |        |
| SBS8: Origen desconocido  |        |        |        |        |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |        |        |        |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |        |        |        | 32,05% |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        | 9,03%  |        |        | 41,67% |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |        |        |        |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        |        |        |        |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |        |        |        |        |
| SBS15: Sistema de reparación MMR defectuoso   |        |        |        |        | 7,58%  |
| SBS16: Origen desconocido   |        |        |        |        |        |
| SBS17b: Origen desconocido  |        |        |        |        |        |
| SBS18: Daño por especies reactivas del oxígeno  |        | 23,23% |        |        |        |
| SBS26: Sistema de reparación MMR defectuoso   |        |        |        |        |        |
| SBS28: Origen desconocido   |        |        |        |        | 10,25% |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        | 12,90% |        |        |        |
| SBS37: Origen desconocido   |        |        |        |        |        |
| SBS40: Origen desconocido (firma plana)   |        |        |        |        |        |
| SBS41: Origen desconocido   |        |        |        | 38,57% |        |
| SBS42: Exposición a haloalcanos   |        |        |        |        |        |
| SBS44: Sistema de reparación MMR defectuoso   |        |        | 16,78% |        |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |        |        |        |        |
| SBS89: Origen desconocido   |        |        |        |        |        |
| SBS92: Tabaco   |        |        |        |        |        |
| SBS93: Origen desconocido   |        |        |        |        |        |
| SBS94: Origen desconocido   |        |        |        |        |        |
| Posible artefacto de secuenciación  |        |        |        |        |        |
| Tratamiento quimioterápico  | 23,62% |        |        |        |        |

## Continuación tabla suplementaria 6.

|   | READ-8 |
|---|--------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 24,32% |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |        |
| SBS5: Origen desconocido (firma plana)  | 40,00% |
| SBS6: Sistema de reparación MMR defectuoso  |        |
| SBS7a: Exposición a luz ultravioleta  |        |
| SBS7b: Exposición a luz ultravioleta  |        |
| SBS7d: Exposición a luz ultravioleta  |        |
| SBS8: Origen desconocido  |        |
| SBS9: Actividad de hipermutación somática de la polimerasa $\eta$                                 |        |
| SBS10a: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |        |
| SBS10d: Defectos en la lectura de prueba de POLD1   |        |
| SBS13: Actividad de la familia APOBEC de citidina desaminasas                                     |        |
| SBS14: Mutación en la polimerasa $\epsilon$ y sistema de reparación MMR defectuoso                |        |
| SBS15: Sistema de reparación MMR defectuoso   |        |
| SBS16: Origen desconocido   |        |
| SBS17b: Origen desconocido  |        |
| SBS18: Daño por especies reactivas del oxígeno  | 35,68% |
| SBS26: Sistema de reparación MMR defectuoso   |        |
| SBS28: Origen desconocido   |        |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |        |
| SBS37: Origen desconocido   |        |
| SBS40: Origen desconocido (firma plana)   |        |
| SBS41: Origen desconocido   |        |
| SBS42: Exposición a haloalcanos   |        |
| SBS44: Sistema de reparación MMR defectuoso   |        |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |        |
| SBS89: Origen desconocido   |        |
| SBS92: Tabaco   |        |
| SBS93: Origen desconocido   |        |
| SBS94: Origen desconocido   |        |
| Posible artefacto de secuenciación  |        |
| Tratamiento quimioterápico  |        |



**Tabla suplementaria 7. Clasificación de los CCR de las cohortes TOGETHER y TCGA en los distintos CMS de acuerdo la herramienta CMSCaller v0.99.2<sup>107</sup>.** Se recogen los datos de clasificación de los distintos tumores obtenidos a partir de los datos de RNA-seq con el paquete de R *CMSCaller* v.0.99.2<sup>107</sup>. En la columna predicción se indica el resultado final de la clasificación, las columnas 3 a 6 indican la distancia de predicción a cada uno de los CMS previamente definidos por Guinney y colaboradores<sup>33</sup>, y se indica el p-valor y FDR de cada una de las predicciones. FDR: tasa de descubrimiento falso, NA: no clasificado en los CMS predefinidos.

| Paciente | Predicción | Distancia a CMS1 | Distancia a CMS2 | Distancia a CMS3 | Distancia a CMS4 | p-valor | FDR   |
|----------|------------|------------------|------------------|------------------|------------------|---------|-------|
| 102      | CMS4       | 0,723            | 0,667            | 0,704            | 0,620            | 0,001   | 0,001 |
| 103      | NA         | 0,721            | 0,687            | 0,761            | 0,788            | 0,326   | 0,333 |
| 104      | CMS3       | 0,776            | 0,679            | 0,631            | 0,823            | 0,001   | 0,001 |
| 105      | CMS1       | 0,651            | 0,729            | 0,714            | 0,801            | 0,001   | 0,001 |
| 107      | CMS3       | 0,723            | 0,681            | 0,604            | 0,832            | 0,001   | 0,001 |
| 108      | CMS3       | 0,738            | 0,726            | 0,642            | 0,856            | 0,001   | 0,001 |
| 109      | NA         | 0,735            | 0,677            | 0,735            | 0,800            | 0,105   | 0,112 |
| 202      | CMS1       | 0,637            | 0,736            | 0,647            | 0,822            | 0,001   | 0,001 |
| 209      | CMS4       | 0,726            | 0,706            | 0,699            | 0,398            | 0,001   | 0,001 |
| 302      | CMS4       | 0,668            | 0,712            | 0,678            | 0,548            | 0,001   | 0,001 |
| 303      | CMS2       | 0,677            | 0,640            | 0,658            | 0,708            | 0,001   | 0,001 |
| 304      | CMS4       | 0,561            | 0,769            | 0,677            | 0,490            | 0,001   | 0,001 |
| 305      | CMS4       | 0,739            | 0,665            | 0,713            | 0,535            | 0,001   | 0,001 |
| 308      | CMS4       | 0,693            | 0,607            | 0,632            | 0,583            | 0,001   | 0,001 |
| 310      | CMS2       | 0,765            | 0,662            | 0,814            | 0,690            | 0,005   | 0,006 |
| 401      | CMS1       | 0,585            | 0,821            | 0,656            | 0,733            | 0,001   | 0,001 |
| 402      | NA         | 0,743            | 0,698            | 0,725            | 0,775            | 0,253   | 0,261 |
| 501      | NA         | 0,801            | 0,674            | 0,791            | 0,865            | 0,109   | 0,115 |
| COAD-1   | CMS2       | 0,795            | 0,657            | 0,770            | 0,701            | 0,004   | 0,005 |
| COAD-2   | CMS2       | 0,786            | 0,594            | 0,775            | 0,652            | 0,001   | 0,001 |
| COAD-3   | CMS4       | 0,623            | 0,739            | 0,720            | 0,469            | 0,001   | 0,001 |
| COAD-5   | CMS4       | 0,788            | 0,651            | 0,703            | 0,612            | 0,001   | 0,001 |
| COAD-6   | CMS3       | 0,702            | 0,685            | 0,659            | 0,755            | 0,004   | 0,005 |
| COAD-7   | CMS4       | 0,652            | 0,662            | 0,735            | 0,586            | 0,001   | 0,001 |
| COAD-8   | CMS4       | 0,743            | 0,650            | 0,726            | 0,443            | 0,001   | 0,001 |
| COAD-9   | CMS4       | 0,666            | 0,720            | 0,759            | 0,574            | 0,001   | 0,001 |
| COAD-11  | CMS4       | 0,597            | 0,771            | 0,613            | 0,421            | 0,001   | 0,001 |

Continuación tabla suplementaria 7.

| Paciente | Predicción | Distancia a CMS1 | Distancia a CMS2 | Distancia a CMS3 | Distancia a CMS4 | p-valor | FDR   |
|----------|------------|------------------|------------------|------------------|------------------|---------|-------|
| COAD-11  | CMS4       | 0,597            | 0,771            | 0,613            | 0,421            | 0,001   | 0,001 |
| COAD-12  | CMS2       | 0,781            | 0,574            | 0,781            | 0,711            | 0,001   | 0,001 |
| COAD-13  | CMS2       | 0,837            | 0,622            | 0,781            | 0,771            | 0,001   | 0,001 |
| COAD-14  | CMS1       | 0,551            | 0,851            | 0,628            | 0,720            | 0,001   | 0,001 |
| COAD-16  | CMS1       | 0,551            | 0,794            | 0,682            | 0,699            | 0,001   | 0,001 |
| COAD-17  | CMS2       | 0,793            | 0,592            | 0,776            | 0,633            | 0,001   | 0,001 |
| COAD-18  | CMS3       | 0,624            | 0,823            | 0,556            | 0,834            | 0,001   | 0,001 |
| COAD-19  | CMS2       | 0,770            | 0,593            | 0,633            | 0,846            | 0,001   | 0,001 |
| COAD-20  | CMS3       | 0,766            | 0,640            | 0,612            | 0,743            | 0,001   | 0,001 |
| COAD-21  | CMS4       | 0,692            | 0,687            | 0,781            | 0,366            | 0,001   | 0,001 |
| COAD-22  | CMS1       | 0,519            | 0,825            | 0,665            | 0,751            | 0,001   | 0,001 |
| COAD-23  | CMS2       | 0,780            | 0,626            | 0,811            | 0,703            | 0,001   | 0,001 |
| COAD-24  | CMS4       | 0,810            | 0,640            | 0,800            | 0,501            | 0,001   | 0,001 |
| COAD-25  | CMS4       | 0,600            | 0,761            | 0,726            | 0,410            | 0,001   | 0,001 |
| COAD-26  | CMS2       | 0,816            | 0,651            | 0,743            | 0,883            | 0,003   | 0,004 |
| COAD-27  | NA         | 0,705            | 0,664            | 0,732            | 0,868            | 0,067   | 0,073 |
| COAD-28  | CMS4       | 0,686            | 0,649            | 0,744            | 0,418            | 0,001   | 0,001 |
| COAD-29  | CMS4       | 0,626            | 0,827            | 0,690            | 0,606            | 0,001   | 0,001 |
| COAD-30  | CMS4       | 0,631            | 0,731            | 0,745            | 0,371            | 0,001   | 0,001 |
| COAD-31  | CMS4       | 0,738            | 0,642            | 0,697            | 0,428            | 0,001   | 0,001 |
| COAD-32  | CMS2       | 0,794            | 0,660            | 0,675            | 0,854            | 0,018   | 0,020 |
| COAD-33  | CMS4       | 0,645            | 0,688            | 0,737            | 0,421            | 0,001   | 0,001 |
| COAD-34  | CMS1       | 0,529            | 0,852            | 0,680            | 0,807            | 0,001   | 0,001 |
| COAD-35  | NA         | 0,696            | 0,702            | 0,794            | 0,692            | 0,153   | 0,160 |
| COAD-36  | CMS2       | 0,751            | 0,633            | 0,703            | 0,799            | 0,002   | 0,003 |
| COAD-37  | CMS2       | 0,802            | 0,651            | 0,721            | 0,882            | 0,002   | 0,003 |
| COAD-38  | CMS4       | 0,697            | 0,690            | 0,705            | 0,475            | 0,001   | 0,001 |
| COAD-39  | CMS2       | 0,804            | 0,594            | 0,724            | 0,866            | 0,001   | 0,001 |
| COAD-40  | CMS3       | 0,726            | 0,682            | 0,588            | 0,794            | 0,001   | 0,001 |
| COAD-41  | CMS2       | 0,812            | 0,652            | 0,766            | 0,880            | 0,001   | 0,001 |
| COAD-42  | CMS2       | 0,781            | 0,661            | 0,729            | 0,707            | 0,011   | 0,013 |
| COAD-43  | CMS1       | 0,547            | 0,856            | 0,670            | 0,805            | 0,001   | 0,001 |
| COAD-44  | CMS3       | 0,712            | 0,735            | 0,618            | 0,917            | 0,001   | 0,001 |

Continuación tabla suplementaria 7.

| Paciente | Predicción | Distancia a CMS1 | Distancia a CMS2 | Distancia a CMS3 | Distancia a CMS4 | p-valor | FDR   |
|----------|------------|------------------|------------------|------------------|------------------|---------|-------|
| COAD-45  | CMS4       | 0,604            | 0,733            | 0,661            | 0,410            | 0,001   | 0,001 |
| COAD-46  | CMS2       | 0,797            | 0,630            | 0,690            | 0,880            | 0,001   | 0,001 |
| COAD-47  | CMS4       | 0,763            | 0,650            | 0,765            | 0,590            | 0,001   | 0,001 |
| COAD-48  | CMS3       | 0,664            | 0,829            | 0,631            | 0,879            | 0,001   | 0,001 |
| COAD-49  | CMS4       | 0,662            | 0,718            | 0,724            | 0,402            | 0,001   | 0,001 |
| COAD-50  | CMS1       | 0,574            | 0,755            | 0,638            | 0,701            | 0,001   | 0,001 |
| COAD-51  | CMS1       | 0,634            | 0,763            | 0,684            | 0,861            | 0,001   | 0,001 |
| COAD-52  | CMS4       | 0,732            | 0,637            | 0,669            | 0,483            | 0,001   | 0,001 |
| COAD-53  | NA         | 0,676            | 0,785            | 0,720            | 0,842            | 0,067   | 0,073 |
| COAD-54  | CMS1       | 0,668            | 0,701            | 0,701            | 0,669            | 0,006   | 0,007 |
| COAD-55  | CMS2       | 0,782            | 0,605            | 0,754            | 0,658            | 0,001   | 0,001 |
| COAD-56  | CMS4       | 0,737            | 0,632            | 0,781            | 0,448            | 0,001   | 0,001 |
| COAD-57  | CMS3       | 0,718            | 0,749            | 0,540            | 0,828            | 0,001   | 0,001 |
| COAD-58  | CMS3       | 0,703            | 0,761            | 0,508            | 0,669            | 0,001   | 0,001 |
| READ-1   | CMS1       | 0,667            | 0,759            | 0,764            | 0,700            | 0,036   | 0,040 |
| READ-4   | NA         | 0,835            | 0,699            | 0,753            | 0,849            | 0,713   | 0,721 |
| READ-5   | CMS2       | 0,794            | 0,602            | 0,729            | 0,756            | 0,001   | 0,001 |
| READ-6   | CMS3       | 0,804            | 0,671            | 0,660            | 0,826            | 0,006   | 0,007 |
| READ-7   | CMS1       | 0,586            | 0,803            | 0,773            | 0,622            | 0,001   | 0,001 |
| READ-8   | CMS4       | 0,658            | 0,786            | 0,649            | 0,397            | 0,001   | 0,001 |
| READ-9   | CMS2       | 0,729            | 0,629            | 0,796            | 0,782            | 0,001   | 0,001 |
| READ-10  | CMS2       | 0,795            | 0,646            | 0,719            | 0,857            | 0,001   | 0,001 |
| READ-11  | CMS1       | 0,649            | 0,765            | 0,766            | 0,798            | 0,001   | 0,001 |
| READ-12  | CMS4       | 0,629            | 0,684            | 0,765            | 0,442            | 0,001   | 0,001 |
| READ-13  | CMS2       | 0,806            | 0,625            | 0,762            | 0,712            | 0,001   | 0,001 |
| READ-14  | CMS4       | 0,777            | 0,651            | 0,774            | 0,493            | 0,001   | 0,001 |
| READ-15  | CMS1       | 0,586            | 0,835            | 0,700            | 0,643            | 0,001   | 0,001 |
| READ-16  | NA         | 0,774            | 0,700            | 0,722            | 0,846            | 0,732   | 0,732 |
| READ-17  | CMS2       | 0,755            | 0,629            | 0,719            | 0,854            | 0,001   | 0,001 |
| READ-18  | NA         | 0,670            | 0,692            | 0,687            | 0,862            | 0,079   | 0,085 |
| READ-19  | CMS2       | 0,808            | 0,579            | 0,729            | 0,691            | 0,001   | 0,001 |
| READ-20  | CMS3       | 0,633            | 0,803            | 0,567            | 0,859            | 0,001   | 0,001 |
| READ-21  | CMS4       | 0,726            | 0,668            | 0,769            | 0,546            | 0,001   | 0,001 |

Continuación tabla suplementaria 7.

| Paciente | Predicción | Distancia a CMS1 | Distancia a CMS2 | Distancia a CMS3 | Distancia a CMS4 | p-valor | FDR   |
|----------|------------|------------------|------------------|------------------|------------------|---------|-------|
| READ-22  | CMS2       | 0,755            | 0,644            | 0,658            | 0,833            | 0,002   | 0,003 |
| READ-23  | CMS2       | 0,742            | 0,627            | 0,730            | 0,825            | 0,001   | 0,001 |

**Tabla suplementaria 8. Resultados del análisis GSA para el CMS1.** Se muestran los resultados (significativos y no significativos) para los distintos conjuntos de datos predefinidos por el paquete de R *CMSCaller* v0.99.2<sup>107</sup>. Up: conjunto de genes con expresión aumentada, Down: conjunto de genes con expresión disminuida, FDR: tasa de descubrimiento falso, MSI: inestabilidad de microsatélites, MSS: estabilidad de microsatélites, EMT: transición epitelio-mesenquimal.

|                    | Total de genes | Dirección | p-valor                | FDR                    |
|--------------------|----------------|-----------|------------------------|------------------------|
| MSI                | 27             | Up        | $2,17 \times 10^{-7}$  | $1,52 \times 10^{-6}$  |
| Reparación ADN     | 149            | Up        | 0,017                  | 0,030                  |
| HNF4A              | 57             | Down      | 0,004                  | 0,008                  |
| MSS                | 65             | Down      | $5,31 \times 10^{-19}$ | $7,44 \times 10^{-18}$ |
| MYC                | 58             | Up        | $4,51 \times 10^{-5}$  | $1,58 \times 10^{-4}$  |
| WNT                | 12             | Down      | 0,641                  | 0,641                  |
| Ciclo celular      | 200            | Up        | $1,76 \times 10^{-5}$  | $8,21 \times 10^{-5}$  |
| Diferenciación     | 440            | Down      | 0,041                  | 0,057                  |
| Glicólisis         | 192            | Up        | 0,001                  | 0,003                  |
| Ácidos grasos      | 148            | Up        | $7,45 \times 10^{-4}$  | 0,002                  |
| CDX2               | 30             | Down      | 0,470                  | 0,506                  |
| Células madre LGR5 | 59             | Down      | 0,027                  | 0,043                  |
| TGF- $\beta$       | 58             | Down      | 0,322                  | 0,375                  |
| EMT                | 197            | Down      | 0,245                  | 0,312                  |

**Tabla suplementaria 9. Resultados del análisis GSA para el CMS2.** Se muestran los resultados (significativos y no significativos) para los distintos conjuntos de datos predefinidos por el paquete de R *CMSCaller* v0.99.2<sup>107</sup>. Up: conjunto de genes con expresión aumentada, Down: conjunto de genes con expresión disminuida, FDR: tasa de descubrimiento falso, MSI: inestabilidad de microsatélites, MSS: estabilidad de microsatélites, EMT: transición epitelio-mesenquimal.

|                    | Total de genes | Dirección | p-valor                | FDR                    |
|--------------------|----------------|-----------|------------------------|------------------------|
| MSI                | 27             | Down      | $3,27 \times 10^{-8}$  | $2,60 \times 10^{-7}$  |
| Reparación ADN     | 149            | Up        | 0,642                  | 0,686                  |
| HNF4A              | 57             | Up        | 0,236                  | 0,300                  |
| MSS                | 65             | Up        | $2,73 \times 10^{-12}$ | $3,82 \times 10^{-11}$ |
| MYC                | 58             | Up        | 0,471                  | 0,549                  |
| WNT                | 12             | Up        | 0,686                  | 0,686                  |
| Ciclo celular      | 200            | Down      | 0,098                  | 0,172                  |
| Diferenciación     | 440            | Down      | 0,041                  | 0,081                  |
| Glicólisis         | 192            | Down      | 0,139                  | 0,216                  |
| Ácidos grasos      | 148            | Down      | 0,007                  | 0,019                  |
| CDX2               | 30             | Down      | 0,007                  | 0,019                  |
| Células madre LGR5 | 59             | Up        | 0,156                  | 0,218                  |
| TGF- $\beta$       | 58             | Down      | 0,008                  | 0,019                  |
| EMT                | 197            | Down      | $7,75 \times 10^{-4}$  | 0,004                  |

**Tabla suplementaria 10. Resultados del análisis GSA para el CMS3.** Se muestran los resultados (significativos y no significativos) para los distintos conjuntos de datos predefinidos por el paquete de R *CMSCaller* v0.99.2<sup>107</sup>. Up: conjunto de genes con expresión aumentada, Down: conjunto de genes con expresión disminuida, FDR: tasa de descubrimiento falso, MSI: inestabilidad de microsatélites, MSS: estabilidad de microsatélites, EMT: transición epitelio-mesenquimal.

|                | Total de genes | Dirección | p-valor               | FDR                   |
|----------------|----------------|-----------|-----------------------|-----------------------|
| MSI            | 27             | Up        | 0,089                 | 0,154                 |
| Reparación ADN | 149            | Down      | 0,750                 | 0,808                 |
| HNF4A          | 57             | Up        | 0,033                 | 0,066                 |
| MSS            | 65             | Up        | 0,243                 | 0,340                 |
| MYC            | 58             | Up        | 0,972                 | 0,972                 |
| WNT            | 12             | Down      | 0,309                 | 0,393                 |
| Ciclo celular  | 200            | Up        | $9,83 \times 10^{-7}$ | $3,44 \times 10^{-6}$ |
| Diferenciación | 440            | Up        | $1,55 \times 10^{-9}$ | $1,09 \times 10^{-8}$ |
| Glicólisis     | 192            | Up        | 0,668                 | 0,779                 |
| Ácidos grasos  | 148            | Up        | 0,015                 | 0,036                 |

Continuación tabla suplementaria 10.

|                    | Total de genes | Dirección | p-valor                | FDR                    |
|--------------------|----------------|-----------|------------------------|------------------------|
| CDX2               | 30             | Up        | 0,005                  | 0,013                  |
| Células madre LGR5 | 59             | Down      | 0,099                  | 0,154                  |
| TGF- $\beta$       | 58             | Down      | $5,34 \times 10^{-9}$  | $2,49 \times 10^{-8}$  |
| EMT                | 197            | Down      | $1,97 \times 10^{-20}$ | $3,76 \times 10^{-19}$ |

**Tabla suplementaria 11. Resultados del análisis GSA para el CMS4.** Se muestran los resultados (significativos y no significativos) para los distintos conjuntos de datos predefinidos por el paquete de R *CMSCaller* v0.99.2<sup>107</sup>. Up: conjunto de genes con expresión aumentada, Down: conjunto de genes con expresión disminuida, FDR: tasa de descubrimiento falso, MSI: inestabilidad de microsatélites, MSS: estabilidad de microsatélites, EMT: transición epitelio-mesenquimal.

|                    | Total de genes | Dirección | p-valor                | FDR                    |
|--------------------|----------------|-----------|------------------------|------------------------|
| MSI                | 27             | Down      | 0,633                  | 0,682                  |
| Reparación ADN     | 149            | Down      | 0,026                  | 0,061                  |
| HNF4A              | 57             | Down      | 0,460                  | 0,583                  |
| MSS                | 65             | Down      | 0,981                  | 0,981                  |
| MYC                | 58             | Down      | $1,40 \times 10^{-5}$  | $4,92 \times 10^{-5}$  |
| WNT                | 12             | Up        | 0,310                  | 0,434                  |
| Ciclo celular      | 200            | Down      | $5,75 \times 10^{-12}$ | $2,68 \times 10^{-11}$ |
| Diferenciación     | 440            | Down      | 0,075                  | 0,125                  |
| Glicólisis         | 192            | Down      | 0,080                  | 0,125                  |
| Ácidos grasos      | 148            | Down      | 0,008                  | 0,022                  |
| CDX2               | 30             | Up        | 0,500                  | 0,583                  |
| Células madre LGR5 | 59             | Up        | 0,034                  | 0,068                  |
| TGF- $\beta$       | 58             | Up        | $2,05 \times 10^{-18}$ | $1,43 \times 10^{-17}$ |
| EMT                | 197            | Up        | $2,15 \times 10^{-37}$ | $3,01 \times 10^{-36}$ |

**Tabla suplementaria 12. Rutas KEGG significativas (p-valor ajustado <0,05) en la cohorte de tumores CMS1.** Listado de rutas de señalización KEGG resultantes del análisis GSA con g:Profiler<sup>116</sup> realizado sobre posiciones diferencialmente metiladas con valores de  $|\log \text{fold change}| \geq 0,10$ . En el ajuste del p-valor se empleó la corrección por el método de Benjamini-Hochberg.

| Ruta KEGG | Nombre de la ruta KEGG                                   | p-valor ajustado       |
|-----------|--|------------------------|
| hsa04080  | Neuroactive ligand-receptor interaction                  | 7,62x10 <sup>-11</sup> |
| hsa04020  | Calcium signaling pathway                                | 2,42x10 <sup>-9</sup>  |
| hsa04724  | Glutamatergic synapse                                    | 3,97x10 <sup>-9</sup>  |
| hsa05032  | Morphine addiction                                       | 8,74x10 <sup>-9</sup>  |
| hsa05200  | Pathways in cancer                                       | 9,19x10 <sup>-9</sup>  |
| hsa04015  | Rap1 signaling pathway                                   | 1,21x10 <sup>-7</sup>  |
| hsa04360  | Axon guidance  | 1,65x10 <sup>-7</sup>  |
| hsa04014  | Ras signaling pathway                                    | 4,16x10 <sup>-7</sup>  |
| hsa04725  | Cholinergic synapse                                      | 9,35x10 <sup>-7</sup>  |
| hsa04072  | Phospholipase D signaling pathway                        | 8,24x10 <sup>-6</sup>  |
| hsa04024  | cAMP signaling pathway                                   | 1,65x10 <sup>-5</sup>  |
| hsa04727  | GABAergic synapse  | 1,91x10 <sup>-5</sup>  |
| hsa04151  | PI3K-Akt signaling pathway                               | 3,94x10 <sup>-5</sup>  |
| hsa04261  | Adrenergic signaling in cardiomyocytes                   | 4,10x10 <sup>-5</sup>  |
| hsa04713  | Circadian entrainment                                    | 8,84x10 <sup>-5</sup>  |
| hsa04726  | Serotonergic synapse                                     | 1,09x10 <sup>-4</sup>  |
| hsa04728  | Dopaminergic synapse                                     | 1,20x10 <sup>-4</sup>  |
| hsa04390  | Hippo signaling pathway                                  | 2,47x10 <sup>-4</sup>  |
| hsa05224  | Breast cancer  | 2,67x10 <sup>-4</sup>  |
| hsa05414  | Dilated cardiomyopathy                                   | 2,91x10 <sup>-4</sup>  |
| hsa04510  | Focal adhesion   | 2,97x10 <sup>-4</sup>  |
| hsa04010  | MAPK signaling pathway                                   | 3,12x10 <sup>-4</sup>  |
| hsa04810  | Regulation of actin cytoskeleton                         | 3,84x10 <sup>-4</sup>  |
| hsa05412  | Arrhythmogenic right ventricular cardiomyopathy          | 4,66x10 <sup>-4</sup>  |
| hsa04514  | Cell adhesion molecules                                  | 7,65x10 <sup>-4</sup>  |
| hsa04928  | Parathyroid hormone synthesis, secretion and action      | 8,24x10 <sup>-4</sup>  |
| hsa04911  | Insulin secretion  | 1,04x10 <sup>-3</sup>  |
| hsa05030  | Cocaine addiction  | 1,19x10 <sup>-3</sup>  |
| hsa04022  | cGMP-PKG signaling pathway                               | 1,57x10 <sup>-3</sup>  |
| hsa05226  | Gastric cancer   | 1,73x10 <sup>-3</sup>  |
| hsa04971  | Gastric acid secretion                                   | 2,45x10 <sup>-3</sup>  |
| hsa04062  | Chemokine signaling pathway                              | 5,94x10 <sup>-3</sup>  |
| hsa04730  | Long-term depression                                     | 6,10x10 <sup>-3</sup>  |
| hsa04520  | Adherens junction  | 6,24x10 <sup>-3</sup>  |
| hsa04611  | Platelet activation                                      | 6,57x10 <sup>-3</sup>  |
| hsa04916  | Melanogenesis  | 6,77x10 <sup>-3</sup>  |
| hsa04934  | Cushing syndrome   | 6,78x10 <sup>-3</sup>  |
| hsa04550  | Signaling pathways regulating pluripotency of stem cells | 7,92x10 <sup>-3</sup>  |
| hsa05165  | Human papillomavirus infection                           | 8,45x10 <sup>-3</sup>  |
| hsa04921  | Oxytocin signaling pathway                               | 8,77x10 <sup>-3</sup>  |
| hsa04935  | Growth hormone synthesis, secretion and action           | 9,32x10 <sup>-3</sup>  |
| hsa04926  | Relaxin signaling pathway                                | 1,29x10 <sup>-2</sup>  |
| hsa04512  | ECM-receptor interaction                                 | 1,40x10 <sup>-2</sup>  |
| hsa05033  | Nicotine addiction                                       | 1,51x10 <sup>-2</sup>  |
| hsa04310  | Wnt signaling pathway                                    | 1,63x10 <sup>-2</sup>  |
| hsa05410  | Hypertrophic cardiomyopathy                              | 2,72x10 <sup>-2</sup>  |
| hsa04340  | Hedgehog signaling pathway                               | 3,80x10 <sup>-2</sup>  |

**Tabla suplementaria 13. Rutas KEGG significativas (p-valor ajustado <0,05) en la cohorte de tumores CMS2.** Listado de rutas de señalización KEGG resultantes del análisis GSA con g:Profiler<sup>116</sup> realizado sobre posiciones diferencialmente metiladas con valores de  $|\log \text{fold change}| \geq 0,10$ . En el ajuste del p-valor se empleó la corrección por el método de Benjamini-Hochberg.

| Ruta KEGG | Nombre de la ruta KEGG                 | p-valor ajustado      |
|-----------|--|-----------------------|
| hsa04360  | Axon guidance                          | $1,42 \times 10^{-3}$ |
| hsa04261  | Adrenergic signaling in cardiomyocytes | $1,53 \times 10^{-3}$ |
| hsa04925  | Aldosterone synthesis and secretion    | $4,09 \times 10^{-3}$ |
| hsa04972  | Pancreatic secretion                   | $5,75 \times 10^{-3}$ |
| hsa04930  | Type II diabetes mellitus              | $8,55 \times 10^{-3}$ |
| hsa04010  | MAPK signaling pathway                 | $1,62 \times 10^{-2}$ |
| hsa04974  | Protein digestion and absorption       | $2,78 \times 10^{-2}$ |
| hsa04072  | Phospholipase D signaling pathway      | $3,22 \times 10^{-2}$ |

**Tabla suplementaria 14. Rutas KEGG significativas (p-valor ajustado <0,05) en la cohorte de tumores CMS3.** Listado de rutas de señalización KEGG resultantes del análisis GSA con g:Profiler<sup>116</sup> realizado sobre posiciones diferencialmente metiladas con valores de  $|\log \text{fold change}| \geq 0,10$ . En el ajuste del p-valor se empleó la corrección por el método de Benjamini-Hochberg.

| Ruta KEGG | Nombre de la ruta KEGG                              | p-valor ajustado      |
|-----------|---|-----------------------|
| hsa04724  | Glutamatergic synapse                               | $3,79 \times 10^{-6}$ |
| hsa04020  | Calcium signaling pathway                           | $3,90 \times 10^{-6}$ |
| hsa04024  | cAMP signaling pathway                              | $1,58 \times 10^{-5}$ |
| hsa04360  | Axon guidance                                       | $3,94 \times 10^{-5}$ |
| hsa04010  | MAPK signaling pathway                              | $1,15 \times 10^{-4}$ |
| hsa05032  | Morphine addiction                                  | $1,25 \times 10^{-4}$ |
| hsa04080  | Neuroactive ligand-receptor interaction             | $2,33 \times 10^{-4}$ |
| hsa05412  | Arrhythmogenic right ventricular cardiomyopathy     | $4,56 \times 10^{-4}$ |
| hsa04510  | Focal adhesion                                      | $1,57 \times 10^{-3}$ |
| hsa04911  | Insulin secretion                                   | $1,70 \times 10^{-3}$ |
| hsa05414  | Dilated cardiomyopathy                              | $1,84 \times 10^{-3}$ |
| hsa04721  | Synaptic vesicle cycle                              | $1,95 \times 10^{-3}$ |
| hsa04928  | Parathyroid hormone synthesis, secretion and action | $6,29 \times 10^{-3}$ |
| hsa04713  | Circadian entrainment                               | $7,91 \times 10^{-3}$ |
| hsa04727  | GABAergic synapse                                   | $9,85 \times 10^{-3}$ |
| hsa04974  | Protein digestion and absorption                    | $1,07 \times 10^{-2}$ |
| hsa05410  | Hypertrophic cardiomyopathy                         | $1,23 \times 10^{-2}$ |
| hsa05033  | Nicotine addiction                                  | $1,52 \times 10^{-2}$ |
| hsa04512  | ECM-receptor interaction                            | $2,11 \times 10^{-2}$ |
| hsa04925  | Aldosterone synthesis and secretion                 | $2,48 \times 10^{-2}$ |
| hsa04934  | Cushing syndrome                                    | $3,41 \times 10^{-2}$ |
| hsa04929  | GnRH secretion                                      | $3,96 \times 10^{-2}$ |



**Tabla suplementaria 15. Rutas KEGG significativas (p-valor ajustado <0,05) en la cohorte de tumores CMS4.** Listado de rutas de señalización KEGG resultantes del análisis GSA con g:Profiler<sup>116</sup> realizado sobre posiciones diferencialmente metiladas con valores de  $|\log \text{fold change}| \geq 0,10$ . En el ajuste del p-valor se empleó la corrección por el método de Benjamini-Hochberg.

| Ruta KEGG | Nombre de la ruta KEGG                          | p-valor ajustado      |
|-----------|---|-----------------------|
| hsa04360  | Axon guidance                                   | $1,22 \times 10^{-8}$ |
| hsa04080  | Neuroactive ligand-receptor interaction         | $1,31 \times 10^{-7}$ |
| hsa04020  | Calcium signaling pathway                       | $3,49 \times 10^{-7}$ |
| hsa04512  | ECM-receptor interaction                        | $1,33 \times 10^{-6}$ |
| hsa04510  | Focal adhesion                                  | $1,53 \times 10^{-6}$ |
| hsa04724  | Glutamatergic synapse                           | $2,81 \times 10^{-6}$ |
| hsa05412  | Arrhythmogenic right ventricular cardiomyopathy | $1,31 \times 10^{-5}$ |
| hsa04014  | Ras signaling pathway                           | $2,36 \times 10^{-5}$ |
| hsa04151  | PI3K-Akt signaling pathway                      | $4,58 \times 10^{-5}$ |
| hsa04974  | Protein digestion and absorption                | $2,05 \times 10^{-4}$ |
| hsa04015  | Rap1 signaling pathway                          | $2,72 \times 10^{-4}$ |
| hsa05033  | Nicotine addiction                              | $3,74 \times 10^{-4}$ |
| hsa04072  | Phospholipase D signaling pathway               | $4,58 \times 10^{-4}$ |
| hsa05410  | Hypertrophic cardiomyopathy                     | $8,76 \times 10^{-4}$ |
| hsa04514  | Cell adhesion molecules                         | $9,88 \times 10^{-4}$ |
| hsa04024  | cAMP signaling pathway                          | $1,12 \times 10^{-3}$ |
| hsa05032  | Morphine addiction                              | $1,67 \times 10^{-3}$ |
| hsa05414  | Dilated cardiomyopathy                          | $1,79 \times 10^{-3}$ |
| hsa04010  | MAPK signaling pathway                          | $2,95 \times 10^{-3}$ |
| hsa04726  | Serotonergic synapse                            | $3,47 \times 10^{-3}$ |
| hsa04713  | Circadian entrainment                           | $8,20 \times 10^{-3}$ |
| hsa05200  | Pathways in cancer                              | $1,18 \times 10^{-2}$ |
| hsa03320  | PPAR signaling pathway                          | $1,35 \times 10^{-2}$ |
| hsa04261  | Adrenergic signaling in cardiomyocytes          | $2,24 \times 10^{-2}$ |
| hsa04921  | Oxytocin signaling pathway                      | $2,24 \times 10^{-2}$ |

**Tabla suplementaria 16. Variantes en genes driver en CCR de la paciente 102.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen          | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de <i>driver</i> | VAF    |
|--------------|------------------------|-----|-------------------|---------------------|---------------------|-----------------------|--------|
| <i>LRP1B</i> | NM_018557.3            | 2   | g.141208136C>T    | c.10057+1G>A        |                     | TSG                   | 14,40% |
| <i>APC</i>   | NM_000038.6            | 5   | g.112128209del    | c.712del            | p.(Gln238Lysfs*55)  | TSG                   | 34,80% |
| <i>HLA-B</i> | NM_005514.8            | 6   | g.31324545G>T     | c.263C>A            | p.(Thr88Lys)        | Ambiguo               | 8,30%  |
| <i>PTPRT</i> | NM_007050.6            | 20  | g.41100981G>A     | c.1375C>T           | p.(Arg459Trp)       | TSG                   | 13,70% |
| <i>P2RY8</i> | NM_178129.5            | X   | g.1584920C>T      | c.532G>A            | p.(Val178Ile)       | OG                    | 27,20% |
| <i>ATRX</i>  | NM_000489.6            | X   | g.76937683C>T     | c.3065G>A           | p.(Arg1022Gln)      | TSG                   | 19,20% |

**Tabla suplementaria 17. Variantes en genes driver en CCR del paciente 103.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de <i>driver</i> | VAF    |
|---------------|------------------------|-----|-------------------|---------------------|---------------------|-----------------------|--------|
| <i>HSPG2</i>  | NM_005529.7            | 1   | g.22211090C>T     | c.1585G>A           | p.(Val529Met)       | Ambiguo               | 20,40% |
| <i>AFF3</i>   | NM_001386135.1         | 2   | g.100210294G>T    | c.1829C>A           | p.(Ala610Asp)       | Ambiguo, fusión       | 19,40% |
| <i>APC</i>    | NM_000038.6            | 5   | g.112175354dup    | c.4063dup           | p.(Ser1355Phefs*20) | TSG                   | 65,00% |
| <i>NRG1</i>   | NM_013964.5            | 8   | g.32599561A>G     | c.668A>G            | p.(Glu223Gly)       | OG                    | 21,30% |
| <i>PREX2</i>  | NM_024870.4            | 8   | g.69058537G>A     | c.4181G>A           | p.(Arg1394Gln)      | OG                    | 18,70% |
| <i>TCF7L2</i> | NM_001367943.1         | 10  | g.114912093G>A    | c.1163G>A           | p.(Trp388*)         | Fusión, OG            | 25,80% |
| <i>KRAS</i>   | NM_004985.5            | 12  | g.25398284C>G     | c.35G>C             | p.(Gly12Ala)        | OG                    | 36,60% |

**Tabla suplementaria 18. Variantes en genes driver en CCR del paciente 104.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen          | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|--------------|-------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>LRP1B</i> | NM_018557.3             | 2   | g.141359173C>A    | c.6835G>T           | p.(Ala2279Ser)      | TSG            | 8,90%  |
| <i>APC</i>   | NM_000038.6             | 5   | g.112164616C>T    | c.1690C>T           | p.(Arg564*)         | TSG            | 13,10% |
| <i>PTPRK</i> | NM_002844.4             | 6   | g.128563698A>T    | c.545T>A            | p.(Ile182Asn)       | TSG            | 5,60%  |
| <i>ARID2</i> | NM_152641.4             | 12  | g.46246305G>T     | c.4399G>T           | p.(Val1467Leu)      | TSG            | 9,40%  |
| <i>TP53</i>  | NM_000546.6             | 17  | g.7577538C>T      | c.743G>A            | p.(Arg248Gln)       | TSG            | 19,20% |
| <i>PLCB4</i> | NM_001377142.1          | 20  | g.9438088dup      | c.3024dup           | p.(Glu1009Argfs*16) | OG             | 15,40% |
| <i>NEFH</i>  | NM_021076.4             | 22  | g.29876735G>A     | c.484G>A            | p.(Ala162Thr)       | Ambiguo        | 48,50% |
| <i>SMC1A</i> | NM_006306.4             | X   | g.53409206C>G     | c.3384G>C           | p.(Leu1128Phe)      | OG             | 25,60% |

**Tabla suplementaria 19. Variantes en genes driver en CCR del paciente 105.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen             | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37)        | Cambio nucleotídico | Cambio aminoacídico  | Tipo de driver | VAF    |
|-----------------|-------------------------|-----|--------------------------|---------------------|----------------------|----------------|--------|
| <i>PCBP1</i>    | NM_006196.4             | 2   | g.70315174T>A            | c.299T>A            | p.(Leu100Gln)        | OG             | 45,70% |
| <i>PIK3R1</i>   | NM_181523.3             | 5   | g.67591137_67591151del   | c.1730_1744del      | p.(Arg577_Leu581del) | TSG            | 64,40% |
| <i>APC</i>      | NM_000038.6             | 5   | g.112175684_112175685del | c.4393_4394del      | p.(Ser1465Trpfs*3)   | TSG            | 69,70% |
| <i>BCLAF1</i>   | NM_014739.3             | 6   | g.136597377C>T           | c.1286G>A           | p.(Arg429Gln)        | OG             | 9,00%  |
| <i>BRAF</i>     | NM_004333.6             | 7   | g.140453136A>T           | c.1799T>A           | p.(Val600Glu)        | Fusión,<br>OG  | 55,50% |
| <i>ARHGEF12</i> | NM_015313.3             | 11  | g.120355778A>C           | c.4627A>C           | p.(Lys1543Gln)       | OG             | 21,90% |
| <i>NBEA</i>     | NM_001385012.1          | 13  | g.35517078G>A            | c.121G>A            | p.(Gly41Arg)         | Ambiguo        | 28,20% |
| <i>TP53</i>     | NM_000546.6             | 17  | g.7577535C>G             | c.746G>C            | p.(Arg249Thr)        | TSG            | 53,40% |
| <i>MAFB</i>     | NM_005461.5             | 20  | g.39317043C>T            | c.448G>A            | p.(Val150Met)        | Driver         | 19,40% |

**Tabla suplementaria 20. Variantes en genes driver en CCR de la paciente 107.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)    | Cambio nucleotídico | Cambio aminoacídico | Tipo de <i>driver</i> | VAF    |
|---------------|------------------------|-----|----------------------|---------------------|---------------------|-----------------------|--------|
| <i>SPEN</i>   | NM_015001.3            | 1   | g.16242743G>A        | c.1364G>A           | p.(Arg455His)       | Ambiguo               | 30,40% |
| <i>PCBP1</i>  | NM_006196.4            | 2   | g.70315174T>A        | c.299T>A            | p.(Leu100Gln)       | OG                    | 26,10% |
| <i>MAP2</i>   | NM_001375505.1         | 2   | g.210557592G>T       | c.698G>T            | p.(Ser233Ile)       | Ambiguo               | 24,00% |
| <i>CTNNB1</i> | NM_001904.4            | 3   | g.41266104G>A        | c.101G>A            | p.(Gly34Glu)        | Fusión, OG            | 45,20% |
| <i>PIK3CA</i> | NM_006218.4            | 3   | g.178936091G>A       | c.1633G>A           | p.(Glu545Lys)       | OG                    | 12,30% |
| <i>KRAS</i>   | NM_004985.5            | 12  | g.25398284C>T        | c.35G>A             | p.(Gly12Asp)        | OG                    | 24,10% |
| <i>TBX3</i>   | NM_005996.4            | 12  | g.115112436G>C       | c.1244C>G           | p.(Ser415*)         | TSG                   | 16,00% |
| <i>TGIF1</i>  | NM_003244.4            | 18  | g.3456374_3456375dup | c.39_40dup          | p.(Glu14Valfs*86)   | TSG                   | 22,50% |

**Tabla suplementaria 21. Variantes en genes driver en CCR del paciente 108.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen         | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)      | Cambio nucleotídico | Cambio aminoacídico | Tipo de <i>driver</i> | VAF    |
|-------------|------------------------|-----|------------------------|---------------------|---------------------|-----------------------|--------|
| <i>APC</i>  | NM_000038.6            | 5   | g.112175239del         | c.3948del           | p.(Glu1317Lysfs*4)  | TSG                   | 52,00% |
| <i>PTEN</i> | NM_000314.8            | 10  | g.89624257_89624269del | c.31_43del          | p.(Arg11Aspfs*9)    | TSG                   | 55,80% |
| <i>KRAS</i> | NM_004985.5            | 12  | g.25398284C>T          | c.35G>A             | p.(Gly12Asp)        | OG                    | 27,20% |
| <i>KLF5</i> | NM_001730.5            | 13  | g.73636638C>A          | c.901C>A            | p.(Pro301Thr)       | Ambiguo               | 32,40% |
| <i>AR</i>   | NM_000044.6            | X   | g.66765170A>T          | c.182A>T            | p.(Gln61Leu)        | OG                    | 75,00% |

**Tabla suplementaria 22. Variantes en genes driver en CCR del paciente 109.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen          | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                    | Cambio nucleotídico      | Cambio aminoacídico                 | Tipo de driver | VAF              |
|--------------|------------------------|-----|--------------------------------------|--------------------------|-------------------------------------|----------------|------------------|
| <i>FBXW7</i> | NM_001349798.2         | 4   | g.153332466T>A<br>g.153332709T>A     | c.490A>T<br>c.247A>T     | p.(Lys164*)<br>p.(Arg83*)           | TSG            | 26,40%<br>27,00% |
| <i>APC</i>   | NM_000038.6            | 5   | g.112175174G>T                       | c.3883G>T                | p.(Glu1295*)                        | TSG            | 35,30%           |
| <i>TP53</i>  | NM_000546.6            | 17  | g.7577580_7577588dup<br>g.7578263G>A | c.695_703dup<br>c.586C>T | p.(Ile232_Tyr234dup)<br>p.(Arg196*) | TSG            | 36,70%<br>14,60% |
| <i>RHPN2</i> | NM_033103.5            | 19  | g.33490500C>T                        | c.1217G>A                | p.(Arg406Gln)                       | Ambiguo        | 18,30%           |

**Tabla suplementaria 23. Variantes en genes driver en CCR del paciente 209.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)            | Cambio nucleotídico  | Cambio aminoacídico              | Tipo de driver | VAF              |
|---------------|------------------------|-----|------------------------------|----------------------|----------------------------------|----------------|------------------|
| <i>SPEN</i>   | NM_015001.3            | 1   | g.16259812_16259813delinsTG  | c.7077_7078delinsTG  | p.(Asn2360Asp)                   | Ambiguo        | 24,70%           |
| <i>PIK3CA</i> | NM_006218.4            | 3   | g.178952085A>G               | c.3140A>G            | p.(His1047Arg)                   | OG             | 28,60%           |
| <i>DROSHA</i> | NM_001382508.1         | 5   | g.31409438_31409445del       | c.3668-6_3669del     | p.(Ser1223Phefs*8)               | TSG            | 11,40%           |
| <i>APC</i>    | NM_000038.6            | 5   | g.112175354del               | c.4063del            | p.(Ser1355Leufs*60)              | TSG            | 30,30%           |
| <i>WNK4</i>   | NM_032387.5            | 17  | g.40946645_40946651del       | c.2309_2315del       | p.(Pro770Leufs*163)              | Ambiguo        | 23,60%           |
| <i>SOX9</i>   | NM_000346.4            | 17  | g.70117882A>C                | c.350A>C             | p.(Gln117Pro)                    | TSG            | 53,60%           |
| <i>TGIF1</i>  | NM_003244.4            | 18  | g.3457438C>T<br>g.3457865del | c.319C>T<br>c.746del | p.(Gln107*)<br>p.(Gly249Aspfs*5) | TSG            | 29,60%<br>24,70% |
| <i>SETBP1</i> | NM_015559.3            | 18  | g.42532219C>T                | c.2914C>T            | p.(Arg972Trp)                    | Fusión,<br>OG  | 30,80%           |

**Tabla suplementaria 24. Variantes en genes driver en CCR del paciente 302.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen             | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)        | Cambio nucleotídico | Cambio aminoacídico | Tipo de <i>driver</i> | VAF    |
|-----------------|------------------------|-----|--------------------------|---------------------|---------------------|-----------------------|--------|
| <i>ACVR2A</i>   | NM_001616.5            | 2   | g.148674977_148674989del | c.798_810del        | p.(Ile266Metfs*41)  | TSG, fusión           | 6,20%  |
| <i>BARD1</i>    | NM_000465.4            | 2   | g.215609789A>G           | c.1903+2T>C         |                     | TSG, OG               | 7,60%  |
| <i>POLQ</i>     | NM_199420.4            | 3   | g.121206866G>C           | c.4912C>G           | p.(Leu1638Val)      | Ambiguo               | 6,60%  |
| <i>NSD2</i>     | NM_001042424.3         | 4   | g.1936914G>T             | c.1599G>T           | p.(Gln533His)       | No clasificado        | 6,80%  |
| <i>FAT4</i>     | NM_001291303.3         | 4   | g.126238099G>A           | c.533G>A            | p.(Arg178His)       | TSG                   | 7,60%  |
| <i>CTNND2</i>   | NM_001332.4            | 5   | g.11159691A>G            | c.2156T>C           | p.(Leu719Pro)       | <i>Driver</i>         | 6,30%  |
| <i>APC</i>      | NM_000038.6            | 5   | g.112175233del           | c.3942del           | p.(Arg1314Serfs*7)  | TSG                   | 10,40% |
| <i>ARHGEF10</i> | NM_014629.4            | 8   | g.1857559C>G             | c.2066C>G           | p.(Ala689Gly)       | Ambiguo               | 6,20%  |
| <i>TP53</i>     | NM_000546.6            | 17  | g.7578548_7578549insA    | c.381_382insT       | p.(Pro128Serfs*21)  | TSG                   | 7,80%  |

**Tabla suplementaria 25. Variantes en genes driver en CCR de la paciente 303.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)        | Cambio nucleotídico | Cambio aminoacídico  | Tipo de <i>driver</i> | VAF    |
|---------------|------------------------|-----|--------------------------|---------------------|----------------------|-----------------------|--------|
| <i>MTOR</i>   | NM_004958.4            | 1   | g.11319442C>T            | c.25G>A             | p.(Ala9Thr)          | OG                    | 14,00% |
| <i>ERCC3</i>  | NM_000122.2            | 2   | g.128046301T>C           | c.962A>G            | p.(Glu321Gly)        | TSG                   | 23,10% |
| <i>APC</i>    | NM_000038.6            | 5   | g.112137082del           | c.834+2del          |                      | TSG                   | 20,00% |
| <i>FAT2</i>   | NM_001447.3            | 5   | g.150906866C>T           | c.10237G>A          | p.(Ala3413Thr)       | OG                    | 12,30% |
| <i>BCL9L</i>  | NM_001378213.1         | 11  | g.118773201_118773206del | c.1248_1253del      | p.(Asp416_Ile417del) | OG, TSG               | 14,50% |
| <i>KDM5A</i>  | NM_001042603.3         | 12  | g.404959_404966del       | c.4235-7_4235del    | p.(Gly1412Valfs*38)  | OG                    | 6,30%  |
| <i>CHD4</i>   | NM_001273.5            | 12  | g.6701928G>A             | c.2708C>T           | p.(Thr903Ile)        | TSG, OG               | 5,20%  |
| <i>KRAS</i>   | NM_004985.5            | 12  | g.25398285C>T            | c.34G>A             | p.(Gly12Ser)         | OG                    | 5,70%  |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7577098T>G             | c.840A>C            | p.(Arg280Ser)        | TSG                   | 16,80% |
| <i>EFTUD2</i> | NM_004247.4            | 17  | g.42929182C>T            | c.2719G>A           | p.(Val907Met)        | Ambiguo               | 6,40%  |

**Tabla suplementaria 26. Variantes en genes driver en CCR del paciente 304.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                | Cambio nucleotídico      | Cambio aminoacídico | Tipo de <i>driver</i> | VAF            |
|---------------|------------------------|-----|----------------------------------|--------------------------|---------------------|-----------------------|----------------|
| <i>SF3B1</i>  | NM_012433.4            | 2   | g.198266716C>T                   | c.2216G>A                | p.(Arg739Lys)       | OG                    | 12,70%         |
| <i>APC</i>    | NM_000038.6            | 5   | g.112162804G>A<br>g.112175957dup | c.1409-1G>A<br>c.4666dup | p.(Thr1556Asnfs*3)  | TSG                   | 9,40%<br>8,70% |
| <i>BRAF</i>   | NM_004333.6            | 7   | g.140453136A>T                   | c.1799T>A                | p.(Val600Glu)       | Fusión, OG            | 16,50%         |
| <i>CDKN2A</i> | NM_000077.5            | 9   | g.21971186G>A                    | c.172C>T                 | p.(Arg58*)          | TSG                   | 7,70%          |
| <i>BCL11B</i> | NM_138576.4            | 14  | g.99642171C>A                    | c.1002G>T                | p.(Glu334Asp)       | Ambiguo, fusión       | 15,30%         |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7577548C>T                     | c.733G>A                 | p.(Gly245Ser)       | TSG                   | 22,00%         |

**Tabla suplementaria 27. Variantes en genes driver en CCR de la paciente 305.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de <i>driver</i> | VAF    |
|---------------|------------------------|-----|-------------------|---------------------|---------------------|-----------------------|--------|
| <i>ARID1A</i> | NM_006015.6            | 1   | g.27057761C>T     | c.1469C>T           | p.(Pro490Leu)       | TSG, fusión           | 11,10% |
| <i>CDH10</i>  | NM_006727.5            | 5   | g.24498554C>A     | c.1468G>T           | p.(Ala490Ser)       | OG                    | 7,50%  |
| <i>PIK3R1</i> | NM_181523.3            | 5   | g.67592086dup     | c.1902dup           | p.(Glu635*)         | TSG                   | 30,20% |
| <i>APC</i>    | NM_000038.6            | 5   | g.112128143C>T    | c.646C>T            | p.(Arg216*)         | TSG                   | 25,50% |
| <i>HGF</i>    | NM_000601.6            | 7   | g.81339515T>G     | c.1489A>C           | p.(Asn497His)       | OG                    | 12,50% |
| <i>CSMD3</i>  | NM_198123.2            | 8   | g.113303750G>T    | c.8963C>A           | p.(Pro2988Gln)      | Driver                | 15,00% |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7577120C>T      | c.818G>A            | p.(Arg273His)       | TSG                   | 26,40% |
| <i>MEF2B</i>  | NM_001145785.2         | 19  | g.19258520C>T     | c.380G>A            | p.(Arg127Gln)       | TSG                   | 10,50% |
| <i>ZNF208</i> | NM_007153.3            | 19  | g.22156261G>C     | c.1575C>G           | p.(Tyr525*)         | Ambiguo               | 9,70%  |
| <i>SALL4</i>  | NM_020436.5            | 20  | g.50407172C>T     | c.1850G>A           | p.(Arg617Gln)       | OG                    | 18,90% |
| <i>AR</i>     | NM_000044.6            | X   | g.66905872G>A     | c.1789G>A           | p.(Ala597Thr)       | OG                    | 9,40%  |

**Tabla suplementaria 28. Variantes en genes driver en CCR de la paciente 308.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen            | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)         | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver  | VAF    |
|----------------|------------------------|-----|---------------------------|---------------------|---------------------|-----------------|--------|
| <i>AFF3</i>    | NM_001386135.1         | 2   | g.100210325C>T            | c.1798G>A           | p.(Ala600Thr)       | Ambiguo         | 25,00% |
| <i>APC</i>     | NM_000038.6            | 5   | g.112175216G>T            | c.3925G>T           | p.(Gly1309*)        | TSG             | 23,70% |
| <i>CSMD3</i>   | NM_198123.2            | 8   | g.113516208_113516209insA | c.4896-3_4896-2insT |                     | Driver          | 10,80% |
| <i>CBL</i>     | NM_005188.4            | 11  | g.119170327G>A            | c.2557G>A           | p.(Ala853Thr)       | Ambiguo, fusión | 16,40% |
| <i>TP53</i>    | NM_000546.6            | 17  | g.7578469_7578481del      | c.454_466del        | p.(Pro152Alafs*14)  | TSG             | 12,60% |
| <i>SMARCA4</i> | NM_003072.5            | 19  | g.11141427G>A             | c.3404G>A           | p.(Arg1135Gln)      | TSG, fusión     | 9,40%  |
| <i>PTPRT</i>   | NM_007050.6            | 20  | g.41076902C>A             | c.1518G>T           | p.(Trp506Cys)       | TSG             | 10,20% |
| <i>SALL4</i>   | NM_020436.5            | 20  | g.50406788C>T             | c.2234G>A           | p.(Arg745His)       | OG              | 7,20%  |
| <i>AR</i>      | NM_000044.6            | X   | g.66766168G>C             | c.1180G>C           | p.(Asp394His)       | OG              | 30,90% |

**Tabla suplementaria 29. Variantes en genes driver en CCR del paciente 310.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen            | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                          | Cambio nucleotídico           | Cambio aminoacídico | Tipo de driver  | VAF              |
|----------------|------------------------|-----|--|-------------------------------|---------------------|-----------------|------------------|
| <i>SPEN</i>    | NM_015001.3            | 1   | g.16248740_16248748del                     | c.1750-4_1754del              | p.(Val584Phefs*26)  | Ambiguo         | 5,40%            |
| <i>UGT2B17</i> | NM_001077.4            | 4   | g.69434039A>G                              | c.164T>C                      | p.(Val55Ala)        | OG              | 13,70%           |
| <i>APC</i>     | NM_000038.6            | 5   | g.112162945G>A<br>g.112175218_112175222del | c.1548+1G>A<br>c.3927_3931del | p.(Glu1309Aspfs*4)  | TSG             | 16,70%<br>32,80% |
| <i>FLT4</i>    | NM_182925.5            | 5   | g.180043968T>A                             | c.3028A>T                     | p.(Thr1010Ser)      | Ambiguo         | 6,10%            |
| <i>HOXA11</i>  | NM_005523.6            | 7   | g.27224131_27224156del                     | c.618_643del                  | p.(Ala207Profs*15)  | Ambiguo, fusión | 25,10%           |
| <i>SETD1B</i>  | NM_001353345.2         | 12  | g.122255792G>A                             | c.3400G>A                     | p.(Glu1134Lys)      | TSG             | 16,60%           |



**Continuación tabla suplementaria 29.**

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|---------------|------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>BRCA2</i>  | NM_000059.4            | 13  | g.32910463G>T     | c.1971G>T           | p.(Leu657Phe)       | TSG            | 11,00% |
| <i>NBEA</i>   | NM_001385012.1         | 13  | g.35716456A>G     | c.2387A>G           | p.(Glu796Gly)       | Ambiguo        | 9,80%  |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7577120C>T      | c.818G>A            | p.(Arg273His)       | TSG            | 40,50% |
| <i>NCOR1</i>  | NM_006311.4            | 17  | g.16022742T>C     | c.1910A>G           | p.(Lys637Arg)       | TSG, fusión    | 23,30% |
| <i>RNF213</i> | NM_001256071.3         | 17  | g.78313089C>T     | c.4922C>T           | p.(Ala1641Val)      | Fusión         | 19,90% |
| <i>SIRPA</i>  | NM_001040023.2         | 20  | g.1895838C>A      | c.173C>A            | p.(Thr58Asn)        | TSG            | 14,40% |

**Tabla suplementaria 30. Variantes en genes driver en CCR del paciente 401.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                          | Cambio nucleotídico         | Cambio aminoacídico                      | Tipo de driver | VAF              |
|---------------|------------------------|-----|--|-----------------------------|--|----------------|------------------|
| <i>PTPRC</i>  | NM_002838.5            | 1   | g.198668825C>G                             | c.431C>G                    | p.(Ala144Gly)                            | OG             | 19,20%           |
| <i>ACVR2A</i> | NM_001616.5            | 2   | g.148676110A>T<br>g.148676114_148676116del | c.911A>T<br>c.915_917del    | p.(His304Leu)<br>p.(Glu305del)           | TSG, fusión    | 18,30%<br>17,90% |
| <i>FAT4</i>   | NM_001291303.3         | 4   | g.126411911A>G                             | c.13940A>G                  | p.(His4647Arg)                           | TSG            | 12,10%           |
| <i>APC</i>    | NM_000038.6            | 5   | g.112163628_112163631dup<br>g.112175957dup | c.1551_1554dup<br>c.4666dup | p.(Leu519Tyrfs*19)<br>p.(Thr1556Asnfs*3) | TSG            | 20,40%<br>20,60% |
| <i>BRAF</i>   | NM_004333.6            | 7   | g.140453136A>T                             | c.1799T>A                   | p.(Val600Glu)                            | Fusión, OG     | 21,20%           |
| <i>KCNJ5</i>  | NM_000890.5            | 11  | g.128781283C>T                             | c.115C>T                    | p.(Arg39Cys)                             | Driver         | 21,60%           |
| <i>AKT1</i>   | NM_001382430.1         | 14  | g.105246551C>T                             | c.49G>A                     | p.(Glu17Lys)                             | OG             | 41,80%           |
| <i>KLHL36</i> | NM_024731.4            | 16  | g.84690558G>A                              | c.145G>A                    | p.(Val49Ile)                             | Ambiguo        | 20,10%           |
| <i>ZNF429</i> | NM_001001415.4         | 19  | g.21720842C>T                              | c.1987C>T                   | p.(Arg663Trp)                            | TSG            | 19,00%           |
| <i>PEG3</i>   | NM_006210.3            | 19  | g.57325375C>A                              | c.4435G>T                   | p.(Ala1479Ser)                           | OG             | 19,20%           |

**Tabla suplementaria 31. Variantes en genes driver en CCR del paciente 402.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)       | Cambio nucleotídico | Cambio aminoacídico       | Tipo de <i>driver</i> | VAF    |
|---------------|------------------------|-----|-------------------------|---------------------|---------------------------|-----------------------|--------|
| <i>ALK</i>    | NM_004304.5            | 2   | g.30143441C>T           | c.85G>A             | p.(Ala29Thr)              | Fusión, OG            | 16,10% |
| <i>PIK3CA</i> | NM_006218.4            | 3   | g.178936082G>A          | c.1624G>A           | p.(Glu542Lys)             | OG                    | 33,70% |
| <i>PDGFRA</i> | NM_006206.6            | 4   | g.55144678C>T           | c.2152C>T           | p.(Arg718Trp)             | OG                    | 23,00% |
| <i>DROSHA</i> | NM_001382508.1         | 5   | g.31495479_31495487del  | c.1669-6_1671del    | p.(Ala557del)             | TSG                   | 5,30%  |
| <i>APC</i>    | NM_000038.6            | 5   | g.112175198C>T          | c.3907C>T           | p.(Gln1303*)              | TSG                   | 39,00% |
| <i>CREBBP</i> | NM_004380.3            | 16  | g.3778301_3778302insGCT | c.6746_6747insAGC   | p.(Arg2249_Met2250insAla) | Ambiguo, fusión       | 44,90% |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7578406C>T            | c.524G>A            | p.(Arg175His)             | TSG                   | 46,50% |

**Tabla suplementaria 32. Variantes en genes driver en CCR del paciente 501.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                          | Cambio nucleotídico         | Cambio aminoacídico               | Tipo de <i>driver</i> | VAF              |
|---------------|------------------------|-----|--|-----------------------------|-----------------------------------|-----------------------|------------------|
| <i>ARID1A</i> | NM_006015.6            | 1   | g.27088795C>T                              | c.2404C>T                   | p.(Gln802*)                       | TSG, fusión           | 33,40%           |
| <i>SRGAP3</i> | NM_014850.3            | 3   | g.9036036A>G                               | c.2399T>C                   | p.(Val800Ala)                     | OG                    | 39,50%           |
| <i>GMPS</i>   | NM_003875.3            | 3   | g.155652739C>T                             | c.1711C>T                   | p.(Pro571Ser)                     | TSG                   | 10,80%           |
| <i>APC</i>    | NM_000038.6            | 5   | g.112174214_112174227del<br>g.112175147G>T | c.2923_2936del<br>c.3856G>T | p.(Lys975Glufs*5)<br>p.(Glu1286*) | TSG                   | 34,00%<br>41,30% |
| <i>UBR5</i>   | NM_015902.6            | 8   | g.103354693_103354701del                   | c.1098_1098+8del            | p.(Asp367Metfs*53)                | Ambiguo               | 5,90%            |
| <i>BCL9L</i>  | NM_001378213.1         | 11  | g.118773227G>A                             | c.1225C>T                   | p.(Arg409Trp)                     | Ambiguo               | 56,10%           |
| <i>KRAS</i>   | NM_004985.5            | 12  | g.25398281C>T                              | c.38G>A                     | p.(Gly13Asp)                      | OG                    | 49,30%           |
| <i>BCL11B</i> | NM_138576.4            | 14  | g.99697736T>C                              | c.586A>G                    | p.(Thr196Ala)                     | Ambiguo, fusión       | 82,90%           |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7577538C>T                               | c.743G>A                    | p.(Arg248Gln)                     | TSG                   | 77,30%           |

**Tabla suplementaria 33. Variantes en genes driver en CCR del paciente COAD-1.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen     | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver  | VAF    |
|---------|-------------------------|-----|-------------------|---------------------|---------------------|-----------------|--------|
| APC     | NM_000038.6             | 5   | g.112175255G>T    | c.3964G>T           | p.(Glu1322*)        | TSG             | 88,40% |
| NCOA2   | NM_006540.4             | 8   | g.71071783T>C     | c.1081A>G           | p.(Thr361Ala)       | Ambiguo, fusión | 48,90% |
| DDX6    | NM_004397.6             | 11  | g.118627953C>T    | c.1037G>A           | p.(Arg346Gln)       | Ambiguo         | 44,20% |
| KRAS    | NM_004985.5             | 12  | g.25398284C>A     | c.35G>T             | p.(Gly12Val)        | OG              | 47,80% |
| CBFA2T3 | NM_005187.6             | 16  | g.88945758C>G     | c.1582G>C           | p.(Glu528Gln)       | Ambiguo, fusión | 50,00% |
| CCR7    | NM_001838.4             | 17  | g.38711505C>T     | c.626G>A            | p.(Arg209Gln)       | TSG             | 47,40% |
| STAG2   | NM_001042750.2          | X   | g.123217392del    | c.3046del           | p.(Arg1016Glufs*11) | TSG             | 92,20% |

**Tabla suplementaria 34. Variantes en genes driver en CCR de la paciente COAD-2.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen   | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|-------|-------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| EPHA3 | NM_005233.6             | 3   | g.89521640G>T     | c.2717G>T           | p.(Ser906Ile)       | Ambiguo        | 21,30% |
| APC   | NM_000038.6             | 5   | g.112175423C>T    | c.4132C>T           | p.(Gln1378*)        | TSG            | 76,90% |
| NCOR2 | NM_006312.6             | 12  | g.124832436G>A    | c.4007C>T           | p.(Pro1336Leu)      | Ambiguo        | 48,30% |
| TP53  | NM_000546.6             | 17  | g.7577538C>T      | c.743G>A            | p.(Arg248Gln)       | TSG            | 70,80% |

**Tabla suplementaria 35. Variantes en genes driver en CCR del paciente COAD-3.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen    | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|--------|-------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| TP53   | NM_000546.6             | 17  | g.7578445A>T      | c.485T>A            | p.(Ile162Asn)       | TSG            | 34,40% |
| ASXL1  | NM_015338.6             | 20  | g.31024758C>T     | c.4243C>T           | p.(Arg1415*)        | TSG            | 32,40% |
| NF2    | NM_000268.4             | 22  | g.30090742T>C     | c.1739T>C           | p.(Leu580Pro)       | TSG            | 9,40%  |
| BCOR   | NM_001123385.2          | X   | g.39923059G>A     | c.3649C>T           | p.(Arg1217*)        | TSG, fusión    | 35,90% |
| BCORL1 | NM_001379451.1          | X   | g.129150139C>T    | c.3391C>T           | p.(Gln1131*)        | TSG, OG        | 60,00% |

**Tabla suplementaria 36. Variantes en genes driver en CCR del paciente COAD-5.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)            | Cambio nucleotídico    | Cambio aminoácido | Tipo de driver | VAF              |
|---------------|------------------------|-----|------------------------------|------------------------|-------------------|----------------|------------------|
| <i>SRGAP3</i> | NM_014850.4            | 3   | g.9099942C>T                 | c.1016G>A              | p.(Gly339Glu)     | OG             | 44,20%           |
| <i>POLQ</i>   | NM_199420.4            | 3   | g.121207891C>T               | c.3887G>A              | p.(Gly1296Asp)    | Ambiguo        | 20,40%           |
| <i>TRAF3</i>  | NM_145725.3            | 14  | g.103342783A>T               | c.491A>T               | p.(Asp164Val)     | TSG            | 13,60%           |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7578370C>T<br>g.7579451G>A | c.559+1G>A<br>c.236C>T | p.(Ala79Val)      | TSG            | 40,40%<br>44,70% |

**Tabla suplementaria 37. Variantes en genes driver en CCR de la paciente COAD-6.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                       | Cambio nucleotídico        | Cambio aminoácido                    | Tipo de driver | VAF              |
|---------------|------------------------|-----|---|----------------------------|--------------------------------------|----------------|------------------|
| <i>PCBP1</i>  | NM_006196.4            | 2   | g.70315174T>A                           | c.299T>A                   | p.(Leu100Gln)                        | OG             | 30,80%           |
| <i>APC</i>    | NM_000038.6            | 5   | g.112128143C>T                          | c.646C>T                   | p.(Arg216*)                          | TSG            | 24,10%           |
| <i>FAS</i>    | NM_000043.6            | 10  | g.90770546T>C                           | c.542T>C                   | p.(Leu181Pro)                        | TSG            | 30,40%           |
| <i>TCF7L2</i> | NM_001367943.1         | 10  | g.114925316dup                          | c.1445dup                  | p.(Cys486Valfs*8)                    | Fusión, OG     | 12,40%           |
| <i>ATM</i>    | NM_000051.4            | 11  | g.108236153G>T                          | c.9089G>T                  | p.(Gly3030Val)                       | TSG            | 19,10%           |
| <i>KRAS</i>   | NM_004985.5            | 12  | g.25398284C>G                           | c.35G>C                    | p.(Gly12Ala)                         | OG             | 31,20%           |
| <i>SOX9</i>   | NM_000346.4            | 17  | g.70117935C>T<br>g.70120346_70120349dup | c.403C>T<br>c.1348_1351dup | p.(Leu135Phe)<br>p.(Tyr451Phefs*128) | TSG            | 28,10%<br>32,10% |
| <i>RNF213</i> | NM_001256071.3         | 17  | g.78333931G>A                           | c.11125G>A                 | p.(Val3709Ile)                       | Fusión         | 25,00%           |
| <i>IRAK1</i>  | NM_001569.4            | X   | g.153281983T>A                          | c.1141A>T                  | p.(Thr381Ser)                        | TSG            | 29,80%           |

**Tabla suplementaria 38. Variantes en genes driver en CCR de la paciente COAD-7.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|---------------|------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>SOS1</i>   | NM_005633.4            | 2   | g.39224494A>G     | c.2864T>C           | p.(Leu955Pro)       | OG             | 28,70% |
| <i>FBXW7</i>  | NM_001349798.2         | 4   | g.153271206G>C    | c.572C>G            | p.(Ser191*)         | TSG            | 21,90% |
| <i>APC</i>    | NM_000038.6            | 5   | g.112174466G>T    | c.3175G>T           | p.(Glu1059*)        | TSG            | 19,80% |
| <i>NUP214</i> | NM_005085.4            | 9   | g.134064512G>A    | c.3568G>A           | p.(Ala1190Thr)      | OG             | 45,90% |
| <i>SMAD3</i>  | NM_005902.4            | 15  | g.67473722C>T     | c.802C>T            | p.(Arg268Cys)       | TSG            | 63,60% |
| <i>IDH2</i>   | NM_002168.4            | 15  | g.90630482C>T     | c.829G>A            | p.(Asp277Asn)       | OG             | 47,50% |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7578526C>T      | c.404G>A            | p.(Cys135Tyr)       | TSG            | 58,50% |

**Tabla suplementaria 39. Variantes en genes driver en CCR de la paciente COAD-8.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen         | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|-------------|------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>IDH1</i> | NM_005896.4            | 2   | g.209104604G>A    | c.974C>T            | p.(Thr325Met)       | OG             | 38,10% |
| <i>APC</i>  | NM_000038.6            | 5   | g.112175235C>A    | c.3944C>A           | p.(Ser1315*)        | TSG            | 87,20% |
| <i>ATM</i>  | NM_000051.4            | 11  | g.108168027dup    | c.4923dup           | p.(Gly1642Trpfs*23) | TSG            | 80,00% |
| <i>KRAS</i> | NM_004985.5            | 12  | g.25398284C>T     | c.35G>A             | p.(Gly12Asp)        | OG             | 37,30% |
| <i>TSC2</i> | NM_000548.5            | 16  | g.2104304G>A      | c.344G>A            | p.(Arg115His)       | TSG            | 25,90% |

**Tabla suplementaria 40. Variantes en genes driver en CCR del paciente COAD-9.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|---------------|------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>DHX9</i>   | NM_001357.5            | 1   | g.182847257G>A    | c.2300G>A           | p.(Arg767Gln)       | OG             | 16,50% |
| <i>PTPRC</i>  | NM_002838.5            | 1   | g.198671524G>A    | c.448G>A            | p.(Gly150Arg)       | OG             | 21,50% |
| <i>PIK3CA</i> | NM_006218.4            | 3   | g.178936091G>A    | c.1633G>A           | p.(Glu545Lys)       | OG             | 10,00% |
| <i>APC</i>    | NM_000038.6            | 5   | g.112178000C>T    | c.6709C>T           | p.(Arg2237*)        | TSG            | 31,00% |
| <i>FAT3</i>   | NM_001367949.2         | 11  | g.92624169G>A     | c.13660G>A          | p.(Gly4554Ser)      | Ambiguo        | 49,20% |
| <i>NBEA</i>   | NM_001385012.1         | 13  | g.36229854G>A     | c.8330G>A           | p.(Arg2777Gln)      | Ambiguo        | 7,60%  |
| <i>PCDH17</i> | NM_001040429.3         | 13  | g.58207258C>A     | c.578C>A            | p.(Pro193Gln)       | TSG            | 10,40% |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7579415C>T      | c.272G>A            | p.(Trp91*)          | TSG            | 61,30% |
| <i>ASXL1</i>  | NM_015338.6            | 20  | g.31024843A>G     | c.4328A>G           | p.(Lys1443Arg)      | TSG            | 8,70%  |

**Tabla suplementaria 41. Variantes en genes driver en CCR de la paciente COAD-11.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|---------------|------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>RPL22</i>  | NM_000983.4            | 1   | g.6257792del      | c.44del             | p.(Lys15Argfs*5)    | OG             | 27,30% |
| <i>SPEN</i>   | NM_015001.3            | 1   | g.16264500C>T     | c.10703C>T          | p.(Thr3568Met)      | Ambiguo        | 20,70% |
| <i>ARID1A</i> | NM_006015.6            | 1   | g.27106354C>T     | c.5965C>T           | p.(Arg1989*)        | TSG, fusión    | 21,40% |
| <i>CDKN2C</i> | NM_078626.3            | 1   | g.51436169G>T     | c.129G>T            | p.(Gln43His)        | TSG            | 12,40% |
| <i>JAK1</i>   | NM_002227.4            | 1   | g.65307004del     | c.2580del           | p.(Lys860Asnfs*16)  | Ambiguo        | 14,50% |
| <i>RBM15</i>  | NM_022768.5            | 1   | g.110883184C>T    | c.1157C>T           | p.(Thr386Met)       | TSG, fusión    | 23,40% |
| <i>ASXL2</i>  | NM_018263.6            | 2   | g.25965049G>A     | c.4157C>T           | p.(Ala1386Val)      | TSG            | 18,80% |
| <i>ALK</i>    | NM_004304.5            | 2   | g.29551273G>A     | c.1357C>T           | p.(Gln453*)         | Fusión, OG     | 21,90% |
| <i>ACVR2A</i> | NM_001616.5            | 2   | g.148683693del    | c.1310del           | p.(Lys437Argfs*5)   | TSG, fusión    | 35,00% |
| <i>CUL3</i>   | NM_003590.5            | 2   | g.225379410C>T    | c.458G>A            | p.(Arg153His)       | TSG            | 23,30% |
| <i>RAF1</i>   | NM_002880.4            | 3   | g.12660103G>A     | c.118C>T            | p.(Arg40Cys)        | Fusión, OG     | 25,20% |
| <i>FGD5</i>   | NM_152536.4            | 3   | g.14861539G>A     | c.961G>A            | p.(Ala321Thr)       | OG             | 25,00% |

Continuación tabla suplementaria 41.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                | Cambio nucleotídico   | Cambio aminoacídico            | Tipo de driver     | VAF              |
|---------------|------------------------|-----|----------------------------------|-----------------------|--------------------------------|--------------------|------------------|
| <i>PRKCD</i>  | NM_006254.4            | 3   | g.53215468C>T                    | c.385C>T              | p.(Gln129*)                    | TSG                | 24,60%           |
| <i>POLQ</i>   | NM_199420.4            | 3   | g.121200622G>A                   | c.6008C>T             | p.(Pro2003Leu)                 | Ambiguo            | 23,50%           |
| <i>PIK3CA</i> | NM_006218.4            | 3   | g.178916876G>A<br>g.178916936G>A | c.263G>A<br>c.323G>A  | p.(Arg88Gln)<br>p.(Arg108His)  | OG                 | 18,00%<br>21,10% |
| <i>KLHL6</i>  | NM_130446.4            | 3   | g.183217531G>A                   | c.994C>T              | p.(Arg332Trp)                  | TSG                | 25,00%           |
| <i>TP63</i>   | NM_003722.5            | 3   | g.189526257C>T                   | c.521C>T              | p.(Pro174Leu)                  | Ambiguo            | 19,60%           |
| <i>FGFR3</i>  | NM_000142.5            | 4   | g.1808985C>T                     | c.2417C>T             | p.(Thr806Met)                  | Fusión,<br>OG      | 60,00%           |
| <i>KDM3B</i>  | NM_016604.4            | 5   | g.137717277G>A                   | c.778G>A              | p.(Glu260Lys)                  | Ambiguo            | 26,10%           |
| <i>CARD11</i> | NM_032415.7            | 7   | g.2978404C>T                     | c.926G>A              | p.(Arg309His)                  | OG                 | 15,40%           |
| <i>ELN</i>    | NM_000501.4            | 7   | g.73461044del                    | c.590del              | p.(Gly197Aspfs*126)            | Ambiguo            | 19,40%           |
| <i>BRAF</i>   | NM_004333.6            | 7   | g.140482933dup                   | c.1208dup             | p.(Ala404Cysfs*9)              | Fusión,<br>OG      | 15,40%           |
| <i>KMT2C</i>  | NM_170606.3            | 7   | g.151845367G>A                   | c.13645C>T            | p.(Arg4549Cys)                 | TSG                | 21,80%           |
| <i>UBR5</i>   | NM_015902.6            | 8   | g.103307431C>T                   | c.4058+1G>A           | p.(Glu2121Lysfs*28)            | Ambiguo            | 18,50%           |
| <i>WNK2</i>   | NM_006648.4            | 9   | g.96015213C>T                    | c.1883C>T             | p.(Ser628Leu)                  | Ambiguo            | 20,00%           |
| <i>TNC</i>    | NM_002160.4            | 9   | g.117838345C>A                   | c.2916G>T             | p.(Lys972Asn)                  | Ambiguo            | 25,00%           |
| <i>TSC1</i>   | NM_000368.5            | 9   | g.135780989G>A                   | c.1976C>T             | p.(Ala659Val)                  | TSG                | 23,50%           |
| <i>LARP4B</i> | NM_015155.3            | 10  | g.890945del                      | c.487del              | p.(Thr163Hisfs*47)             | Driver             | 23,70%           |
| <i>NRP1</i>   | NM_003873.7            | 10  | g.33559762C>T                    | c.271G>A              | p.(Asp91Asn)                   | TSG                | 24,00%           |
| <i>EGR2</i>   | NM_000399.5            | 10  | g.64573838G>A                    | c.560C>T              | p.(Ala187Val)                  | Ambiguo            | 19,80%           |
| <i>KMT2A</i>  | NM_001197104.2         | 11  | g.118352594G>A                   | c.3799G>A             | p.(Val1267Ile)                 | Ambiguo,<br>fusión | 19,50%           |
| <i>KDM5A</i>  | NM_001042603.3         | 12  | g.416960del                      | c.3597del             | p.(Gly1200Aspfs*9)             | OG                 | 8,00%            |
| <i>KMT2D</i>  | NM_003482.4            | 12  | g.49416416C>T                    | c.16295G>A            | p.(Arg5432Gln)                 | Ambiguo            | 23,80%           |
| <i>SETD1B</i> | NM_001353345.2         | 12  | g.122242665del                   | c.22del               | p.(His8Thrfs*27)               | TSG                | 23,80%           |
| <i>POLE</i>   | NM_006231.4            | 12  | g.133202738C>T                   | c.6496G>A             | p.(Asp2166Asn)                 | Ambiguo            | 17,20%           |
| <i>BRCA2</i>  | NM_000059.4            | 13  | g.32906571del                    | c.956del              | p.(Asn319Ilefs*5)              | TSG                | 20,60%           |
| <i>BCL11B</i> | NM_138576.4            | 14  | g.99640651C>T<br>g.99642394G>A   | c.2522G>A<br>c.779C>T | p.(Arg841His)<br>p.(Thr260Met) | Ambiguo,<br>fusión | 35,60%<br>27,30% |
| <i>HERC2</i>  | NM_004667.6            | 15  | g.28397918G>A                    | c.10805C>T            | p.(Ser3602Leu)                 | Ambiguo            | 12,50%           |
| <i>SIN3A</i>  | NM_001145358.2         | 15  | g.75668178C>T                    | c.3419G>A             | p.(Arg1140Gln)                 | OG                 | 23,00%           |
| <i>CHD2</i>   | NM_001271.4            | 15  | g.93563304T>C                    | c.4969T>C             | p.(Trp1657Arg)                 | TSG                | 16,70%           |

Continuación tabla suplementaria 41.

| Gen            | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)      | Cambio nucleotídico | Cambio aminoácido   | Tipo de driver | VAF    |
|----------------|------------------------|-----|------------------------|---------------------|---------------------|----------------|--------|
| <i>AXIN1</i>   | NM_003502.4            | 16  | g.359972C>T            | c.1116+1G>A         |                     | TSG            | 15,00% |
| <i>MYH11</i>   | NM_002474.3            | 16  | g.15844140G>A          | c.1913C>T           | p.(Ser638Leu)       | OG             | 20,30% |
| <i>CTCF</i>    | NM_006565.4            | 16  | g.67645904C>T          | c.832C>T            | p.(Arg278Cys)       | TSG            | 23,70% |
| <i>ZFH3</i>    | NM_006885.4            | 16  | g.72993432C>T          | c.613G>A            | p.(Gly205Arg)       | TSG            | 15,40% |
| <i>ZFH3</i>    | NM_006885.4            | 16  | g.72993432C>T          | c.613G>A            | p.(Gly205Arg)       | TSG            | 15,40% |
| <i>RNF43</i>   | NM_017763.6            | 17  | g.56440904G>A          | c.433C>T            | p.(Arg145*)         | TSG            | 24,50% |
| <i>AXIN2</i>   | NM_004655.4            | 17  | g.63554450_63554451del | c.288_289del        | p.(Thr97Phefs*43)   | TSG            | 21,40% |
| <i>MAP2K7</i>  | NM_145185.4            | 19  | g.7976364C>T           | c.980C>T            | p.(Thr327Met)       | TSG            | 28,00% |
| <i>ZNF429</i>  | NM_001001415.4         | 19  | g.21720016_21720017del | c.1161_1162del      | p.(His387Glnfs*13)  | TSG            | 22,40% |
| <i>KMT2B</i>   | NM_014727.3            | 19  | g.36222943C>T          | c.5572C>T           | p.(Arg1858Cys)      | TSG            | 23,50% |
| <i>POLD1</i>   | NM_002691.4            | 19  | g.50912817G>A          | c.2048G>A           | p.(Arg683His)       | Ambiguo        | 20,00% |
| <i>SMARCB1</i> | NM_003073.5            | 22  | g.24176338C>T          | c.1129C>T           | p.(Arg377Cys)       | TSG, fusión    | 28,60% |
| <i>ZNRF3</i>   | NM_001206998.2         | 22  | g.29446744C>T          | c.2575C>T           | p.(Arg859*)         | TSG            | 20,00% |
| <i>BCOR</i>    | NM_001123385.2         | X   | g.39930286del          | c.3182del           | p.(Lys1061Serfs*52) | TSG, fusión    | 26,30% |
| <i>USP9X</i>   | NM_001039591.3         | X   | g.41076578G>A          | c.6191G>A           | p.(Arg2064His)      | TSG            | 26,20% |

Tabla suplementaria 42. Variantes en genes driver en CCR de la paciente COAD-12. TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen          | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoácido | Tipo de driver | VAF    |
|--------------|------------------------|-----|-------------------|---------------------|-------------------|----------------|--------|
| <i>KIT</i>   | NM_000222.3            | 4   | g.55564660G>C     | c.548G>C            | p.(Cys183Ser)     | OG             | 25,30% |
| <i>APC</i>   | NM_000038.6            | 5   | g.112128191C>T    | c.694C>T            | p.(Arg232*)       | TSG            | 23,60% |
| <i>KRAS</i>  | NM_004985.5            | 12  | g.25398218G>A     | c.101C>T            | p.(Pro34Leu)      | OG             | 15,40% |
| <i>TP53</i>  | NM_000546.6            | 17  | g.7577141C>A      | c.797G>T            | p.(Gly266Val)     | TSG            | 35,00% |
| <i>ASXL1</i> | NM_015338.6            | 20  | g.31022289C>T     | c.1774C>T           | p.(Gln592*)       | TSG            | 29,00% |



**Tabla suplementaria 43. Variantes en genes driver en CCR de la paciente COAD-13.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen          | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37)           | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|--------------|-------------------------|-----|-----------------------------|---------------------|---------------------|----------------|--------|
| <i>PPP6C</i> | NM_002721.5             | 9   | g.127912170 <sup>a</sup> >C | c.700T>G            | p.(Cys234Gly)       | TSG            | 32,80% |
| <i>KRAS</i>  | NM_004985.5             | 12  | g.25378562C>T               | c.436G>A            | p.(Ala146Thr)       | OG             | 83,60% |
| <i>AKT1</i>  | NM_001382430.1          | 14  | g.105246457C>T              | c.143G>A            | p.(Arg48His)        | OG             | 38,60% |
| <i>BRD7</i>  | NM_013263.5             | 16  | g.50383933C>T               | c.591+1G>A          |                     | TSG            | 50,30% |
| <i>PEG3</i>  | NM_006210.3             | 19  | g.57326722G>A               | c.3088C>T           | p.(Arg1030Trp)      | OG             | 26,50% |
| <i>SALL4</i> | NM_020436.5             | 20  | g.50405551C>T               | c.2591G>A           | p.(Arg864His)       | OG             | 41,20% |

**Tabla suplementaria 44. Variantes en genes driver en CCR de la paciente COAD-17.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tráncrito (MANE GRCh37) | Chr | Variante (GRCh37)      | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|---------------|-------------------------|-----|------------------------|---------------------|---------------------|----------------|--------|
| <i>TET2</i>   | NM_001127208.3          | 4   | g.106164818T>C         | c.3686T>C           | p.(Leu1229Pro)      | TSG            | 7,30%  |
| <i>APC</i>    | NM_000038.6             | 5   | g.112173517del         | c.2226del           | p.(Met743Cysfs*18)  | TSG            | 27,50% |
| <i>KMT2C</i>  | NM_170606.3             | 7   | g.151932922C>T         | c.2749G>A           | p.(Gly917Arg)       | TSG            | 28,60% |
| <i>NUP214</i> | NM_005085.4             | 9   | g.134008000del         | c.681del            | p.(Val228Serfs*29)  | OG             | 7,00%  |
| <i>BRCA2</i>  | NM_000059.4             | 13  | g.32899311C>G          | c.415C>G            | p.(Leu139Val)       | TSG            | 20,40% |
| <i>PDP1</i>   | NM_017990.5             | 16  | g.70162989G>A          | c.571G>A            | p.(Val191Met)       | Ambiguo        | 12,40% |
| <i>TP53</i>   | NM_000546.6             | 17  | g.7574017C>A           | c.1010G>T           | p.(Arg337Leu)       | TSG            | 53,60% |
| <i>SOX9</i>   | NM_000346.4             | 17  | g.70117603_70117604dup | c.71_72dup          | p.(Thr25Profs*37)   | TSG            | 19,40% |
| <i>RGL3</i>   | NM_001035223.4          | 19  | g.11527598C>T          | c.283G>A            | p.(Ala95Thr)        | OG             | 16,90% |
| <i>CRNKL1</i> | NM_001278628.2          | 20  | g.20028411C>T          | c.620G>A            | p.(Arg207Gln)       | OG             | 23,70% |

**Tabla suplementaria 45. Variantes en genes driver en CCR del paciente READ-1.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen            | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                | Cambio nucleotídico   | Cambio aminoacídico         | Tipo de driver | VAF              |
|----------------|------------------------|-----|----------------------------------|-----------------------|-----------------------------|----------------|------------------|
| <i>ELF3</i>    | NM_004433.5            | 1   | g.201983019dup                   | c.868dup              | p.(Glu290Glyfs*11)          | TSG            | 26,10%           |
| <i>ASXL2</i>   | NM_018263.6            | 2   | g.25982386G>A                    | c.904C>T              | p.(Arg302*)                 | TSG            | 85,10%           |
| <i>MAP2</i>    | NM_001375505.1         | 2   | g.210574917A>G                   | c.5012A>G             | p.(Lys1671Arg)              | Ambiguo        | 8,40%            |
| <i>APC</i>     | NM_000038.6            | 5   | g.112151204C>T<br>g.112175507C>T | c.847C>T<br>c.4216C>T | p.(Arg283*)<br>p.(Gln1406*) | TSG            | 15,00%<br>66,30% |
| <i>CARD11</i>  | NM_032415.7            | 7   | g.2962953G>A                     | c.1955C>T             | p.(Ser652Leu)               | OG             | 39,40%           |
| <i>NRP1</i>    | NM_003873.7            | 10  | g.33502450T>G                    | c.1478A>C             | p.(Lys493Thr)               | TSG            | 23,70%           |
| <i>RELA</i>    | NM_021975.4            | 11  | g.65427176G>A                    | c.520C>T              | p.(Arg174Cys)               | OG             | 81,80%           |
| <i>FAT3</i>    | NM_001367949.2         | 11  | g.92565169C>A                    | c.9863C>A             | p.(Pro3288His)              | Ambiguo        | 62,50%           |
| <i>RNF6</i>    | NM_005977.4            | 13  | g.26788761T>A                    | c.1258A>T             | p.(Arg420*)                 | TSG            | 19,70%           |
| <i>PCDH17</i>  | NM_001040429.3         | 13  | g.58299342T>A                    | c.3394T>A             | p.(Ser1132Thr)              | TSG            | 7,70%            |
| <i>MGA</i>     | NM_001080541.1         | 15  | g.42042703G>C                    | c.7045G>C             | p.(Asp2349His)              | TSG            | 24,30%           |
| <i>GRIN2A</i>  | NM_001134407.3         | 16  | g.10031897T>G                    | c.926A>C              | p.(Lys309Thr)               | TSG, OG        | 24,00%           |
| <i>TP53</i>    | NM_000546.6            | 17  | g.7579361A>C                     | c.326T>G              | p.(Phe109Cys)               | TSG            | 74,60%           |
| <i>PEG3</i>    | NM_006210.3            | 19  | g.57325595T>G                    | c.4215A>C             | p.(Glu1405Asp)              | OG             | 22,60%           |
| <i>PTPRT</i>   | NM_007050.6            | 20  | g.40727077T>C                    | c.3830A>G             | p.(Glu1277Gly)              | TSG            | 30,00%           |
| <i>SMARCA1</i> | NM_001282874.2         | X   | g.128614769G>C                   | c.2351C>G             | p.(Pro784Arg)               | TSG            | 37,80%           |

**Tabla suplementaria 46. Variantes en genes driver en CCR de la paciente READ-4.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen          | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37) | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver | VAF    |
|--------------|------------------------|-----|-------------------|---------------------|---------------------|----------------|--------|
| <i>APC</i>   | NM_000038.6            | 5   | g.112175216G>T    | c.3925G>T           | p.(Glu1309*)        | TSG            | 84,20% |
| <i>TP53</i>  | NM_000546.6            | 17  | g.7578392C>A      | c.538G>T            | p.(Glu180*)         | TSG            | 51,60% |
| <i>PTPRT</i> | NM_007050.6            | 20  | g.41419939C>T     | c.382G>A            | p.(Gly128Ser)       | TSG            | 17,90% |

**Tabla suplementaria 47. Variantes en genes driver en CCR del paciente READ-5.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen           | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)        | Cambio nucleotídico | Cambio aminoacídico | Tipo de driver  | VAF    |
|---------------|------------------------|-----|--------------------------|---------------------|---------------------|-----------------|--------|
| <i>FUBP1</i>  | NM_003902.5            | 1   | g.78435700C>A            | c.121-1G>T          |                     | Ambiguo         | 25,00% |
| <i>RANBP2</i> | NM_006267.5            | 2   | g.109384704G>T           | c.7709G>T           | p.(Ser2570Ile)      | TSG, fusión     | 46,20% |
| <i>APC</i>    | NM_000038.6            | 5   | g.112173854_112173855dup | c.2563_2564dup      | p.(Arg856Asnfs*6)   | TSG             | 20,00% |
| <i>RET</i>    | NM_020975.6            | 10  | g.43597981C>T            | c.529C>T            | p.(Arg177Trp)       | Ambiguo, fusión | 37,00% |
| <i>TCF7L2</i> | NM_001367943.1         | 10  | g.114912188C>T           | c.1258C>T           | p.(Arg420Trp)       | Fusión, OG      | 53,30% |
| <i>KRAS</i>   | NM_004985.5            | 12  | g.25398285C>A            | c.34G>T             | p.(Gly12Cys)        | OG              | 30,80% |
| <i>TP53</i>   | NM_000546.6            | 17  | g.7577568C>T             | c.713G>A            | p.(Cys238Tyr)       | TSG             | 50,00% |

**Tabla suplementaria 48. Variantes en genes driver en CCR de la paciente READ-6.** TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.

| Gen            | Tránsito (MANE GRCh37) | Chr | Variante (GRCh37)                | Cambio nucleotídico    | Cambio aminoacídico          | Tipo de driver | VAF              |
|----------------|------------------------|-----|----------------------------------|------------------------|------------------------------|----------------|------------------|
| <i>APC</i>     | NM_000038.6            | 5   | g.112174631C>T<br>g.112175216G>T | c.3340C>T<br>c.3925G>T | p.(Arg1114*)<br>p.(Glu1309*) | TSG            | 28,60%<br>37,50% |
| <i>FAM135B</i> | NM_015912.4            | 8   | g.139165293T>G                   | c.1425A>C              | p.(Glu475Asp)                | OG             | 35,20%           |
| <i>BCL9L</i>   | NM_001378213.1         | 11  | g.118770659dup                   | c.3380dup              | p.(Pro1128Thrfs*42)          | TSG, OG        | 77,30%           |
| <i>TP53</i>    | NM_000546.6            | 17  | g.7578370C>T                     | c.559+1G>A             |                              | TSG            | 84,40%           |
| <i>SOX9</i>    | NM_000346.4            | 17  | g.70120178dup                    | c.1180dup              | p.(Arg394Profs*184)          | TSG            | 67,90%           |
| <i>RNF213</i>  | NM_001256071.3         | 17  | g.78360568del                    | c.14799del             | p.(Leu4934Serfs*38)          | Fusión         | 82,80%           |
| <i>SMAD4</i>   | NM_005359.6            | 18  | g.48604662T>G                    | c.1484T>G              | p.(Leu495Arg)                | TSG            | 66,30%           |

**Tabla suplementaria 49. Variantes en genes driver en CCR de la paciente READ-8. TSG: gen supresor tumoral, OG: oncogén, VAF: frecuencia alélica de la variante alterada.**

| Gen          | Tránsito<br>(MANE<br>GRCh37) | Chr | Variante (GRCh37) | Cambio<br>nucleotídico | Cambio<br>aminoacídico | Tipo de<br><i>driver</i> | VAF    |
|--------------|------------------------------|-----|-------------------|------------------------|------------------------|--------------------------|--------|
| <i>ERBB4</i> | NM_005235.3                  | 2   | g.212295749T>G    | c.2564A>C              | p.(Asn855Thr)          | Ambiguo                  | 54,10% |
| <i>ATM</i>   | NM_000051.4                  | 11  | g.108218045A>G    | c.8624A>G              | p.(Asn2875Ser)         | TSG                      | 27,90% |
| <i>RNF43</i> | NM_017763.6                  | 17  | g.56492812del     | c.127del               | p.(Glu43Asnfs*8)       | TSG                      | 65,40% |

**Tabla suplementaria 50. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 105.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón;  $F$ : estadístico  $F$  del  $F$  test de cuasiverosimilitud;  $p$ -valor:  $p$ -valor nominal derivado del valor  $F$  sin ningún tipo de corrección múltiple;  $\text{FDR}$ :  $p$ -valor con corrección Benjamin-Hochberg.

|                 | $\log_2\text{FC}$ | $\log_2\text{CPM}$     | $F$    | $p$ -valor             | $\text{FDR}$           |
|-----------------|-------------------|------------------------|--------|------------------------|------------------------|
| <i>SLC6A6</i>   | 3,48              | 4,43                   | 119,51 | $1,67 \times 10^{-16}$ | $6,12 \times 10^{-14}$ |
| <i>FBXO41</i>   | 2,38              | 3,10                   | 73,87  | $2,14 \times 10^{-12}$ | $3,10 \times 10^{-10}$ |
| <i>GDPD5</i>    | 2,41              | 2,80                   | 57,13  | $1,56 \times 10^{-10}$ | $1,59 \times 10^{-8}$  |
| <i>ALDH3B2</i>  | 4,89              | -2,08                  | 44,59  | $5,84 \times 10^{-9}$  | $3,82 \times 10^{-7}$  |
| <i>MDFI</i>     | 4,03              | 0,38                   | 42,49  | $1,12 \times 10^{-8}$  | $6,75 \times 10^{-7}$  |
| <i>DNAH14</i>   | -3,63             | 1,50                   | 36,79  | $6,96 \times 10^{-8}$  | $3,37 \times 10^{-6}$  |
| <i>ADAMTSL1</i> | -2,93             | 3,76                   | 27,90  | $1,51 \times 10^{-6}$  | $4,75 \times 10^{-5}$  |
| <i>RNF130</i>   | -1,54             | 5,57                   | 27,61  | $1,68 \times 10^{-6}$  | $5,16 \times 10^{-5}$  |
| <i>HTR4</i>     | -4,33             | 2,56                   | 26,38  | $2,64 \times 10^{-6}$  | $7,39 \times 10^{-5}$  |
| <i>DNAH10</i>   | 2,41              | $-3,09 \times 10^{-3}$ | 24,76  | $4,85 \times 10^{-6}$  | $1,22 \times 10^{-4}$  |
| <i>LRRC34</i>   | 2,45              | -0,58                  | 24,74  | $4,88 \times 10^{-6}$  | $1,22 \times 10^{-4}$  |
| <i>CTU2</i>     | 1,59              | 3,14                   | 23,64  | $7,43 \times 10^{-6}$  | $1,73 \times 10^{-4}$  |
| <i>TAF4B</i>    | 1,65              | 2,63                   | 22,68  | $1,07 \times 10^{-5}$  | $2,38 \times 10^{-4}$  |
| <i>AKR1C1</i>   | 3,54              | 4,06                   | 20,95  | $2,12 \times 10^{-5}$  | $4,07 \times 10^{-4}$  |
| <i>PDGFB</i>    | 1,80              | 3,31                   | 18,81  | $5,00 \times 10^{-5}$  | $8,10 \times 10^{-4}$  |
| <i>IQUB</i>     | 3,14              | -3,05                  | 18,26  | $6,27 \times 10^{-5}$  | $9,67 \times 10^{-4}$  |
| <i>KIF19</i>    | -3,66             | 2,34                   | 17,24  | $9,58 \times 10^{-5}$  | $1,36 \times 10^{-3}$  |
| <i>PLSCR4</i>   | -1,62             | 5,15                   | 15,31  | $2,17 \times 10^{-4}$  | $2,63 \times 10^{-3}$  |
| <i>NAP1L2</i>   | -3,47             | 1,69                   | 15,07  | $2,40 \times 10^{-4}$  | $2,84 \times 10^{-3}$  |
| <i>SGO2</i>     | 2,01              | 2,80                   | 13,36  | $5,09 \times 10^{-4}$  | $5,16 \times 10^{-3}$  |
| <i>TULP1</i>    | 3,23              | -2,92                  | 12,68  | $6,87 \times 10^{-4}$  | $6,49 \times 10^{-3}$  |
| <i>PTP4A3</i>   | 2,21              | 3,99                   | 12,22  | $8,48 \times 10^{-4}$  | $7,59 \times 10^{-3}$  |
| <i>AZML1</i>    | 3,25              | -1,21                  | 12,16  | $8,69 \times 10^{-4}$  | $7,73 \times 10^{-3}$  |
| <i>TRPV5</i>    | 3,25              | -4,38                  | 10,37  | $1,98 \times 10^{-3}$  | $1,47 \times 10^{-2}$  |
| <i>CCDC150</i>  | 1,85              | 0,82                   | 10,17  | $2,18 \times 10^{-3}$  | $1,58 \times 10^{-2}$  |
| <i>LHX4</i>     | -1,57             | 2,98                   | 9,69   | $2,73 \times 10^{-3}$  | $1,86 \times 10^{-2}$  |
| <i>CEMIP</i>    | 3,50              | 2,70                   | 9,01   | $3,78 \times 10^{-3}$  | $2,37 \times 10^{-2}$  |

**Tabla suplementaria 51. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 401.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | logFC | logCPM | F     | p-valor               | FDR                   |
|----------------|-------|--------|-------|-----------------------|-----------------------|
| <i>ZNF350</i>  | -2,35 | 3,38   | 31,38 | $4,37 \times 10^{-7}$ | $6,45 \times 10^{-5}$ |
| <i>ZNF793</i>  | -2,72 | 2,70   | 27,18 | $1,97 \times 10^{-6}$ | $2,17 \times 10^{-4}$ |
| <i>TNFSF9</i>  | 4,87  | 0,07   | 24,68 | $4,99 \times 10^{-6}$ | $4,39 \times 10^{-4}$ |
| <i>MROH7</i>   | -3,22 | 4,34   | 17,70 | $7,90 \times 10^{-5}$ | $3,27 \times 10^{-3}$ |
| <i>OR2W3</i>   | -8,74 | -0,46  | 16,99 | $1,06 \times 10^{-4}$ | $4,03 \times 10^{-3}$ |
| <i>SCN5A</i>   | -3,41 | 0,02   | 16,33 | $1,41 \times 10^{-4}$ | $4,91 \times 10^{-3}$ |
| <i>ERBIN</i>   | -1,75 | 6,61   | 14,47 | $3,11 \times 10^{-4}$ | $8,61 \times 10^{-3}$ |
| <i>ANK1</i>    | 2,93  | -0,24  | 13,25 | $5,33 \times 10^{-4}$ | $1,24 \times 10^{-2}$ |
| <i>SPATA46</i> | -6,17 | -1,75  | 13,19 | $5,47 \times 10^{-4}$ | $1,26 \times 10^{-2}$ |
| <i>SPON2</i>   | 1,52  | 5,32   | 11,98 | $9,45 \times 10^{-4}$ | $1,83 \times 10^{-2}$ |
| <i>SLCO4A1</i> | 1,93  | 4,19   | 10,79 | $1,63 \times 10^{-3}$ | $2,64 \times 10^{-2}$ |
| <i>WDR4</i>    | 1,74  | 2,93   | 10,66 | $1,73 \times 10^{-3}$ | $2,76 \times 10^{-2}$ |
| <i>IGSF9</i>   | -2,79 | 6,10   | 9,07  | $3,66 \times 10^{-3}$ | $4,52 \times 10^{-2}$ |
| <i>PKD1L1</i>  | -1,59 | 0,39   | 8,97  | $3,85 \times 10^{-3}$ | $4,62 \times 10^{-2}$ |
| <i>TTL6</i>    | -2,49 | 4,11   | 8,91  | $3,96 \times 10^{-3}$ | $4,69 \times 10^{-2}$ |
| <i>VWA5A</i>   | -1,65 | 6,36   | 8,75  | $4,29 \times 10^{-3}$ | $4,90 \times 10^{-2}$ |

**Tabla suplementaria 52. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente READ-1.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón;  $F$ : estadístico  $F$  del  $F$  test de cuasiverosimilitud;  $p$ -valor:  $p$ -valor nominal derivado del valor  $F$  sin ningún tipo de corrección múltiple;  $\text{FDR}$ :  $p$ -valor con corrección Benjamin-Hochberg.

|                 | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | $F$    | $p$ -valor             | $\text{FDR}$           |
|-----------------|-------------------|--------------------|--------|------------------------|------------------------|
| <i>ASB4</i>     | 8,31              | -0,25              | 437,51 | $5,62 \times 10^{-31}$ | $9,98 \times 10^{-28}$ |
| <i>IQGAP1</i>   | 3,05              | 8,32               | 324,07 | $2,79 \times 10^{-27}$ | $2,48 \times 10^{-24}$ |
| <i>KIAA1549</i> | 3,29              | 2,70               | 169,47 | $5,74 \times 10^{-20}$ | $1,35 \times 10^{-17}$ |
| <i>COPA</i>     | 1,62              | 7,58               | 143,30 | $2,94 \times 10^{-18}$ | $4,75 \times 10^{-16}$ |
| <i>HABP2</i>    | 5,56              | -1,70              | 124,39 | $6,98 \times 10^{-17}$ | $8,36 \times 10^{-15}$ |
| <i>PCNX2</i>    | 2,62              | 3,44               | 112,93 | $5,57 \times 10^{-16}$ | $5,74 \times 10^{-14}$ |
| <i>SPRTN</i>    | 2,18              | 3,21               | 86,47  | $1,18 \times 10^{-13}$ | $7,52 \times 10^{-12}$ |
| <i>DNAH2</i>    | 3,40              | 0,10               | 84,04  | $2,03 \times 10^{-13}$ | $1,23 \times 10^{-11}$ |
| <i>TBCA</i>     | 1,61              | 5,41               | 73,65  | $2,26 \times 10^{-12}$ | $1,02 \times 10^{-10}$ |
| <i>HDAC2</i>    | 1,58              | 6,10               | 61,97  | $4,25 \times 10^{-11}$ | $1,43 \times 10^{-9}$  |
| <i>FGF18</i>    | 4,51              | -0,95              | 60,87  | $5,68 \times 10^{-11}$ | $1,85 \times 10^{-9}$  |
| <i>MUC4</i>     | -8,40             | 8,47               | 58,03  | $1,22 \times 10^{-10}$ | $3,68 \times 10^{-9}$  |
| <i>KLC3</i>     | 2,68              | -1,22              | 46,09  | $3,70 \times 10^{-9}$  | $7,93 \times 10^{-8}$  |
| <i>RHBDL2</i>   | -4,18             | 4,54               | 41,54  | $1,50 \times 10^{-8}$  | $2,74 \times 10^{-7}$  |
| <i>RILP</i>     | -4,01             | 4,74               | 36,18  | $8,51 \times 10^{-8}$  | $1,28 \times 10^{-6}$  |
| <i>MAMDC4</i>   | -4,00             | 5,47               | 35,49  | $1,07 \times 10^{-7}$  | $1,56 \times 10^{-6}$  |
| <i>IL6R</i>     | -2,57             | 5,95               | 34,99  | $1,27 \times 10^{-7}$  | $1,82 \times 10^{-6}$  |
| <i>POLR1B</i>   | 1,89              | 4,24               | 34,49  | $1,50 \times 10^{-7}$  | $2,11 \times 10^{-6}$  |
| <i>TAF1B</i>    | 1,69              | 2,34               | 29,95  | $7,24 \times 10^{-7}$  | $8,51 \times 10^{-6}$  |
| <i>GALM</i>     | -2,37             | 5,92               | 29,92  | $7,30 \times 10^{-7}$  | $8,57 \times 10^{-6}$  |
| <i>UGT1A6</i>   | -6,68             | 6,61               | 27,60  | $1,69 \times 10^{-6}$  | $1,79 \times 10^{-5}$  |
| <i>CCDC17</i>   | -4,29             | 1,51               | 27,17  | $1,97 \times 10^{-6}$  | $2,05 \times 10^{-5}$  |
| <i>SLC9A9</i>   | -3,24             | 3,14               | 27,12  | $2,01 \times 10^{-6}$  | $2,08 \times 10^{-5}$  |
| <i>IL15</i>     | -2,67             | 2,87               | 25,11  | $4,25 \times 10^{-6}$  | $3,99 \times 10^{-5}$  |
| <i>SEC23B</i>   | 1,55              | 5,86               | 24,31  | $5,74 \times 10^{-6}$  | $5,21 \times 10^{-5}$  |
| <i>GCG</i>      | -12,51            | 5,77               | 23,34  | $8,33 \times 10^{-6}$  | $7,20 \times 10^{-5}$  |
| <i>IGSF10</i>   | 2,87              | 2,26               | 22,71  | $1,06 \times 10^{-5}$  | $8,80 \times 10^{-5}$  |
| <i>NEXMIF</i>   | -6,00             | 1,43               | 22,31  | $1,24 \times 10^{-5}$  | $1,00 \times 10^{-4}$  |
| <i>WNT10B</i>   | -4,17             | 0,42               | 22,18  | $1,31 \times 10^{-5}$  | $1,05 \times 10^{-4}$  |
| <i>PCNA</i>     | 1,71              | 5,96               | 21,91  | $1,45 \times 10^{-5}$  | $1,15 \times 10^{-4}$  |
| <i>FAM135B</i>  | -5,92             | 1,44               | 21,18  | $1,94 \times 10^{-5}$  | $1,48 \times 10^{-4}$  |
| <i>BRCA1</i>    | 2,02              | 3,43               | 20,72  | $2,32 \times 10^{-5}$  | $1,73 \times 10^{-4}$  |
| <i>SLAMF9</i>   | 4,11              | -3,91              | 20,56  | $2,47 \times 10^{-5}$  | $1,82 \times 10^{-4}$  |
| <i>IL12RB2</i>  | -3,94             | -0,40              | 19,09  | $4,46 \times 10^{-5}$  | $3,02 \times 10^{-4}$  |
| <i>TRPM6</i>    | -7,82             | 6,57               | 18,40  | $5,93 \times 10^{-5}$  | $3,86 \times 10^{-4}$  |
| <i>KCNAB3</i>   | -2,64             | 1,87               | 17,81  | $7,57 \times 10^{-5}$  | $4,77 \times 10^{-4}$  |
| <i>U2AF1L4</i>  | -1,83             | 3,62               | 17,15  | $9,95 \times 10^{-5}$  | $6,01 \times 10^{-4}$  |
| <i>FOXA2</i>    | 2,07              | 4,10               | 17,10  | $1,02 \times 10^{-4}$  | $6,12 \times 10^{-4}$  |
| <i>MYSM1</i>    | -1,77             | 6,08               | 16,20  | $1,48 \times 10^{-4}$  | $8,43 \times 10^{-4}$  |
| <i>KSR1</i>     | -1,99             | 4,74               | 16,16  | $1,51 \times 10^{-4}$  | $8,56 \times 10^{-4}$  |

Continuación tabla suplementaria 52.

|          | logFC | logCPM | F     | p-valor               | FDR                   |
|----------|-------|--------|-------|-----------------------|-----------------------|
| LAMC3    | -3,37 | 2,85   | 12,37 | 7,90x10 <sup>-4</sup> | 3,51x10 <sup>-3</sup> |
| SH3RF2   | -1,52 | 6,20   | 11,49 | 1,18x10 <sup>-3</sup> | 4,90x10 <sup>-3</sup> |
| RTN4RL1  | -3,37 | -0,45  | 11,33 | 1,27x10 <sup>-3</sup> | 5,21x10 <sup>-3</sup> |
| RAD51AP2 | 2,69  | -2,24  | 11,23 | 1,33x10 <sup>-3</sup> | 5,40x10 <sup>-3</sup> |
| CCDC62   | -3,07 | -1,78  | 10,87 | 1,57x10 <sup>-3</sup> | 6,18x10 <sup>-3</sup> |
| SCN7A    | -5,52 | 4,53   | 10,76 | 1,65x10 <sup>-3</sup> | 6,46x10 <sup>-3</sup> |
| OGN      | -5,39 | 3,79   | 10,59 | 1,79x10 <sup>-3</sup> | 6,92x10 <sup>-3</sup> |
| EQTN     | -6,94 | -2,42  | 10,14 | 2,21x10 <sup>-3</sup> | 8,27x10 <sup>-3</sup> |
| DCLK1    | -3,48 | 2,87   | 9,50  | 2,98x10 <sup>-3</sup> | 1,07x10 <sup>-2</sup> |
| PTCH2    | -2,64 | 1,58   | 9,40  | 3,14x10 <sup>-3</sup> | 1,11x10 <sup>-2</sup> |
| CYP46A1  | -2,94 | -0,69  | 9,18  | 3,48x10 <sup>-3</sup> | 1,21x10 <sup>-2</sup> |
| RMDN2    | -1,62 | 4,02   | 8,81  | 4,16x10 <sup>-3</sup> | 1,41x10 <sup>-2</sup> |
| SGO2     | 1,61  | 2,70   | 8,49  | 4,87x10 <sup>-3</sup> | 1,60x10 <sup>-2</sup> |
| CYP4B1   | -9,49 | 3,61   | 8,30  | 5,33x10 <sup>-3</sup> | 1,73x10 <sup>-2</sup> |
| COL9A2   | -2,86 | 4,04   | 7,96  | 6,30x10 <sup>-3</sup> | 1,99x10 <sup>-2</sup> |
| ABCD2    | -2,51 | 0,05   | 7,92  | 6,41x10 <sup>-3</sup> | 2,01x10 <sup>-2</sup> |
| ZNF667   | -2,78 | 1,22   | 7,48  | 8,00x10 <sup>-3</sup> | 2,42x10 <sup>-2</sup> |
| ANKRD36  | -1,59 | 2,31   | 7,44  | 8,16x10 <sup>-3</sup> | 2,46x10 <sup>-2</sup> |
| SLC44A3  | -1,52 | 5,70   | 7,43  | 8,19x10 <sup>-3</sup> | 2,46x10 <sup>-2</sup> |
| KCNB1    | -5,49 | 2,66   | 7,37  | 8,43x10 <sup>-3</sup> | 2,52x10 <sup>-2</sup> |
| FAT2     | -1,96 | 0,69   | 7,35  | 8,52x10 <sup>-3</sup> | 2,55x10 <sup>-2</sup> |
| TEX14    | -1,59 | -0,89  | 7,29  | 8,79x10 <sup>-3</sup> | 2,61x10 <sup>-2</sup> |
| RTP5     | -7,43 | -2,18  | 7,28  | 8,84x10 <sup>-3</sup> | 2,62x10 <sup>-2</sup> |
| FLG2     | -8,28 | -1,33  | 6,66  | 1,21x10 <sup>-2</sup> | 3,36x10 <sup>-2</sup> |
| SYNE1    | -2,23 | 5,58   | 6,64  | 1,22x10 <sup>-2</sup> | 3,39x10 <sup>-2</sup> |
| SARDH    | -2,07 | 0,50   | 6,35  | 1,41x10 <sup>-2</sup> | 3,82x10 <sup>-2</sup> |
| PER3     | -3,32 | 4,47   | 5,99  | 1,70x10 <sup>-2</sup> | 4,43x10 <sup>-2</sup> |
| PTCD     | -3,90 | -1,07  | 5,94  | 1,75x10 <sup>-2</sup> | 4,53x10 <sup>-2</sup> |
| DNAH6    | 1,59  | 1,16   | 5,92  | 1,77x10 <sup>-2</sup> | 4,56x10 <sup>-2</sup> |
| SLC47A2  | -3,23 | -1,48  | 5,77  | 1,91x10 <sup>-2</sup> | 4,85x10 <sup>-2</sup> |
| ADRA1A   | -9,02 | -0,65  | 15,66 | 1,86x10 <sup>-4</sup> | 1,03x10 <sup>-3</sup> |
| PNMA3    | -5,91 | 9,70   | 14,98 | 2,50x10 <sup>-4</sup> | 1,32x10 <sup>-3</sup> |
| SPECC1   | -2,15 | 5,94   | 14,52 | 3,06x10 <sup>-4</sup> | 1,57x10 <sup>-3</sup> |
| PLIN5    | -3,50 | 1,98   | 14,29 | 3,37x10 <sup>-4</sup> | 1,71x10 <sup>-3</sup> |
| LRRN2    | -4,47 | 4,71   | 14,24 | 3,44x10 <sup>-4</sup> | 1,74x10 <sup>-3</sup> |
| USP9Y    | -4,13 | 4,03   | 14,21 | 3,49x10 <sup>-4</sup> | 1,76x10 <sup>-3</sup> |
| CDH26    | -2,47 | -0,18  | 14,04 | 3,76x10 <sup>-4</sup> | 1,87x10 <sup>-3</sup> |
| FAM13A   | -1,85 | 5,86   | 13,92 | 3,96x10 <sup>-4</sup> | 1,96x10 <sup>-3</sup> |
| ADGB     | -4,42 | -0,13  | 13,87 | 4,06x10 <sup>-4</sup> | 2,00x10 <sup>-3</sup> |
| SCIN     | -3,30 | 5,84   | 12,64 | 7,02x10 <sup>-4</sup> | 3,18x10 <sup>-3</sup> |
| RIC3     | -5,07 | 2,56   | 12,56 | 7,25x10 <sup>-4</sup> | 3,27x10 <sup>-3</sup> |
| HJURP    | 2,10  | 3,00   | 12,54 | 7,31x10 <sup>-4</sup> | 3,29x10 <sup>-3</sup> |



**Tabla suplementaria 53. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente 303.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                 | logFC | logCPM | F     | p-valor                | FDR                    |
|-----------------|-------|--------|-------|------------------------|------------------------|
| <i>ASAP1</i>    | 2,30  | 4,48   | 74,71 | $1,76 \times 10^{-12}$ | $4,06 \times 10^{-10}$ |
| <i>TRIM54</i>   | 4,80  | -1,00  | 70,68 | $4,65 \times 10^{-12}$ | $9,80 \times 10^{-10}$ |
| <i>RELL2</i>    | 2,09  | 2,65   | 49,76 | $1,25 \times 10^{-9}$  | $1,41 \times 10^{-7}$  |
| <i>PLEKHN1</i>  | 2,76  | 0,20   | 47,03 | $2,79 \times 10^{-9}$  | $2,91 \times 10^{-7}$  |
| <i>ABCC1</i>    | 1,87  | 5,14   | 44,09 | $6,81 \times 10^{-9}$  | $6,36 \times 10^{-7}$  |
| <i>KCNJ15</i>   | 5,04  | -1,42  | 38,11 | $4,52 \times 10^{-8}$  | $3,30 \times 10^{-6}$  |
| <i>PLEKHA4</i>  | 2,90  | 3,67   | 25,71 | $3,39 \times 10^{-6}$  | $1,31 \times 10^{-4}$  |
| <i>COL5A3</i>   | 2,12  | 4,76   | 25,42 | $3,77 \times 10^{-6}$  | $1,43 \times 10^{-4}$  |
| <i>CCN4</i>     | 3,97  | 0,90   | 22,77 | $1,04 \times 10^{-5}$  | $3,27 \times 10^{-4}$  |
| <i>SERPINF2</i> | 2,36  | 1,75   | 21,09 | $2,00 \times 10^{-5}$  | $5,54 \times 10^{-4}$  |
| <i>DNAH10</i>   | 2,29  | -0,01  | 20,98 | $2,09 \times 10^{-5}$  | $5,76 \times 10^{-4}$  |
| <i>ERAP1</i>    | -1,73 | 6,81   | 19,80 | $3,36 \times 10^{-5}$  | $8,24 \times 10^{-4}$  |
| <i>CA9</i>      | 3,30  | -0,57  | 18,74 | $5,16 \times 10^{-5}$  | $1,18 \times 10^{-3}$  |
| <i>CDH16</i>    | 3,84  | 0,10   | 18,50 | $5,69 \times 10^{-5}$  | $1,28 \times 10^{-3}$  |
| <i>ITGBL1</i>   | 3,28  | -0,71  | 15,69 | $1,84 \times 10^{-4}$  | $3,29 \times 10^{-3}$  |
| <i>MTFR2</i>    | 2,08  | 1,34   | 13,90 | $4,01 \times 10^{-4}$  | $5,95 \times 10^{-3}$  |
| <i>LEF1</i>     | 2,17  | 2,03   | 13,46 | $4,86 \times 10^{-4}$  | $6,90 \times 10^{-3}$  |
| <i>ELOVL3</i>   | 3,44  | -3,05  | 10,57 | $1,80 \times 10^{-3}$  | $1,81 \times 10^{-2}$  |
| <i>POLQ</i>     | 2,02  | 2,99   | 10,47 | $1,89 \times 10^{-3}$  | $1,87 \times 10^{-2}$  |
| <i>SELP</i>     | 1,71  | 2,16   | 10,45 | $1,91 \times 10^{-3}$  | $1,88 \times 10^{-2}$  |
| <i>SCIN</i>     | -2,83 | 5,84   | 9,95  | $2,41 \times 10^{-3}$  | $2,21 \times 10^{-2}$  |
| <i>BICD1</i>    | 1,62  | 2,15   | 9,88  | $2,50 \times 10^{-3}$  | $2,27 \times 10^{-2}$  |
| <i>CPXM1</i>    | 2,54  | 1,63   | 9,67  | $2,75 \times 10^{-3}$  | $2,42 \times 10^{-2}$  |
| <i>SLC6A20</i>  | 3,41  | 1,32   | 9,63  | $2,81 \times 10^{-3}$  | $2,46 \times 10^{-2}$  |
| <i>PNPLA7</i>   | -1,93 | 5,19   | 9,11  | $3,60 \times 10^{-3}$  | $2,92 \times 10^{-2}$  |
| <i>LGALS4</i>   | -1,89 | 11,43  | 8,40  | $5,08 \times 10^{-3}$  | $3,76 \times 10^{-2}$  |
| <i>MEFV</i>     | 1,89  | 0,46   | 8,20  | $5,59 \times 10^{-3}$  | $4,01 \times 10^{-2}$  |
| <i>ARHGEF4</i>  | 2,89  | 2,23   | 7,81  | $6,77 \times 10^{-3}$  | $4,58 \times 10^{-2}$  |
| <i>MYO1H</i>    | 1,78  | -2,19  | 7,64  | $7,39 \times 10^{-3}$  | $4,85 \times 10^{-2}$  |
| <i>DLC1</i>     | 1,79  | 4,65   | 7,55  | $7,70 \times 10^{-3}$  | $4,99 \times 10^{-2}$  |

**Tabla suplementaria 54. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 310.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                 | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | F      | p-valor                | FDR                    |
|-----------------|-------------------|--------------------|--------|------------------------|------------------------|
| <i>TMEM132C</i> | 6,41              | 1,71               | 272,60 | $3,14 \times 10^{-25}$ | $1,34 \times 10^{-21}$ |
| <i>ASPCR1</i>   | 2,08              | 4,25               | 122,15 | $1,04 \times 10^{-16}$ | $4,42 \times 10^{-14}$ |
| <i>MOCS3</i>    | 1,72              | 2,82               | 74,77  | $1,73 \times 10^{-12}$ | $2,68 \times 10^{-10}$ |
| <i>PKLR</i>     | 4,86              | 0,50               | 60,84  | $5,73 \times 10^{-11}$ | $5,97 \times 10^{-9}$  |
| <i>COL18A1</i>  | 2,91              | 6,53               | 59,89  | $7,38 \times 10^{-11}$ | $7,55 \times 10^{-9}$  |
| <i>FMO1</i>     | 5,24              | -2,21              | 52,91  | $5,05 \times 10^{-10}$ | $3,98 \times 10^{-8}$  |
| <i>PCDH17</i>   | 3,60              | 1,54               | 50,61  | $9,75 \times 10^{-10}$ | $6,93 \times 10^{-8}$  |
| <i>L3MBTL1</i>  | 1,95              | 3,48               | 48,68  | $1,71 \times 10^{-9}$  | $1,14 \times 10^{-7}$  |
| <i>FAM9A</i>    | 4,97              | -4,87              | 48,73  | $2,07 \times 10^{-9}$  | $1,34 \times 10^{-7}$  |
| <i>NOTCH4</i>   | 2,36              | 3,85               | 47,66  | $2,31 \times 10^{-9}$  | $1,47 \times 10^{-7}$  |
| <i>OSM</i>      | 5,71              | -0,42              | 44,67  | $5,70 \times 10^{-9}$  | $3,19 \times 10^{-7}$  |
| <i>WFS1</i>     | 2,76              | 5,10               | 42,06  | $1,28 \times 10^{-8}$  | $6,50 \times 10^{-7}$  |
| <i>ENG</i>      | 2,01              | 6,18               | 38,61  | $3,83 \times 10^{-8}$  | $1,73 \times 10^{-6}$  |
| <i>LRCH1</i>    | 1,77              | 4,58               | 37,27  | $5,94 \times 10^{-8}$  | $2,57 \times 10^{-6}$  |
| <i>PALM3</i>    | 2,61              | 0,01               | 20,98  | $2,09 \times 10^{-5}$  | $3,56 \times 10^{-4}$  |
| <i>RP1L1</i>    | 3,95              | -2,65              | 20,32  | $2,72 \times 10^{-5}$  | $4,41 \times 10^{-4}$  |
| <i>HECW1</i>    | 2,46              | -1,27              | 18,95  | $4,74 \times 10^{-5}$  | $6,97 \times 10^{-4}$  |
| <i>NUTM1</i>    | -6,56             | -0,89              | 16,63  | $1,24 \times 10^{-4}$  | $1,53 \times 10^{-3}$  |
| <i>SKA3</i>     | 2,58              | 2,04               | 14,50  | $3,08 \times 10^{-4}$  | $3,20 \times 10^{-3}$  |
| <i>TST</i>      | -2,60             | 7,95               | 14,39  | $3,23 \times 10^{-4}$  | $3,32 \times 10^{-3}$  |
| <i>MUCL1</i>    | 5,31              | -3,67              | 13,78  | $4,35 \times 10^{-4}$  | $4,22 \times 10^{-3}$  |
| <i>CARD10</i>   | -1,54             | 5,95               | 12,57  | $7,21 \times 10^{-4}$  | $6,31 \times 10^{-3}$  |
| <i>SLC26A9</i>  | 3,71              | -3,86              | 12,43  | $7,69 \times 10^{-4}$  | $6,62 \times 10^{-3}$  |
| <i>LRIT2</i>    | -3,75             | 1,89               | 11,22  | $1,34 \times 10^{-3}$  | $1,02 \times 10^{-2}$  |
| <i>ABCA10</i>   | -2,20             | 3,03               | 10,73  | $1,67 \times 10^{-3}$  | $1,21 \times 10^{-2}$  |
| <i>MUC5B</i>    | -3,64             | 8,42               | 10,61  | $1,77 \times 10^{-3}$  | $1,26 \times 10^{-2}$  |
| <i>STAP2</i>    | -1,85             | 6,91               | 9,84   | $2,54 \times 10^{-3}$  | $1,67 \times 10^{-2}$  |
| <i>WNK4</i>     | 1,91              | 5,94               | 8,74   | $4,30 \times 10^{-3}$  | $2,48 \times 10^{-2}$  |
| <i>TG</i>       | 3,58              | 2,05               | 8,62   | $4,55 \times 10^{-3}$  | $2,59 \times 10^{-2}$  |
| <i>SDK2</i>     | -2,95             | 2,10               | 8,42   | $5,10 \times 10^{-3}$  | $2,79 \times 10^{-2}$  |
| <i>SLC4A11</i>  | 1,76              | 0,05               | 8,37   | $5,14 \times 10^{-3}$  | $2,84 \times 10^{-2}$  |
| <i>KIAA1614</i> | 1,64              | 2,58               | 7,96   | $6,31 \times 10^{-3}$  | $3,31 \times 10^{-2}$  |

**Tabla suplementaria 55. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente COAD-1.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón;  $F$ : estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | F      | p-valor                | FDR                    |
|----------------|-------------------|--------------------|--------|------------------------|------------------------|
| <i>COG6</i>    | 2,33              | 4,65               | 135,16 | $1,11 \times 10^{-17}$ | $1,49 \times 10^{-15}$ |
| <i>ESR2</i>    | -4,58             | 0,80               | 35,40  | $1,10 \times 10^{-7}$  | $1,84 \times 10^{-6}$  |
| <i>RUVBL1</i>  | 1,52              | 5,00               | 32,83  | $2,64 \times 10^{-7}$  | $3,91 \times 10^{-6}$  |
| <i>DEPDC5</i>  | -1,71             | 4,72               | 30,18  | $6,67 \times 10^{-7}$  | $8,67 \times 10^{-6}$  |
| <i>SIGLEC1</i> | -4,53             | 4,91               | 29,61  | $8,16 \times 10^{-7}$  | $1,03 \times 10^{-5}$  |
| <i>RNF207</i>  | -2,96             | 4,91               | 24,82  | $4,73 \times 10^{-6}$  | $4,80 \times 10^{-5}$  |
| <i>MAST2</i>   | -2,11             | 7,34               | 21,84  | $1,49 \times 10^{-5}$  | $1,27 \times 10^{-4}$  |
| <i>ADSS1</i>   | -2,33             | 1,74               | 21,55  | $1,67 \times 10^{-5}$  | $1,40 \times 10^{-4}$  |
| <i>TTN</i>     | -2,95             | 4,37               | 20,76  | $2,28 \times 10^{-5}$  | $1,83 \times 10^{-4}$  |
| <i>RASGRP2</i> | -4,64             | 4,30               | 19,14  | $4,37 \times 10^{-5}$  | $3,18 \times 10^{-4}$  |
| <i>APLF</i>    | 1,74              | 1,08               | 17,95  | $7,14 \times 10^{-5}$  | $4,84 \times 10^{-4}$  |
| <i>CPT2</i>    | -1,80             | 6,27               | 17,60  | $8,24 \times 10^{-5}$  | $5,48 \times 10^{-4}$  |
| <i>SCNN1D</i>  | -2,93             | 2,54               | 17,29  | $9,38 \times 10^{-5}$  | $6,12 \times 10^{-4}$  |
| <i>RBM6</i>    | -1,54             | 7,15               | 16,84  | $1,13 \times 10^{-4}$  | $7,18 \times 10^{-4}$  |
| <i>AHNAK</i>   | -2,54             | 10,53              | 15,99  | $1,62 \times 10^{-4}$  | $9,75 \times 10^{-4}$  |
| <i>CLIC6</i>   | -5,46             | 4,06               | 14,73  | $2,78 \times 10^{-4}$  | $1,54 \times 10^{-3}$  |
| <i>PYGL</i>    | -2,99             | 5,36               | 14,35  | $3,28 \times 10^{-4}$  | $1,76 \times 10^{-3}$  |
| <i>NFATC4</i>  | -2,95             | 5,29               | 13,10  | $5,71 \times 10^{-4}$  | $2,82 \times 10^{-3}$  |
| <i>STOX2</i>   | -2,42             | 2,67               | 12,71  | $6,80 \times 10^{-4}$  | $3,25 \times 10^{-3}$  |
| <i>ASIC1</i>   | 2,44              | 2,12               | 10,34  | $2,01 \times 10^{-3}$  | $8,00 \times 10^{-3}$  |
| <i>NTRK3</i>   | -3,66             | 1,63               | 10,11  | $2,24 \times 10^{-3}$  | $8,75 \times 10^{-3}$  |
| <i>RUFY4</i>   | -6,06             | -0,85              | 9,89   | $2,48 \times 10^{-3}$  | $9,51 \times 10^{-3}$  |
| <i>KIFC3</i>   | 1,71              | 3,44               | 9,38   | $3,16 \times 10^{-3}$  | $1,16 \times 10^{-2}$  |
| <i>CASP5</i>   | -4,87             | 3,75               | 9,33   | $3,23 \times 10^{-3}$  | $1,18 \times 10^{-2}$  |
| <i>CPA1</i>    | -12,06            | 2,12               | 8,22   | $5,55 \times 10^{-3}$  | $1,86 \times 10^{-2}$  |
| <i>SDK2</i>    | -2,80             | 2,10               | 7,81   | $6,79 \times 10^{-3}$  | $2,19 \times 10^{-2}$  |
| <i>ZNF540</i>  | -1,52             | 1,76               | 6,56   | $1,27 \times 10^{-2}$  | $3,64 \times 10^{-2}$  |
| <i>CNTN4</i>   | -2,15             | 3,01               | 6,34   | $1,42 \times 10^{-2}$  | $3,99 \times 10^{-2}$  |
| <i>CNGA3</i>   | -4,99             | 1,01               | 6,06   | $1,64 \times 10^{-2}$  | $4,48 \times 10^{-2}$  |
| <i>SUGCT</i>   | -2,02             | 2,54               | 6,03   | $1,67 \times 10^{-2}$  | $4,54 \times 10^{-2}$  |

**Tabla suplementaria 56. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-2.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|          | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | F      | p-valor                | FDR                    |
|----------|-------------------|--------------------|--------|------------------------|------------------------|
| PRKDC    | 3,07              | 6,85               | 212,36 | $2,15 \times 10^{-22}$ | $1,53 \times 10^{-19}$ |
| PKP4     | 1,54              | 5,76               | 94,07  | $2,32 \times 10^{-14}$ | $2,03 \times 10^{-12}$ |
| STAU2    | 1,54              | 4,68               | 85,68  | $1,41 \times 10^{-13}$ | $1,01 \times 10^{-11}$ |
| TRAM1    | 1,88              | 6,80               | 74,30  | $1,93 \times 10^{-12}$ | $1,11 \times 10^{-10}$ |
| NUP155   | 1,87              | 4,44               | 68,90  | $7,21 \times 10^{-12}$ | $3,56 \times 10^{-10}$ |
| SPAG1    | 2,12              | 5,20               | 49,64  | $1,29 \times 10^{-9}$  | $3,48 \times 10^{-8}$  |
| ABCA10   | -5,63             | 3,02               | 42,93  | $9,73 \times 10^{-9}$  | $2,03 \times 10^{-7}$  |
| ENC1     | 2,40              | 5,60               | 39,28  | $3,09 \times 10^{-8}$  | $5,58 \times 10^{-7}$  |
| ZNF518B  | -2,54             | 3,95               | 37,25  | $5,97 \times 10^{-8}$  | $9,98 \times 10^{-7}$  |
| PAN2     | -2,02             | 5,95               | 35,95  | $9,17 \times 10^{-8}$  | $1,47 \times 10^{-6}$  |
| PLTP     | 2,79              | 5,84               | 31,83  | $3,74 \times 10^{-7}$  | $5,12 \times 10^{-6}$  |
| PITPNM3  | -2,51             | 5,36               | 28,56  | $1,19 \times 10^{-6}$  | $1,39 \times 10^{-5}$  |
| TBX18    | 3,09              | -0,31              | 24,16  | $6,09 \times 10^{-6}$  | $5,88 \times 10^{-5}$  |
| SPAG8    | -2,90             | 1,66               | 21,14  | $1,97 \times 10^{-5}$  | $1,62 \times 10^{-4}$  |
| TIAM1    | -2,09             | 2,74               | 18,86  | $4,91 \times 10^{-5}$  | $3,54 \times 10^{-4}$  |
| BRCA2    | 2,40              | 2,00               | 17,94  | $7,15 \times 10^{-5}$  | $4,89 \times 10^{-4}$  |
| ADAMTS10 | -3,14             | 3,35               | 17,60  | $8,24 \times 10^{-5}$  | $5,52 \times 10^{-4}$  |
| TRPM5    | -5,00             | 2,97               | 17,35  | $9,15 \times 10^{-5}$  | $6,04 \times 10^{-4}$  |
| TTN      | -2,41             | 4,37               | 15,05  | $2,42 \times 10^{-4}$  | $1,37 \times 10^{-3}$  |
| ROBO2    | 2,55              | 0,63               | 14,76  | $2,75 \times 10^{-4}$  | $1,53 \times 10^{-3}$  |
| SLC6A4   | 3,64              | 0,65               | 13,43  | $4,92 \times 10^{-4}$  | $2,50 \times 10^{-3}$  |
| FRMD4B   | -2,07             | 4,75               | 12,10  | $8,92 \times 10^{-4}$  | $4,11 \times 10^{-3}$  |
| AGMAT    | 1,55              | 4,25               | 11,50  | $1,18 \times 10^{-3}$  | $5,16 \times 10^{-3}$  |
| SIGLEC11 | -3,81             | -0,50              | 11,37  | $1,25 \times 10^{-3}$  | $5,43 \times 10^{-3}$  |
| RGS6     | -3,79             | -0,34              | 11,35  | $1,26 \times 10^{-3}$  | $5,48 \times 10^{-3}$  |
| CACNG1   | 3,71              | -4,12              | 11,25  | $1,32 \times 10^{-3}$  | $5,70 \times 10^{-3}$  |
| TMEM17   | 1,63              | -0,15              | 11,22  | $1,33 \times 10^{-3}$  | $5,75 \times 10^{-3}$  |
| RYR1     | -2,25             | 1,47               | 10,86  | $1,58 \times 10^{-3}$  | $6,59 \times 10^{-3}$  |
| PTX3     | -4,67             | 2,30               | 10,46  | $1,90 \times 10^{-3}$  | $7,68 \times 10^{-3}$  |
| KLRC3    | -4,26             | -0,89              | 10,44  | $1,92 \times 10^{-3}$  | $7,74 \times 10^{-3}$  |
| LY9      | -3,26             | 1,96               | 10,27  | $2,08 \times 10^{-3}$  | $8,27 \times 10^{-3}$  |
| ROBO3    | -2,51             | 3,13               | 9,84   | $2,54 \times 10^{-3}$  | $9,77 \times 10^{-3}$  |
| ADAMTS15 | -3,48             | 3,15               | 9,64   | $2,80 \times 10^{-3}$  | $1,06 \times 10^{-2}$  |
| ASPM     | 1,89              | 3,50               | 9,58   | $2,87 \times 10^{-3}$  | $1,08 \times 10^{-2}$  |
| SLC9C1   | -2,84             | -1,30              | 9,24   | $3,38 \times 10^{-3}$  | $1,23 \times 10^{-2}$  |
| MUC6     | -4,13             | -0,21              | 8,93   | $3,93 \times 10^{-3}$  | $1,39 \times 10^{-2}$  |
| CCDC178  | -4,22             | -2,41              | 8,69   | $4,42 \times 10^{-3}$  | $1,53 \times 10^{-2}$  |
| VWA5B1   | 2,90              | -1,06              | 8,67   | $4,44 \times 10^{-3}$  | $1,54 \times 10^{-2}$  |
| TRPV3    | -5,32             | 3,40               | 7,33   | $8,62 \times 10^{-3}$  | $2,64 \times 10^{-2}$  |
| THY1     | 2,08              | 4,34               | 7,28   | $8,85 \times 10^{-3}$  | $2,70 \times 10^{-2}$  |
| TRIM7    | -2,15             | 2,08               | 6,94   | $1,04 \times 10^{-2}$  | $3,09 \times 10^{-2}$  |
| MUC17    | -6,12             | 5,28               | 6,79   | $1,13 \times 10^{-2}$  | $3,30 \times 10^{-2}$  |

Continuación tabla suplementaria 56.

|                 | logFC | logCPM | F    | p-valor               | FDR                   |
|-----------------|-------|--------|------|-----------------------|-----------------------|
| <i>PLEKHG4B</i> | -5,74 | 0,24   | 6,47 | 1,33x10 <sup>-2</sup> | 3,75x10 <sup>-2</sup> |
| <i>RIMS1</i>    | 1,90  | -1,83  | 5,91 | 1,78x10 <sup>-2</sup> | 4,77x10 <sup>-2</sup> |
| <i>CRB1</i>     | -6,06 | -3,01  | 5,90 | 1,78x10 <sup>-2</sup> | 4,78x10 <sup>-2</sup> |

**Tabla suplementaria 57. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-12.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                 | logFC | logCPM | F      | p-valor                | FDR                    |
|-----------------|-------|--------|--------|------------------------|------------------------|
| <i>RAE1</i>     | 2,1   | 4,62   | 174,58 | 2,80x10 <sup>-20</sup> | 1,57x10 <sup>-17</sup> |
| <i>NEBL</i>     | 3,76  | 4,28   | 107,02 | 1,71x10 <sup>-15</sup> | 2,99x10 <sup>-13</sup> |
| <i>TCF20</i>    | 1,53  | 5,62   | 93,93  | 2,39x10 <sup>-14</sup> | 3,03x10 <sup>-12</sup> |
| <i>MAGEA4</i>   | 6,64  | -4,78  | 77,63  | 1,22x10 <sup>-11</sup> | 8,27x10 <sup>-10</sup> |
| <i>MPP1</i>     | 1,80  | 4,60   | 49,61  | 1,30x10 <sup>-9</sup>  | 5,12x10 <sup>-8</sup>  |
| <i>HOXA3</i>    | -4,94 | 4,49   | 44,39  | 6,20x10 <sup>-9</sup>  | 2,00x10 <sup>-7</sup>  |
| <i>DNHD1</i>    | -2,91 | 4,67   | 42,68  | 1,05x10 <sup>-8</sup>  | 3,19x10 <sup>-7</sup>  |
| <i>SLC13A3</i>  | 3,51  | -0,46  | 40,18  | 2,31x10 <sup>-8</sup>  | 6,34x10 <sup>-7</sup>  |
| <i>EVPL</i>     | 1,57  | 5,37   | 39,67  | 2,72x10 <sup>-8</sup>  | 7,30x10 <sup>-7</sup>  |
| <i>FGD5</i>     | 1,71  | 3,82   | 32,70  | 2,76x10 <sup>-7</sup>  | 5,51x10 <sup>-6</sup>  |
| <i>ADSS1</i>    | -3,03 | 1,74   | 31,96  | 3,57x10 <sup>-7</sup>  | 6,85x10 <sup>-6</sup>  |
| <i>PKHD1L1</i>  | -7,66 | 2,45   | 30,21  | 6,60x10 <sup>-7</sup>  | 1,18x10 <sup>-5</sup>  |
| <i>HOXA9</i>    | -2,64 | 4,84   | 22,84  | 1,01x10 <sup>-5</sup>  | 1,21x10 <sup>-4</sup>  |
| <i>TAP1</i>     | 1,81  | 6,62   | 21,44  | 1,74x10 <sup>-5</sup>  | 1,94x10 <sup>-4</sup>  |
| <i>SLC46A3</i>  | -2,27 | 5,68   | 20,70  | 2,34x10 <sup>-5</sup>  | 2,49x10 <sup>-4</sup>  |
| <i>SI</i>       | -8,09 | 6,45   | 20,70  | 2,34x10 <sup>-5</sup>  | 2,49x10 <sup>-4</sup>  |
| <i>FAM83H</i>   | 1,63  | 6,27   | 18,94  | 4,74x10 <sup>-5</sup>  | 4,46x10 <sup>-4</sup>  |
| <i>NAALADL2</i> | 1,60  | 2,94   | 18,92  | 4,80x10 <sup>-5</sup>  | 4,50x10 <sup>-4</sup>  |
| <i>ISG15</i>    | 2,40  | 3,63   | 18,46  | 5,77x10 <sup>-5</sup>  | 5,24x10 <sup>-4</sup>  |
| <i>COL28A1</i>  | -4,18 | 3,29   | 16,42  | 1,35x10 <sup>-4</sup>  | 1,07x10 <sup>-3</sup>  |
| <i>ADAMTSL5</i> | -2,56 | 1,83   | 16,1   | 1,55x10 <sup>-4</sup>  | 1,19x10 <sup>-3</sup>  |
| <i>APCDD1</i>   | 2,13  | 4,65   | 16,06  | 1,57x10 <sup>-4</sup>  | 1,21x10 <sup>-3</sup>  |
| <i>SNED1</i>    | -1,63 | 4,81   | 13,68  | 4,40x10 <sup>-4</sup>  | 2,81x10 <sup>-3</sup>  |
| <i>CASQ1</i>    | -5,74 | -0,38  | 12,06  | 9,10x10 <sup>-4</sup>  | 5,07x10 <sup>-3</sup>  |
| <i>SLC14A1</i>  | -2,51 | 1,34   | 11,21  | 1,34x10 <sup>-3</sup>  | 6,98x10 <sup>-3</sup>  |
| <i>PLCD1</i>    | -2,18 | 5,86   | 10,51  | 1,86x10 <sup>-3</sup>  | 9,14x10 <sup>-3</sup>  |
| <i>NFATC4</i>   | -2,49 | 5,28   | 10,39  | 1,96x10 <sup>-3</sup>  | 9,54x10 <sup>-3</sup>  |
| <i>SLC27A6</i>  | -3,39 | -0,32  | 10,26  | 2,08x10 <sup>-3</sup>  | 1,00x10 <sup>-2</sup>  |
| <i>IL22RA1</i>  | 1,87  | 5,08   | 10,22  | 2,12x10 <sup>-3</sup>  | 1,01x10 <sup>-2</sup>  |
| <i>TTN</i>      | -1,89 | 4,37   | 10,12  | 2,23x10 <sup>-3</sup>  | 1,06x10 <sup>-2</sup>  |
| <i>PLCH2</i>    | -3,36 | 2,80   | 9,96   | 2,41x10 <sup>-3</sup>  | 1,12x10 <sup>-2</sup>  |
| <i>PIK3C2G</i>  | -7,89 | -1,36  | 9,94   | 2,43x10 <sup>-3</sup>  | 1,13x10 <sup>-2</sup>  |
| <i>GPR150</i>   | -3,54 | -0,93  | 9,34   | 3,23x10 <sup>-3</sup>  | 1,43x10 <sup>-2</sup>  |
| <i>TNXB</i>     | -3,30 | 5,49   | 9,23   | 3,40x10 <sup>-3</sup>  | 1,49x10 <sup>-2</sup>  |
| <i>STAB2</i>    | -3,54 | 1,27   | 8,99   | 3,81x10 <sup>-3</sup>  | 1,63x10 <sup>-2</sup>  |

Continuación tabla suplementaria 57.

|                  | logFC | logCPM | F    | p-valor               | FDR                   |
|------------------|-------|--------|------|-----------------------|-----------------------|
| <i>CDHR5</i>     | -2,16 | 8,76   | 8,85 | 4,09x10 <sup>-3</sup> | 1,71x10 <sup>-2</sup> |
| <i>ASPG</i>      | -5,24 | 3,00   | 8,37 | 5,15x10 <sup>-3</sup> | 2,06x10 <sup>-2</sup> |
| <i>ITGA10</i>    | -4,13 | 3,69   | 8,19 | 5,62x10 <sup>-3</sup> | 2,21x10 <sup>-2</sup> |
| <i>STPG3</i>     | -3,76 | 0,05   | 8,09 | 5,91x10 <sup>-3</sup> | 2,29x10 <sup>-2</sup> |
| <i>TEX38</i>     | -6,46 | -3,14  | 7,91 | 6,46x10 <sup>-3</sup> | 2,46x10 <sup>-2</sup> |
| <i>SASH1</i>     | -1,55 | 4,97   | 7,90 | 6,47x10 <sup>-3</sup> | 2,46x10 <sup>-2</sup> |
| <i>HMGCS2</i>    | -2,91 | 9,31   | 7,87 | 6,57x10 <sup>-3</sup> | 2,49x10 <sup>-2</sup> |
| <i>HHIPL2</i>    | -3,56 | -1,96  | 7,86 | 6,61x10 <sup>-3</sup> | 2,50x10 <sup>-2</sup> |
| <i>KRT222</i>    | -2,74 | -0,66  | 7,82 | 6,74x10 <sup>-3</sup> | 2,54x10 <sup>-2</sup> |
| <i>CATSPERG</i>  | -2,13 | 1,40   | 7,74 | 7,01x10 <sup>-3</sup> | 2,63x10 <sup>-2</sup> |
| <i>SHISA9</i>    | 2,07  | 1,06   | 7,73 | 7,07x10 <sup>-3</sup> | 2,64x10 <sup>-2</sup> |
| <i>C16orf89</i>  | -3,53 | 1,56   | 7,69 | 7,20x10 <sup>-3</sup> | 2,68x10 <sup>-2</sup> |
| <i>SPATA31E1</i> | -6,73 | -2,53  | 7,68 | 7,22x10 <sup>-3</sup> | 2,68x10 <sup>-2</sup> |
| <i>SKOR1</i>     | -2,46 | 0,01   | 6,21 | 1,52x10 <sup>-2</sup> | 4,80x10 <sup>-2</sup> |

**Tabla suplementaria 58. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-13.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamini-Hochberg.

|                | logFC  | logCPM | F      | p-valor                | FDR                    |
|----------------|--------|--------|--------|------------------------|------------------------|
| <i>NONO</i>    | 2,06   | 7,68   | 298,32 | 2,73x10 <sup>-26</sup> | 3,74x10 <sup>-23</sup> |
| <i>ZNRF3</i>   | 3,34   | 3,92   | 195,02 | 1,85x10 <sup>-21</sup> | 6,91x10 <sup>-19</sup> |
| <i>URI1</i>    | 1,64   | 5,61   | 119,94 | 1,54x10 <sup>-16</sup> | 1,68x10 <sup>-14</sup> |
| <i>HAS2</i>    | 3,76   | 2,33   | 83,84  | 2,12x10 <sup>-13</sup> | 1,13x10 <sup>-11</sup> |
| <i>MTBP</i>    | 2,38   | 1,69   | 62,69  | 3,52x10 <sup>-11</sup> | 1,06x10 <sup>-9</sup>  |
| <i>FOXRED2</i> | 2,54   | 2,89   | 61,90  | 4,33x10 <sup>-11</sup> | 1,28x10 <sup>-9</sup>  |
| <i>WDR75</i>   | 1,69   | 4,99   | 59,36  | 8,50x10 <sup>-11</sup> | 2,35x10 <sup>-9</sup>  |
| <i>MACC1</i>   | 3,24   | 3,70   | 59,33  | 8,59x10 <sup>-11</sup> | 2,37x10 <sup>-9</sup>  |
| <i>NCKAP5</i>  | 2,96   | 1,79   | 52,22  | 6,13x10 <sup>-1</sup>  | 1,39x10 <sup>-8</sup>  |
| <i>CAD</i>     | 1,58   | 5,36   | 48,83  | 1,64x10 <sup>-9</sup>  | 3,39x10 <sup>-8</sup>  |
| <i>TRIM28</i>  | 1,51   | 7,45   | 46,23  | 3,55x10 <sup>-9</sup>  | 6,72x10 <sup>-8</sup>  |
| <i>STAB1</i>   | -4,47  | 7,47   | 38,94  | 3,44x10 <sup>-8</sup>  | 4,96x10 <sup>-7</sup>  |
| <i>MAMDC4</i>  | -4,13  | 5,47   | 36,70  | 7,17x10 <sup>-8</sup>  | 9,54x10 <sup>-7</sup>  |
| <i>ATAD2</i>   | 2,17   | 4,11   | 36,23  | 8,36x10 <sup>-8</sup>  | 1,09x10 <sup>-6</sup>  |
| <i>PKHD1L1</i> | -12,05 | 2,44   | 36,23  | 8,38x10 <sup>-8</sup>  | 1,09x10 <sup>-6</sup>  |
| <i>SLC8B1</i>  | -2,24  | 5,68   | 32,94  | 2,54x10 <sup>-7</sup>  | 2,92x10 <sup>-6</sup>  |
| <i>AMT</i>     | -2,21  | 5,28   | 30,36  | 6,25x10 <sup>-7</sup>  | 6,50x10 <sup>-6</sup>  |
| <i>QPCT</i>    | 2,79   | 3,42   | 26,94  | 2,15x10 <sup>-6</sup>  | 1,93x10 <sup>-5</sup>  |
| <i>AHNAK</i>   | -3,55  | 10,54  | 23,46  | 2,57x10 <sup>-6</sup>  | 2,24x10 <sup>-5</sup>  |
| <i>BYSL</i>    | 1,93   | 3,29   | 25,14  | 4,19x10 <sup>-6</sup>  | 3,44x10 <sup>-5</sup>  |
| <i>SPZ1</i>    | 4,60   | -4,61  | 23,14  | 8,98x10 <sup>-6</sup>  | 6,72x10 <sup>-5</sup>  |
| <i>FAM214A</i> | -1,82  | 5,26   | 22,97  | 9,59x10 <sup>-6</sup>  | 7,09x10 <sup>-5</sup>  |
| <i>HOGA1</i>   | 3,28   | -2,05  | 22,92  | 9,79x10 <sup>-6</sup>  | 7,21x10 <sup>-5</sup>  |
| <i>DNAH14</i>  | 1,69   | 1,56   | 22,86  | 1,00x10 <sup>-4</sup>  | 7,37x10 <sup>-5</sup>  |

Continuación tabla suplementaria 58.

|          | logFC | logCPM | F     | p-valor               | FDR                   |
|----------|-------|--------|-------|-----------------------|-----------------------|
| KLHL23   | 2,33  | 4,41   | 22,81 | 1,02x10 <sup>-5</sup> | 7,48x10 <sup>-5</sup> |
| KIAA1109 | -1,55 | 6,89   | 22,70 | 1,07x10 <sup>-5</sup> | 7,77x10 <sup>-5</sup> |
| TET1     | 2,54  | 0,10   | 22,28 | 1,25x10 <sup>-5</sup> | 8,94x10 <sup>-5</sup> |
| ANLN     | 2,80  | 3,49   | 22,05 | 1,37x10 <sup>-5</sup> | 9,66x10 <sup>-5</sup> |
| MIB2     | -1,94 | 5,76   | 22,02 | 1,39x10 <sup>-5</sup> | 9,76x10 <sup>-5</sup> |
| NUDT6    | -1,56 | 2,61   | 21,54 | 1,68x10 <sup>-5</sup> | 1,15x10 <sup>-4</sup> |
| DCST1    | -2,92 | 0,29   | 2,86  | 2,20x10 <sup>-5</sup> | 1,45x10 <sup>-4</sup> |
| EPHB6    | -3,75 | 2,78   | 20,09 | 2,98x10 <sup>-5</sup> | 1,89x10 <sup>-4</sup> |
| NOD2     | 2,37  | 1,16   | 18,84 | 4,95x10 <sup>-5</sup> | 2,93x10 <sup>-4</sup> |
| ARHGEF37 | -3,29 | 4,92   | 17,29 | 9,36x10 <sup>-5</sup> | 5,07x10 <sup>-4</sup> |
| KLF9     | -3,13 | 6,29   | 15,91 | 1,68x10 <sup>-4</sup> | 8,43x10 <sup>-4</sup> |
| FSTL4    | 2,19  | -1,08  | 15,35 | 2,13x10 <sup>-4</sup> | 1,03x10 <sup>-3</sup> |
| DGKQ     | -1,71 | 6,44   | 14,94 | 2,55x10 <sup>-4</sup> | 1,21x10 <sup>-3</sup> |
| MYO1H    | 2,24  | -2,13  | 14,81 | 2,70x10 <sup>-4</sup> | 1,27x10 <sup>-3</sup> |
| CECR2    | -3,91 | 0,81   | 14,15 | 3,59x10 <sup>-4</sup> | 1,61x10 <sup>-3</sup> |
| CSMD1    | -7,34 | -1,51  | 13,93 | 3,95x10 <sup>-4</sup> | 1,75x10 <sup>-3</sup> |
| HTR3E    | -9,46 | 0,38   | 12,92 | 6,19x10 <sup>-4</sup> | 2,58x10 <sup>-3</sup> |
| FAM166A  | 1,64  | 3,27   | 12,84 | 6,41x10 <sup>-4</sup> | 2,66x10 <sup>-3</sup> |
| SERPINH1 | 1,92  | 6,09   | 12,16 | 8,68x10 <sup>-4</sup> | 3,46x10 <sup>-3</sup> |
| ASXL3    | -3,32 | 2,32   | 11,83 | 1,01x10 <sup>-3</sup> | 3,93x10 <sup>-3</sup> |
| EXOC3L4  | -2,61 | 3,21   | 11,77 | 1,04x10 <sup>-3</sup> | 4,02x10 <sup>-3</sup> |
| ZNF415   | -2,07 | 2,46   | 11,51 | 1,17x10 <sup>-3</sup> | 4,47x10 <sup>-3</sup> |
| GFI1B    | -3,12 | 0,08   | 10,83 | 1,60x10 <sup>-3</sup> | 5,81x10 <sup>-3</sup> |
| ITGA10   | -5,07 | 3,69   | 10,74 | 1,66x10 <sup>-3</sup> | 6,02x10 <sup>-3</sup> |
| PTPRM    | -1,95 | 5,01   | 10,57 | 1,80x10 <sup>-3</sup> | 6,44x10 <sup>-3</sup> |
| NLRP7    | -7,58 | -1,36  | 10,44 | 1,92x10 <sup>-3</sup> | 6,80x10 <sup>-3</sup> |
| FOXN1    | -7,88 | -0,88  | 9,35  | 3,21x10 <sup>-3</sup> | 1,05x10 <sup>-2</sup> |
| JPH3     | -4,85 | -0,04  | 9,32  | 3,25x10 <sup>-3</sup> | 1,06x10 <sup>-2</sup> |
| CIDEC    | -6,48 | 3,49   | 9,25  | 3,36x10 <sup>-3</sup> | 1,09x10 <sup>-2</sup> |
| DUSP26   | -3,39 | 0,67   | 9,15  | 3,53x10 <sup>-3</sup> | 1,14x10 <sup>-2</sup> |
| ATG9B    | 1,57  | 1,33   | 8,35  | 5,20x10 <sup>-3</sup> | 1,57x10 <sup>-2</sup> |
| VEPH1    | -3,25 | 0,43   | 8,13  | 5,79x10 <sup>-3</sup> | 1,72x10 <sup>-2</sup> |
| HYDIN    | -4,56 | 0,09   | 7,99  | 6,21x10 <sup>-3</sup> | 1,83x10 <sup>-2</sup> |
| KDM4D    | 1,69  | -1,35  | 6,65  | 1,21x10 <sup>-2</sup> | 3,18x10 <sup>-2</sup> |
| CHODL    | 1,59  | 0,23   | 6,54  | 1,29x10 <sup>-2</sup> | 3,33x10 <sup>-2</sup> |
| CMYA5    | -2,59 | 1,97   | 6,36  | 1,41x10 <sup>-2</sup> | 3,58x10 <sup>-2</sup> |
| SYTL2    | -1,59 | 7,55   | 6,24  | 1,49x10 <sup>-2</sup> | 3,76x10 <sup>-2</sup> |
| GPRASP1  | -2,85 | 4,42   | 6,13  | 1,58x10 <sup>-2</sup> | 3,94x10 <sup>-2</sup> |
| GPR78    | 1,81  | 1,86   | 5,96  | 1,73x10 <sup>-2</sup> | 4,25x10 <sup>-2</sup> |
| CST2     | 3,89  | -3,46  | 5,94  | 1,75x10 <sup>-2</sup> | 4,29x10 <sup>-2</sup> |
| ANO2     | -1,82 | 0,23   | 5,93  | 1,75x10 <sup>-2</sup> | 4,30x10 <sup>-2</sup> |

**Tabla suplementaria 59. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-17. Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.**

|          | logFC | logCPM | F      | p-valor                | FDR                    |
|----------|-------|--------|--------|------------------------|------------------------|
| NOTUM    | 7,40  | -1,17  | 186,84 | $5,33 \times 10^{-21}$ | $4,37 \times 10^{-18}$ |
| SRPK1    | 2,28  | 6,57   | 74,90  | $1,67 \times 10^{-12}$ | $1,52 \times 10^{-1}$  |
| TMEM97   | 2,19  | 4,51   | 39,62  | $2,76 \times 10^{-8}$  | $7,28 \times 10^{-7}$  |
| LIFR     | -5,60 | 4,88   | 35,48  | $1,08 \times 10^{-7}$  | $2,39 \times 10^{-6}$  |
| MCM3     | 1,83  | 6,00   | 34,90  | $1,31 \times 10^{-7}$  | $2,84 \times 10^{-6}$  |
| VWA2     | 4,09  | 0,92   | 31,52  | $4,17 \times 10^{-7}$  | $7,67 \times 10^{-6}$  |
| GREB1L   | 3,47  | -0,44  | 30,22  | $6,57 \times 10^{-7}$  | $1,15 \times 10^{-5}$  |
| ATP11A   | 1,60  | 4,40   | 29,27  | $9,21 \times 10^{-7}$  | $1,55 \times 10^{-5}$  |
| DZANK1   | -2,15 | 1,50   | 28,67  | $1,14 \times 10^{-6}$  | $1,85 \times 10^{-5}$  |
| UVSSA    | -2,67 | 4,71   | 28,03  | $1,44 \times 10^{-6}$  | $2,23 \times 10^{-5}$  |
| CDT1     | 2,45  | 3,45   | 23,72  | $7,20 \times 10^{-6}$  | $8,81 \times 10^{-5}$  |
| ATOH8    | -3,10 | 3,43   | 19,83  | $3,31 \times 10^{-5}$  | $3,22 \times 10^{-4}$  |
| TIAM1    | -1,87 | 2,74   | 15,66  | $1,86 \times 10^{-4}$  | $1,35 \times 10^{-3}$  |
| BEX1     | -5,10 | 0,40   | 12,91  | $6,21 \times 10^{-4}$  | $3,65 \times 10^{-3}$  |
| SLC17A1  | 2,56  | -2,72  | 12,38  | $7,89 \times 10^{-4}$  | $4,44 \times 10^{-3}$  |
| ANK2     | -4,92 | 4,44   | 12,17  | $8,64 \times 10^{-4}$  | $4,79 \times 10^{-3}$  |
| CAMK4    | -2,06 | 2,63   | 12,06  | $9,12 \times 10^{-4}$  | $5,00 \times 10^{-3}$  |
| CDHR5    | -2,47 | 8,75   | 11,03  | $1,46 \times 10^{-3}$  | $7,30 \times 10^{-3}$  |
| RGL3     | -2,26 | 3,43   | 10,01  | $2,34 \times 10^{-3}$  | $1,08 \times 10^{-2}$  |
| COL12A1  | 2,54  | 6,04   | 9,51   | $2,97 \times 10^{-3}$  | $1,31 \times 10^{-2}$  |
| LETM1    | -1,64 | 6,90   | 9,18   | $3,48 \times 10^{-3}$  | $1,49 \times 10^{-2}$  |
| COL11A2  | -2,02 | 1,01   | 8,94   | $3,90 \times 10^{-3}$  | $1,63 \times 10^{-2}$  |
| CYP3A7   | -4,89 | -0,94  | 8,10   | $5,88 \times 10^{-3}$  | $2,25 \times 10^{-2}$  |
| HKDC1    | 2,30  | 4,60   | 7,90   | $6,47 \times 10^{-3}$  | $2,44 \times 10^{-2}$  |
| CPZ      | 2,01  | 1,92   | 7,67   | $7,28 \times 10^{-3}$  | $2,68 \times 10^{-2}$  |
| PROM1    | -2,80 | 6,46   | 7,59   | $7,57 \times 10^{-3}$  | $2,76 \times 10^{-2}$  |
| NOX5     | -3,13 | -1,50  | 7,58   | $7,59 \times 10^{-3}$  | $2,77 \times 10^{-2}$  |
| HPDL     | 1,56  | 3,06   | 6,75   | $1,16 \times 10^{-2}$  | $3,83 \times 10^{-2}$  |
| TRDN     | -7,76 | -1,99  | 6,65   | $1,21 \times 10^{-2}$  | $3,99 \times 10^{-2}$  |
| HEPACAM2 | -2,34 | 6,64   | 6,52   | $1,30 \times 10^{-2}$  | $4,20 \times 10^{-2}$  |
| STRC     | -2,05 | -1,17  | 6,47   | $1,33 \times 10^{-2}$  | $4,28 \times 10^{-2}$  |



**Tabla suplementaria 60. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente READ-5.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón;  $F$ : estadístico  $F$  del  $F$  test de cuasiverosimilitud;  $p$ -valor:  $p$ -valor nominal derivado del valor  $F$  sin ningún tipo de corrección múltiple;  $\text{FDR}$ :  $p$ -valor con corrección Benjamin-Hochberg.

|                  | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | $F$    | $p$ -valor             | $\text{FDR}$           |
|------------------|-------------------|--------------------|--------|------------------------|------------------------|
| <i>SNTB1</i>     | 2,98              | 4,01               | 223,44 | $5,88 \times 10^{-23}$ | $3,57 \times 10^{-20}$ |
| <i>KARS1</i>     | 1,78              | 6,66               | 99,12  | $8,21 \times 10^{-15}$ | $7,97 \times 10^{-13}$ |
| <i>HM13</i>      | 2,35              | 6,23               | 92,08  | $3,53 \times 10^{-14}$ | $2,94 \times 10^{-12}$ |
| <i>LZTR1</i>     | -2,06             | 6,20               | 74,11  | $2,02 \times 10^{-12}$ | $1,14 \times 10^{-10}$ |
| <i>MXRA5</i>     | 3,38              | 4,36               | 63,90  | $2,57 \times 10^{-11}$ | $1,15 \times 10^{-9}$  |
| <i>ABHD2</i>     | 2,14              | 6,92               | 61,99  | $4,23 \times 10^{-11}$ | $1,82 \times 10^{-9}$  |
| <i>TMEM97</i>    | 2,37              | 4,53               | 50,69  | $9,53 \times 10^{-10}$ | $2,82 \times 10^{-8}$  |
| <i>TTLL3</i>     | -2,31             | 6,04               | 42,57  | $1,09 \times 10^{-8}$  | $2,46 \times 10^{-7}$  |
| <i>STAB1</i>     | -4,61             | 7,47               | 40,76  | $1,92 \times 10^{-8}$  | $4,5 \times 10^{-7}$   |
| <i>CCDC157</i>   | -1,88             | 2,00               | 36,64  | $7,32 \times 10^{-8}$  | $1,32 \times 10^{-6}$  |
| <i>CADM3</i>     | -4,90             | 4,19               | 28,22  | $1,35 \times 10^{-6}$  | $1,67 \times 10^{-5}$  |
| <i>ADK</i>       | 1,75              | 3,75               | 26,05  | $2,98 \times 10^{-6}$  | $3,32 \times 10^{-5}$  |
| <i>RASGRP2</i>   | -4,46             | 4,30               | 18,1   | $6,70 \times 10^{-5}$  | $4,86 \times 10^{-4}$  |
| <i>FAT2</i>      | -3,29             | 0,68               | 16,40  | $1,36 \times 10^{-4}$  | $8,88 \times 10^{-4}$  |
| <i>LAMC3</i>     | -3,87             | 2,85               | 15,28  | $2,19 \times 10^{-4}$  | $1,32 \times 10^{-3}$  |
| <i>KIAA1549L</i> | 4,31              | -0,09              | 15,15  | $2,32 \times 10^{-4}$  | $1,39 \times 10^{-3}$  |
| <i>CDH26</i>     | -2,57             | -0,18              | 14,66  | $2,87 \times 10^{-4}$  | $1,67 \times 10^{-3}$  |
| <i>SYT2</i>      | -2,88             | 1,20               | 13,57  | $4,64 \times 10^{-4}$  | $2,49 \times 10^{-3}$  |
| <i>SDK1</i>      | -2,93             | 3,48               | 13,51  | $4,76 \times 10^{-4}$  | $2,55 \times 10^{-3}$  |
| <i>SLC7A2</i>    | -3,53             | 3,60               | 12,32  | $8,08 \times 10^{-4}$  | $3,95 \times 10^{-3}$  |
| <i>MCM2</i>      | 1,66              | 4,47               | 11,99  | $9,39 \times 10^{-4}$  | $4,46 \times 10^{-3}$  |
| <i>GREB1L</i>    | -5,12             | -0,68              | 11,09  | $1,42 \times 10^{-3}$  | $6,26 \times 10^{-3}$  |
| <i>VCAN</i>      | 1,96              | 4,72               | 1,96   | $1,51 \times 10^{-3}$  | $6,55 \times 10^{-3}$  |
| <i>SHCBP1</i>    | 2,03              | 2,03               | 10,85  | $1,59 \times 10^{-3}$  | $6,83 \times 10^{-3}$  |
| <i>HYAL1</i>     | -3,48             | 3,11               | 10,69  | $1,70 \times 10^{-3}$  | $7,24 \times 10^{-3}$  |
| <i>TTN</i>       | -1,91             | 4,37               | 10,38  | $1,97 \times 10^{-3}$  | $8,19 \times 10^{-3}$  |
| <i>CENPF</i>     | 1,84              | 4,74               | 7,71   | $7,12 \times 10^{-3}$  | $2,32 \times 10^{-2}$  |
| <i>CMYA5</i>     | -2,90             | 1,97               | 7,61   | $7,50 \times 10^{-3}$  | $2,42 \times 10^{-2}$  |
| <i>PTPN22</i>    | -1,98             | 3,18               | 7,24   | $9,00 \times 10^{-3}$  | $2,81 \times 10^{-2}$  |
| <i>FSTL4</i>     | -2,27             | -1,16              | 6,03   | $1,67 \times 10^{-2}$  | $4,60 \times 10^{-2}$  |
| <i>AJAP1</i>     | -2,41             | -0,08              | 6,01   | $1,68 \times 10^{-2}$  | $4,62 \times 10^{-2}$  |
| <i>GLP1R</i>     | -5,91             | -3,25              | 5,93   | $1,76 \times 10^{-2}$  | $4,77 \times 10^{-2}$  |

**Tabla suplementaria 61. Valores procedentes del análisis de expresión diferencial *All vs one* de las muestras *CMS2* para los genes *CTNNB1* y *LEF1*.**  
logFC: cambio en la expresión del gen expresada en logaritmo en base 2; FDR: p-valor con corrección Benjamin-Hochberg.

|         | <i>CTNNB1</i> |                       | <i>LEF1</i> |                       |
|---------|---------------|-----------------------|-------------|-----------------------|
|         | logFC         | FDR                   | logFC       | FDR                   |
| 303     | 0,51          | 0,113                 | 2,16        | 6,90x10 <sup>-3</sup> |
| 310     | 0,82          | 3,07x10 <sup>-3</sup> | 1,83        | 0,022                 |
| COAD-1  | 1,47          | 4,79x10 <sup>-9</sup> | 1,05        | 0,181                 |
| COAD-2  | 0,76          | 3,15x10 <sup>-3</sup> | 0,90        | 0,280                 |
| COAD-12 | 1,47          | 1,03x10 <sup>-8</sup> | 0,82        | 0,347                 |
| COAD-13 | 1,04          | 2,54x10 <sup>-5</sup> | 3,20        | 3,22x10 <sup>-7</sup> |
| COAD-17 | 0,90          | 5,24x10 <sup>-4</sup> | 1,87        | 0,011                 |
| READ-5  | 1,21          | 1,11x10 <sup>-6</sup> | 1,99        | 3,86x10 <sup>-3</sup> |

**Tabla suplementaria 62. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 104.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|               | logFC | logCPM | F     | p-valor                | FDR                    |
|---------------|-------|--------|-------|------------------------|------------------------|
| <i>TULP3</i>  | 2,71  | 4,01   | 93,91 | 2,40x10 <sup>-14</sup> | 1,96x10 <sup>-11</sup> |
| <i>COL5A3</i> | 1,90  | 4,76   | 20,28 | 2,76x10 <sup>-5</sup>  | 1,20x10 <sup>-3</sup>  |
| <i>KIF14</i>  | 2,66  | 2,11   | 16,74 | 1,18x10 <sup>-4</sup>  | 4,00x10 <sup>-3</sup>  |
| <i>LAMA5</i>  | 2,66  | 7,22   | 15,88 | 1,70x10 <sup>-4</sup>  | 5,43x10 <sup>-3</sup>  |
| <i>MACC1</i>  | 1,88  | 3,70   | 13,39 | 5,01x10 <sup>-4</sup>  | 1,28x10 <sup>-2</sup>  |
| <i>UNC93A</i> | 3,33  | -1,59  | 13,35 | 5,10x10 <sup>-4</sup>  | 1,30x10 <sup>-2</sup>  |
| <i>FANCD2</i> | 1,55  | 3,70   | 12,47 | 7,57x10 <sup>-4</sup>  | 1,74x10 <sup>-2</sup>  |
| <i>CEL</i>    | 2,53  | 1,60   | 11,85 | 9,99x10 <sup>-4</sup>  | 2,11x10 <sup>-2</sup>  |
| <i>CSF2</i>   | 3,95  | -4,14  | 11,24 | 1,32x10 <sup>-3</sup>  | 2,60x10 <sup>-2</sup>  |
| <i>CLEC5A</i> | 2,70  | -1,75  | 10,10 | 2,25x10 <sup>-3</sup>  | 3,76x10 <sup>-2</sup>  |
| <i>MYOM2</i>  | -3,14 | 1,12   | 9,73  | 2,68x10 <sup>-3</sup>  | 4,26x10 <sup>-2</sup>  |
| <i>DNAH5</i>  | 2,30  | 0,39   | 9,41  | 3,12x10 <sup>-3</sup>  | 4,77x10 <sup>-2</sup>  |

**Tabla suplementaria 63. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente 107.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                 | logFC | logCPM | F      | p-valor                | FDR                    |
|-----------------|-------|--------|--------|------------------------|------------------------|
| <i>OTOP1</i>    | 10,96 | -3,59  | 196,15 | $1,60 \times 10^{-21}$ | $2,84 \times 10^{-18}$ |
| <i>DNAH2</i>    | 3,44  | 0,07   | 71,20  | $4,09 \times 10^{-12}$ | $1,01 \times 10^{-9}$  |
| <i>MUC16</i>    | 5,58  | -1,96  | 45,75  | $4,11 \times 10^{-9}$  | $4,96 \times 10^{-7}$  |
| <i>ACAN</i>     | 4,87  | -1,16  | 40,78  | $1,91 \times 10^{-8}$  | $2,02 \times 10^{-6}$  |
| <i>COL18A1</i>  | 2,21  | 6,51   | 29,04  | $1,00 \times 10^{-6}$  | $6,76 \times 10^{-5}$  |
| <i>BMP4</i>     | 2,26  | 4,22   | 25,67  | $3,45 \times 10^{-6}$  | $1,94 \times 10^{-4}$  |
| <i>DNAH14</i>   | 1,79  | 1,58   | 24,18  | $6,03 \times 10^{-6}$  | $3,20 \times 10^{-4}$  |
| <i>SLC16A4</i>  | 2,68  | 1,19   | 22,90  | $9,87 \times 10^{-6}$  | $4,86 \times 10^{-4}$  |
| <i>TNK2</i>     | 1,61  | 5,34   | 19,64  | $3,58 \times 10^{-5}$  | $1,41 \times 10^{-3}$  |
| <i>DMXL2</i>    | -1,55 | 5,85   | 19,24  | $4,21 \times 10^{-5}$  | $1,62 \times 10^{-3}$  |
| <i>THNSL2</i>   | -6,87 | 4,30   | 17,32  | $9,26 \times 10^{-5}$  | $3,01 \times 10^{-3}$  |
| <i>ADAMTS18</i> | 2,72  | -1,80  | 11,95  | $9,58 \times 10^{-4}$  | $1,83 \times 10^{-2}$  |

**Tabla suplementaria 64. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente 108.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                 | logFC | logCPM | F      | p-valor                | FDR                    |
|-----------------|-------|--------|--------|------------------------|------------------------|
| <i>CELSR3</i>   | 2,90  | 2,58   | 122,21 | $1,03 \times 10^{-16}$ | $8,08 \times 10^{-14}$ |
| <i>ADGRF5</i>   | 2,33  | 3,97   | 39,07  | $3,31 \times 10^{-8}$  | $2,53 \times 10^{-6}$  |
| <i>NR3C2</i>    | -2,77 | 6,28   | 36,32  | $8,12 \times 10^{-8}$  | $5,33 \times 10^{-6}$  |
| <i>SPIB</i>     | -6,08 | 5,80   | 30,68  | $5,58 \times 10^{-7}$  | $2,80 \times 10^{-5}$  |
| <i>SH3PXD2B</i> | 2,01  | 5,17   | 29,93  | $7,29 \times 10^{-7}$  | $3,45 \times 10^{-5}$  |
| <i>MMS22L</i>   | 1,75  | 2,85   | 26,86  | $2,21 \times 10^{-6}$  | $8,62 \times 10^{-5}$  |
| <i>CENPQ</i>    | 1,93  | 1,61   | 26,10  | $2,93 \times 10^{-6}$  | $1,09 \times 10^{-4}$  |
| <i>SLC7A5</i>   | 3,43  | 4,33   | 24,73  | $4,90 \times 10^{-6}$  | $1,64 \times 10^{-4}$  |
| <i>SLC4A4</i>   | -4,18 | 7,69   | 24,16  | $6,08 \times 10^{-6}$  | $1,96 \times 10^{-4}$  |
| <i>ADAMTS2</i>  | 2,41  | 4,14   | 22,57  | $1,12 \times 10^{-5}$  | $3,18 \times 10^{-4}$  |
| <i>IL17RC</i>   | -1,76 | 5,10   | 20,46  | $2,57 \times 10^{-5}$  | $6,03 \times 10^{-4}$  |
| <i>FANCD2</i>   | 1,85  | 3,71   | 18,72  | $5,20 \times 10^{-5}$  | $1,06 \times 10^{-3}$  |
| <i>OPTN</i>     | -1,67 | 6,66   | 17,57  | $8,35 \times 10^{-5}$  | $1,52 \times 10^{-3}$  |
| <i>PNPLA7</i>   | -2,84 | 5,18   | 17,05  | $1,04 \times 10^{-4}$  | $1,79 \times 10^{-3}$  |
| <i>SIGLEC1</i>  | -3,10 | 4,91   | 16,95  | $1,08 \times 10^{-4}$  | $1,84 \times 10^{-3}$  |
| <i>NRAP</i>     | -4,76 | 1,18   | 14,65  | $2,89 \times 10^{-4}$  | $3,91 \times 10^{-3}$  |

Continuación tabla suplementaria 64.

|         | logFC | logCPM | F     | p-valor               | FDR                   |
|---------|-------|--------|-------|-----------------------|-----------------------|
| TCEA3   | -1,71 | 6,96   | 13,94 | 3,94x10 <sup>-4</sup> | 4,94x10 <sup>-3</sup> |
| IL2RG   | -3,06 | 6,39   | 13,69 | 4,40x10 <sup>-4</sup> | 5,37x10 <sup>-3</sup> |
| POLE2   | 1,85  | 1,86   | 13,54 | 4,69x10 <sup>-4</sup> | 5,62x10 <sup>-3</sup> |
| LRRRC66 | -2,88 | 3,79   | 13,52 | 4,74x10 <sup>-4</sup> | 5,66x10 <sup>-3</sup> |
| CTH     | -1,85 | 2,85   | 12,69 | 6,84x10 <sup>-4</sup> | 7,51x10 <sup>-3</sup> |
| GRAMD2B | -1,52 | 5,57   | 10,75 | 1,66x10 <sup>-3</sup> | 1,41x10 <sup>-2</sup> |
| TTLL6   | -2,50 | 4,11   | 9,48  | 3,01x10 <sup>-3</sup> | 2,19x10 <sup>-2</sup> |
| TRPM4   | -1,51 | 8,11   | 9,43  | 3,09x10 <sup>-3</sup> | 2,24x10 <sup>-2</sup> |
| SLIT1   | -2,95 | -0,08  | 7,56  | 7,69x10 <sup>-3</sup> | 4,33x10 <sup>-2</sup> |
| TMEM54  | -1,66 | 8,11   | 7,31  | 8,71x10 <sup>-3</sup> | 4,73x10 <sup>-2</sup> |
| EXOC3L4 | -1,99 | 3,21   | 7,31  | 8,72x10 <sup>-3</sup> | 4,73x10 <sup>-2</sup> |
| ABCB11  | -5,32 | 2,75   | 7,17  | 9,32x10 <sup>-2</sup> | 4,96x10 <sup>-2</sup> |

**Tabla suplementaria 65. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-6.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|          | logFC | logCPM | F      | p-valor                | FDR                    |
|----------|-------|--------|--------|------------------------|------------------------|
| LAMP1    | 1,71  | 7,61   | 126,76 | 4,62x10 <sup>-17</sup> | 4,20x10 <sup>-14</sup> |
| ROBO2    | 4,55  | 0,79   | 72,13  | 3,26x10 <sup>-12</sup> | 4,94x10 <sup>-10</sup> |
| LIN9     | 2,01  | 2,17   | 40,35  | 2,19x10 <sup>-8</sup>  | 8,11x10 <sup>-7</sup>  |
| PKDCC    | -3,19 | 5,97   | 38,89  | 3,50x10 <sup>-8</sup>  | 1,18x10 <sup>-6</sup>  |
| WDHD1    | 2,46  | 2,78   | 33,03  | 2,47x10 <sup>-7</sup>  | 5,93x10 <sup>-6</sup>  |
| NEIL1    | -2,46 | 4,48   | 32,47  | 2,99x10 <sup>-7</sup>  | 6,95x10 <sup>-6</sup>  |
| LGI4     | -5,68 | 5,31   | 32,02  | 3,50x10 <sup>-7</sup>  | 7,99x10 <sup>-6</sup>  |
| ZC3H12C  | -2,79 | 4,59   | 29,74  | 7,78x10 <sup>-7</sup>  | 1,58x10 <sup>-5</sup>  |
| CACNB2   | -3,14 | 4,09   | 27,83  | 1,55x10 <sup>-6</sup>  | 2,77x10 <sup>-5</sup>  |
| GIGYF1   | -1,59 | 6,50   | 25,52  | 3,64x10 <sup>-6</sup>  | 5,53x10 <sup>-5</sup>  |
| TRPV1    | -2,92 | 4,07   | 23,00  | 9,47x10 <sup>-6</sup>  | 1,24x10 <sup>-4</sup>  |
| CDK3     | -3,97 | 4,11   | 22,81  | 1,02x10 <sup>-5</sup>  | 1,32x10 <sup>-4</sup>  |
| ADCY6    | -1,51 | 7,65   | 20,52  | 2,51x10 <sup>-5</sup>  | 2,77x10 <sup>-4</sup>  |
| RGPD3    | -7,08 | -2,14  | 16,92  | 1,09x10 <sup>-4</sup>  | 9,27x10 <sup>-4</sup>  |
| MROH7    | -2,91 | 4,34   | 16,74  | 1,18x10 <sup>-4</sup>  | 9,87x10 <sup>-4</sup>  |
| LAPTM4B  | 1,51  | 5,51   | 16,73  | 1,19x10 <sup>-4</sup>  | 9,90x10 <sup>-4</sup>  |
| CFTR     | 1,51  | 7,57   | 15,91  | 1,68x10 <sup>-4</sup>  | 1,31x10 <sup>-3</sup>  |
| NUTM1    | -4,52 | -0,89  | 15,63  | 1,89x10 <sup>-4</sup>  | 1,45x10 <sup>-3</sup>  |
| KLF15    | -5,25 | 3,08   | 15,31  | 2,17x10 <sup>-4</sup>  | 1,61x10 <sup>-3</sup>  |
| SLC16A11 | -3,34 | 0,05   | 14,54  | 3,02x10 <sup>-4</sup>  | 2,11x10 <sup>-3</sup>  |
| ASPM     | 2,20  | 3,52   | 13,95  | 3,92x10 <sup>-4</sup>  | 2,59x10 <sup>-3</sup>  |
| KNL1     | 2,11  | 3,22   | 13,30  | 5,21x10 <sup>-4</sup>  | 3,25x10 <sup>-3</sup>  |
| TTN      | -2,23 | 4,37   | 13,25  | 5,33x10 <sup>-4</sup>  | 3,31x10 <sup>-3</sup>  |
| MS4A12   | -6,91 | 7,81   | 12,31  | 8,11x10 <sup>-4</sup>  | 4,64x10 <sup>-3</sup>  |
| AZIN2    | -2,82 | 3,85   | 12,22  | 8,47x10 <sup>-4</sup>  | 4,80x10 <sup>-3</sup>  |
| GRIN1    | -3,31 | 3,11   | 12,21  | 8,50x10 <sup>-4</sup>  | 4,81x10 <sup>-3</sup>  |

**Continuación tabla suplementaria 65.**

|          | logFC | logCPM | F     | p-valor               | FDR                   |
|----------|-------|--------|-------|-----------------------|-----------------------|
| ZNF285   | -2,53 | 0,81   | 11,21 | 1,34x10 <sup>-3</sup> | 7,01x10 <sup>-3</sup> |
| CMYA5    | -3,48 | 1,96   | 10,13 | 2,22x10 <sup>-3</sup> | 1,04x10 <sup>-2</sup> |
| AKR1C3   | -4,14 | 7,68   | 10,09 | 2,26x10 <sup>-3</sup> | 1,06x10 <sup>-2</sup> |
| CATSPERE | -2,38 | -1,07  | 9,50  | 2,99x10 <sup>-3</sup> | 1,33x10 <sup>-2</sup> |
| KLK4     | -5,42 | 0,78   | 9,23  | 3,39x10 <sup>-3</sup> | 1,47x10 <sup>-2</sup> |
| SLC4A10  | -4,24 | 2,60   | 7,63  | 7,41x10 <sup>-3</sup> | 2,73x10 <sup>-2</sup> |
| FTCD     | -4,47 | -1,07  | 7,10  | 9,65x10 <sup>-3</sup> | 3,37x10 <sup>-2</sup> |
| CCN3     | -2,22 | 2,70   | 6,85  | 1,10x10 <sup>-2</sup> | 3,72x10 <sup>-2</sup> |
| RCVRN    | -3,66 | -2,20  | 6,82  | 1,11x10 <sup>-2</sup> | 3,76x10 <sup>-2</sup> |
| CCR5     | 1,71  | 2,08   | 6,81  | 1,12x10 <sup>-2</sup> | 3,78x10 <sup>-2</sup> |
| EFS      | -2,36 | 3,30   | 6,37  | 1,40x10 <sup>-2</sup> | 4,49x10 <sup>-2</sup> |

**Tabla suplementaria 66. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente READ-6.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|          | logFC | logCPM | F     | p-valor               | FDR                   |
|----------|-------|--------|-------|-----------------------|-----------------------|
| SYCP2    | 2,09  | 1,19   | 42,59 | 1,08x10 <sup>-8</sup> | 3,30x10 <sup>-7</sup> |
| PKDCC    | -3,23 | 5,98   | 39,62 | 2,77x10 <sup>-8</sup> | 7,25x10 <sup>-7</sup> |
| STK31    | 3,33  | -0,44  | 38,92 | 3,46x10 <sup>-8</sup> | 8,76x10 <sup>-7</sup> |
| ATP13A3  | 1,78  | 6,56   | 27,38 | 1,83x10 <sup>-6</sup> | 2,47x10 <sup>-5</sup> |
| TGM1     | -2,94 | 0,54   | 27,13 | 2,01x10 <sup>-6</sup> | 2,68x10 <sup>-5</sup> |
| ADCY4    | -2,51 | 4,05   | 25,97 | 3,08x10 <sup>-6</sup> | 3,83x10 <sup>-5</sup> |
| L1TD1    | 4,50  | 3,67   | 24,96 | 4,49x10 <sup>-6</sup> | 5,27x10 <sup>-5</sup> |
| PGM2L1   | 1,66  | 4,14   | 24,20 | 5,98x10 <sup>-6</sup> | 6,82x10 <sup>-5</sup> |
| SLFN13   | -3,41 | 3,80   | 21,94 | 1,43x10 <sup>-5</sup> | 1,43x10 <sup>-4</sup> |
| KIF26A   | 2,03  | 3,58   | 19,74 | 3,44x10 <sup>-5</sup> | 3,04x10 <sup>-4</sup> |
| CNTNAP2  | -3,97 | 2,64   | 18,68 | 5,27x10 <sup>-5</sup> | 4,38x10 <sup>-4</sup> |
| SORCS1   | -5,73 | 2,81   | 16,94 | 1,08x10 <sup>-4</sup> | 8,06x10 <sup>-4</sup> |
| DCLK3    | 2,51  | -1,93  | 16,64 | 1,23x10 <sup>-4</sup> | 8,98x10 <sup>-4</sup> |
| KCTD1    | -2,21 | 2,79   | 16,38 | 1,37x10 <sup>-4</sup> | 9,82x10 <sup>-4</sup> |
| EPHX2    | -2,17 | 6,57   | 15,92 | 1,67x10 <sup>-4</sup> | 1,16x10 <sup>-3</sup> |
| FBN2     | 1,86  | 2,11   | 15,23 | 2,24x10 <sup>-4</sup> | 1,48x10 <sup>-3</sup> |
| MAMDC4   | -2,24 | 5,47   | 14,69 | 2,83x10 <sup>-4</sup> | 1,80x10 <sup>-3</sup> |
| ZFYVE28  | -2,13 | 4,91   | 13,78 | 4,23x10 <sup>-4</sup> | 2,53x10 <sup>-3</sup> |
| IQGAP3   | 2,05  | 4,46   | 13,01 | 5,92x10 <sup>-4</sup> | 3,35x10 <sup>-3</sup> |
| SHANK3   | -1,94 | 4,76   | 12,99 | 6,00x10 <sup>-4</sup> | 3,38x10 <sup>-3</sup> |
| TRPM2    | 1,53  | 2,97   | 11,77 | 1,04x10 <sup>-3</sup> | 5,33x10 <sup>-3</sup> |
| KIF18B   | 2,19  | 3,21   | 11,71 | 1,07x10 <sup>-3</sup> | 5,45x10 <sup>-3</sup> |
| TTN      | -1,81 | 4,37   | 9,39  | 3,15x10 <sup>-3</sup> | 1,32x10 <sup>-2</sup> |
| TMEM150B | 1,99  | 3,00   | 9,10  | 3,61x10 <sup>-3</sup> | 1,47x10 <sup>-2</sup> |
| VEPH1    | -3,34 | 0,43   | 8,66  | 4,48x10 <sup>-3</sup> | 1,74x10 <sup>-2</sup> |
| COL3A1   | 1,54  | 8,94   | 8,57  | 4,68x10 <sup>-3</sup> | 1,81x10 <sup>-2</sup> |
| GRIN1    | -2,56 | 3,11   | 8,29  | 5,35x10 <sup>-3</sup> | 2,01x10 <sup>-2</sup> |

## Continuación tabla suplementaria 66.

|        | logFC | logCPM | F    | p-valor               | FDR                   |
|--------|-------|--------|------|-----------------------|-----------------------|
| STAT4  | -1,73 | 2,22   | 7,19 | 9,23x10 <sup>-3</sup> | 3,11x10 <sup>-2</sup> |
| VCAN   | 1,62  | 4,70   | 6,83 | 1,11x10 <sup>-2</sup> | 3,58x10 <sup>-2</sup> |
| CCDC60 | -3,77 | 0,09   | 6,02 | 1,68x10 <sup>-2</sup> | 4,97x10 <sup>-2</sup> |

**Tabla suplementaria 67. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente 102.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamín-Hochberg.

|          | logFC | logCPM | F      | p-valor                | FDR                    |
|----------|-------|--------|--------|------------------------|------------------------|
| ESM1     | 8,77  | -0,92  | 135,42 | 1,06x10 <sup>-17</sup> | 1,14x10 <sup>-14</sup> |
| MMP7     | 8,75  | 1,11   | 84,36  | 1,89x10 <sup>-13</sup> | 6,49x10 <sup>-11</sup> |
| NOTCH3   | 3,77  | 5,10   | 63,63  | 2,76x10 <sup>-11</sup> | 5,28x10 <sup>-9</sup>  |
| VWF      | 3,01  | 6,78   | 56,83  | 1,69x10 <sup>-10</sup> | 2,58x10 <sup>-8</sup>  |
| AFAP1L1  | 2,82  | 2,32   | 49,45  | 1,36x10 <sup>-9</sup>  | 1,59x10 <sup>-7</sup>  |
| CEP164   | 1,53  | 4,33   | 42,60  | 1,08x10 <sup>-8</sup>  | 9,45x10 <sup>-7</sup>  |
| ABCC2    | 2,53  | 0,91   | 37,07  | 6,34x10 <sup>-8</sup>  | 4,14x10 <sup>-6</sup>  |
| SLC22A12 | 5,57  | -4,27  | 32,94  | 2,55x10 <sup>-7</sup>  | 1,38x10 <sup>-5</sup>  |
| PTPRS    | 2,54  | 5,02   | 28,06  | 1,43x10 <sup>-6</sup>  | 5,88x10 <sup>-5</sup>  |
| CACNA1A  | -3,94 | 1,96   | 24,40  | 5,56x10 <sup>-6</sup>  | 1,84x10 <sup>-4</sup>  |
| CSMD2    | 2,69  | -0,44  | 23,69  | 7,58x10 <sup>-6</sup>  | 2,32x10 <sup>-4</sup>  |
| MROH7    | -3,60 | 4,34   | 22,24  | 1,28x10 <sup>-5</sup>  | 3,67x10 <sup>-4</sup>  |
| LMCD1    | 2,16  | 2,76   | 20,87  | 2,19x10 <sup>-5</sup>  | 5,74x10 <sup>-4</sup>  |
| INSC     | 1,82  | 1,86   | 18,77  | 5,09x10 <sup>-5</sup>  | 1,15x10 <sup>-3</sup>  |
| MGAT4B   | -2,22 | 8,50   | 16,24  | 1,46x10 <sup>-4</sup>  | 2,74x10 <sup>-3</sup>  |
| GPR78    | 2,75  | 1,80   | 14,59  | 2,96x10 <sup>-4</sup>  | 4,73x10 <sup>-3</sup>  |
| DQX1     | -2,42 | 5,07   | 14,11  | 3,65x10 <sup>-4</sup>  | 5,58x10 <sup>-3</sup>  |
| NINL     | 2,26  | 3,23   | 14,10  | 3,67x10 <sup>-4</sup>  | 5,60x10 <sup>-3</sup>  |
| MFGE8    | 1,69  | 6,44   | 14,09  | 3,68x10 <sup>-4</sup>  | 5,60x10 <sup>-3</sup>  |
| PDZD7    | 1,57  | 0,48   | 12,11  | 8,89x10 <sup>-4</sup>  | 1,11x10 <sup>-2</sup>  |
| SLC28A1  | 3,89  | -3,69  | 11,96  | 9,51x10 <sup>-4</sup>  | 1,16x10 <sup>-2</sup>  |
| UBAP1L   | -2,42 | 0,12   | 11,87  | 9,90x10 <sup>-4</sup>  | 1,20x10 <sup>-2</sup>  |
| UGT1A4   | -3,53 | 6,15   | 11,73  | 1,06x10 <sup>-3</sup>  | 1,26x10 <sup>-2</sup>  |
| ADRB1    | -2,50 | 1,02   | 10,17  | 2,18x10 <sup>-3</sup>  | 2,19x10 <sup>-2</sup>  |
| BFSP1    | 1,64  | -0,43  | 10,16  | 2,18x10 <sup>-3</sup>  | 2,19x10 <sup>-2</sup>  |
| PDZD2    | 1,60  | 2,58   | 10,02  | 2,34x10 <sup>-3</sup>  | 2,31x10 <sup>-2</sup>  |
| HAL      | 1,68  | 0,33   | 9,53   | 2,94x10 <sup>-3</sup>  | 2,72x10 <sup>-2</sup>  |
| ITGAX    | 1,71  | 3,07   | 7,98   | 6,24x10 <sup>-3</sup>  | 4,60x10 <sup>-2</sup>  |

**Tabla suplementaria 68. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 209.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|         | logFC | logCPM | F     | p-valor                | FDR                   |
|---------|-------|--------|-------|------------------------|-----------------------|
| ADAMTS2 | 4,31  | 4,33   | 51,83 | $6,85 \times 10^{-10}$ | $2,80 \times 10^{-7}$ |
| BICC1   | 2,63  | 2,81   | 39,68 | $2,72 \times 10^{-8}$  | $5,55 \times 10^{-6}$ |
| LRRC32  | 2,65  | 4,92   | 34,97 | $1,28 \times 10^{-7}$  | $1,85 \times 10^{-5}$ |
| ACAN    | 6,27  | -0,40  | 34,88 | $1,31 \times 10^{-7}$  | $1,89 \times 10^{-5}$ |
| P4HA3   | 2,75  | 0,65   | 29,23 | $9,35 \times 10^{-7}$  | $9,07 \times 10^{-5}$ |
| COL5A1  | 3,32  | 6,68   | 23,47 | $7,93 \times 10^{-6}$  | $4,82 \times 10^{-4}$ |
| MYO1A   | -2,52 | 7,77   | 22,60 | $1,11 \times 10^{-5}$  | $6,39 \times 10^{-4}$ |
| GPAT3   | -3,66 | 4,87   | 17,59 | $8,26 \times 10^{-5}$  | $3,01 \times 10^{-3}$ |
| COL2A1  | -5,56 | 0,74   | 17,42 | $8,88 \times 10^{-5}$  | $3,19 \times 10^{-3}$ |
| MYO7A   | 1,61  | 2,96   | 14,46 | $3,13 \times 10^{-4}$  | $8,24 \times 10^{-3}$ |
| PIEZO1  | 1,54  | 7,05   | 14,31 | $3,34 \times 10^{-4}$  | $8,69 \times 10^{-3}$ |
| ITGAM   | 2,06  | 2,82   | 12,72 | $6,77 \times 10^{-4}$  | $1,45 \times 10^{-2}$ |
| SFRP1   | -4,59 | 4,31   | 11,94 | $9,63 \times 10^{-4}$  | $1,85 \times 10^{-2}$ |
| ABCA9   | -3,14 | 2,92   | 10,43 | $1,93 \times 10^{-3}$  | $2,97 \times 10^{-2}$ |
| PTPRO   | 2,06  | 4,54   | 9,12  | $3,59 \times 10^{-3}$  | $4,54 \times 10^{-2}$ |

**Tabla suplementaria 69. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 302.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|         | logFC | logCPM | F      | p-valor                | FDR                    |
|---------|-------|--------|--------|------------------------|------------------------|
| HEPHL1  | 6,69  | -0,37  | 123,18 | $8,68 \times 10^{-17}$ | $6,16 \times 10^{-14}$ |
| ROBO4   | 3,29  | 3,54   | 118,96 | $1,85 \times 10^{-16}$ | $1,15 \times 10^{-13}$ |
| PALD1   | 3,52  | 3,54   | 98,20  | $9,92 \times 10^{-15}$ | $3,20 \times 10^{-12}$ |
| SPARC   | 3,35  | 7,88   | 85,73  | $1,40 \times 10^{-13}$ | $3,03 \times 10^{-22}$ |
| INSC    | 2,51  | 1,88   | 42,38  | $1,16 \times 10^{-8}$  | $7,20 \times 10^{-7}$  |
| SORD    | 1,88  | 4,29   | 29,31  | $9,09 \times 10^{-7}$  | $3,51 \times 10^{-5}$  |
| ITGA11  | 2,29  | 2,22   | 26,66  | $2,38 \times 10^{-6}$  | $8,18 \times 10^{-5}$  |
| TOX     | -2,15 | 4,30   | 26,65  | $2,39 \times 10^{-6}$  | $8,19 \times 10^{-5}$  |
| TP73    | 1,95  | -0,14  | 23,32  | $8,39 \times 10^{-4}$  | $2,41 \times 10^{-4}$  |
| MUC16   | 4,28  | 2,32   | 22,43  | $1,18 \times 10^{-5}$  | $3,22 \times 10^{-4}$  |
| CFAP46  | 3,36  | -2,11  | 21,72  | $1,56 \times 10^{-5}$  | $4,06 \times 10^{-4}$  |
| COL27A1 | 2,01  | 4,41   | 18,94  | $4,75 \times 10^{-5}$  | $1,04 \times 10^{-3}$  |
| PLEKHO2 | 1,89  | 4,68   | 18,15  | $6,58 \times 10^{-5}$  | $1,37 \times 10^{-3}$  |

Continuación tabla suplementaria 69.

|                | logFC | logCPM | F     | p-valor               | FDR                   |
|----------------|-------|--------|-------|-----------------------|-----------------------|
| <i>ADGRV1</i>  | -2,29 | 1,68   | 16,27 | 1,44x10 <sup>-4</sup> | 2,61x10 <sup>-3</sup> |
| <i>COL15A1</i> | 3,17  | 5,39   | 15,67 | 1,86x10 <sup>-4</sup> | 3,20x10 <sup>-3</sup> |
| <i>DOP1B</i>   | -1,55 | 6,24   | 12,67 | 6,91x10 <sup>-4</sup> | 8,98x10 <sup>-3</sup> |
| <i>FN1</i>     | 3,26  | 8,65   | 12,26 | 8,30x10 <sup>-4</sup> | 1,04x10 <sup>-2</sup> |
| <i>SH3TC2</i>  | 2,04  | 0,54   | 11,65 | 1,09x10 <sup>-3</sup> | 1,29x10 <sup>-2</sup> |
| <i>SULF1</i>   | 2,82  | 4,74   | 10,12 | 2,23x10 <sup>-3</sup> | 2,15x10 <sup>-2</sup> |
| <i>CHI3L2</i>  | 2,66  | 0,82   | 8,82  | 4,13x10 <sup>-3</sup> | 3,39x10 <sup>-2</sup> |
| <i>SACS</i>    | 1,78  | 3,64   | 8,38  | 5,12x10 <sup>-3</sup> | 3,97x10 <sup>-2</sup> |

**Tabla suplementaria 70. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 304.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamín-Hochberg.

|                 | logFC | logCPM | F      | p-valor                | FDR                    |
|-----------------|-------|--------|--------|------------------------|------------------------|
| <i>THBS2</i>    | 8,46  | 4,45   | 153,41 | 6,11x10 <sup>-19</sup> | 2,71x10 <sup>-16</sup> |
| <i>TFR2</i>     | 3,69  | -0,13  | 116,09 | 3,11x10 <sup>-16</sup> | 8,59x10 <sup>-14</sup> |
| <i>SYCP2</i>    | 3,15  | 1,20   | 110,94 | 8,11x10 <sup>-16</sup> | 2,03x10 <sup>-13</sup> |
| <i>MET</i>      | 3,13  | 5,74   | 89,03  | 6,79x10 <sup>-14</sup> | 9,70x10 <sup>-12</sup> |
| <i>LRRC32</i>   | 3,13  | 4,90   | 80,86  | 4,17x10 <sup>-13</sup> | 4,96x10 <sup>-11</sup> |
| <i>KIAA1549</i> | 2,61  | 2,65   | 79,28  | 6,00x10 <sup>-13</sup> | 6,65x10 <sup>-11</sup> |
| <i>FST</i>      | 4,54  | 2,96   | 57,44  | 1,43x10 <sup>-10</sup> | 1,02x10 <sup>-8</sup>  |
| <i>AEBP1</i>    | 3,47  | 7,03   | 55,51  | 2,43x10 <sup>-10</sup> | 1,67x10 <sup>-8</sup>  |
| <i>PPL</i>      | 2,86  | 4,85   | 53,01  | 4,91x10 <sup>-10</sup> | 3,16x10 <sup>-8</sup>  |
| <i>ESAM</i>     | 1,85  | 4,27   | 45,82  | 4,02x10 <sup>-9</sup>  | 2,04x10 <sup>-7</sup>  |
| <i>CEP170</i>   | 2,21  | 3,37   | 42,87  | 9,93x10 <sup>-9</sup>  | 4,68x10 <sup>-7</sup>  |
| <i>DOK7</i>     | 3,38  | 0,01   | 32,89  | 2,59x10 <sup>-7</sup>  | 9,40x10 <sup>-6</sup>  |
| <i>GLI2</i>     | 3,53  | 2,32   | 32,69  | 2,78x10 <sup>-7</sup>  | 9,95x10 <sup>-6</sup>  |
| <i>MPO</i>      | 4,25  | -2,56  | 25,28  | 3,99x10 <sup>-6</sup>  | 1,10x10 <sup>-4</sup>  |
| <i>ALPK3</i>    | 4,43  | 3,50   | 23,70  | 7,25x10 <sup>-6</sup>  | 1,88x10 <sup>-4</sup>  |
| <i>RP1L1</i>    | 3,95  | -2,65  | 20,41  | 2,62x10 <sup>-5</sup>  | 5,82x10 <sup>-4</sup>  |
| <i>TNC</i>      | 5,04  | 6,94   | 18,91  | 4,80x10 <sup>-5</sup>  | 9,73x10 <sup>-4</sup>  |
| <i>HECTD3</i>   | -1,62 | 7,03   | 18,54  | 5,60x10 <sup>-5</sup>  | 1,12x10 <sup>-3</sup>  |
| <i>TRABD2A</i>  | -2,20 | 6,42   | 16,35  | 1,39x10 <sup>-4</sup>  | 2,38x10 <sup>-3</sup>  |
| <i>LRRC17</i>   | 2,39  | 0,43   | 15,62  | 1,90x10 <sup>-4</sup>  | 3,09x10 <sup>-3</sup>  |
| <i>ANKLE1</i>   | 1,85  | -0,18  | 11,00  | 1,48x10 <sup>-3</sup>  | 1,55x10 <sup>-2</sup>  |
| <i>LAMC1</i>    | 1,64  | 7,21   | 10,91  | 1,54x10 <sup>-3</sup>  | 1,61x10 <sup>-2</sup>  |
| <i>PCDHB14</i>  | 1,67  | 1,59   | 10,79  | 1,63x10 <sup>-3</sup>  | 1,68x10 <sup>-2</sup>  |
| <i>ABCA10</i>   | -2,08 | 3,03   | 9,79   | 2,61x10 <sup>-3</sup>  | 2,39x10 <sup>-2</sup>  |
| <i>PDE6A</i>    | -3,50 | 3,31   | 9,56   | 2,90x10 <sup>-3</sup>  | 2,59x10 <sup>-2</sup>  |
| <i>CCDC68</i>   | -1,62 | 5,49   | 8,19   | 5,62x10 <sup>-3</sup>  | 4,14x10 <sup>-2</sup>  |



**Tabla suplementaria 71. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente 305.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | logFC | logCPM | F     | p-valor                | FDR                   |
|----------------|-------|--------|-------|------------------------|-----------------------|
| <i>APLN</i>    | 5,99  | 0,39   | 60,05 | $7,08 \times 10^{-11}$ | $1,01 \times 10^{-8}$ |
| <i>VTCN1</i>   | 6,50  | -4,29  | 45,24 | $4,79 \times 10^{-9}$  | $4,45 \times 10^{-7}$ |
| <i>PLEKHG4</i> | 2,39  | 4,28   | 33,10 | $2,40 \times 10^{-7}$  | $1,32 \times 10^{-5}$ |
| <i>PLVAP</i>   | 1,89  | 6,51   | 27,63 | $1,67 \times 10^{-6}$  | $6,74 \times 10^{-5}$ |
| <i>VWA5B2</i>  | -5,81 | 1,61   | 26,52 | $2,51 \times 10^{-6}$  | $9,34 \times 10^{-5}$ |
| <i>PCED1A</i>  | 1,57  | 5,26   | 23,47 | $7,93 \times 10^{-6}$  | $2,40 \times 10^{-4}$ |
| <i>NTRK2</i>   | -7,11 | 4,18   | 23,37 | $8,23 \times 10^{-6}$  | $2,47 \times 10^{-4}$ |
| <i>LYPD3</i>   | 2,01  | 0,17   | 22,91 | $9,84 \times 10^{-6}$  | $2,86 \times 10^{-4}$ |
| <i>SLC19A1</i> | 1,97  | 5,39   | 20,44 | $2,59 \times 10^{-5}$  | $6,41 \times 10^{-4}$ |
| <i>PLD1</i>    | -2,25 | 5,71   | 18,79 | $5,05 \times 10^{-5}$  | $1,11 \times 10^{-3}$ |
| <i>PYGO1</i>   | 3,33  | 1,95   | 17,49 | $8,62 \times 10^{-5}$  | $1,69 \times 10^{-3}$ |
| <i>ACSS2</i>   | -1,70 | 7,97   | 12,34 | $8,02 \times 10^{-4}$  | $1,00 \times 10^{-2}$ |
| <i>FAM227A</i> | 2,10  | 0,24   | 12,13 | $8,81 \times 10^{-4}$  | $1,08 \times 10^{-2}$ |
| <i>SCN5A</i>   | 1,72  | 0,06   | 11,22 | $1,34 \times 10^{-3}$  | $1,47 \times 10^{-2}$ |
| <i>TRPM4</i>   | -1,53 | 8,11   | 9,75  | $2,66 \times 10^{-3}$  | $2,45 \times 10^{-2}$ |
| <i>FAM110C</i> | -2,23 | 4,98   | 9,73  | $2,68 \times 10^{-3}$  | $2,46 \times 10^{-2}$ |
| <i>INSL3</i>   | -2,25 | -1,37  | 7,90  | $6,48 \times 10^{-3}$  | $4,61 \times 10^{-2}$ |

**Tabla suplementaria 72. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente 308.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|               | logFC | logCPM | F     | p-valor                | FDR                   |
|---------------|-------|--------|-------|------------------------|-----------------------|
| <i>ACAN</i>   | 5,90  | -1,00  | 66,51 | $1,32 \times 10^{-11}$ | $3,42 \times 10^{-9}$ |
| <i>KCNT1</i>  | 4,47  | -0,93  | 51,27 | $8,06 \times 10^{-10}$ | $1,18 \times 10^{-7}$ |
| <i>CSMD2</i>  | 3,21  | -0,42  | 37,71 | $5,15 \times 10^{-8}$  | $4,60 \times 10^{-6}$ |
| <i>DNAH10</i> | 2,12  | -0,02  | 17,39 | $9,01 \times 10^{-5}$  | $2,68 \times 10^{-3}$ |
| <i>DCST2</i>  | -3,01 | 1,80   | 13,52 | $4,73 \times 10^{-4}$  | $1,00 \times 10^{-2}$ |
| <i>ECEL1</i>  | 3,75  | -1,69  | 13,02 | $5,91 \times 10^{-4}$  | $1,19 \times 10^{-2}$ |
| <i>ADGRV1</i> | -1,82 | 1,68   | 11,18 | $1,36 \times 10^{-3}$  | $2,20 \times 10^{-2}$ |
| <i>LRRC70</i> | 1,59  | 0,45   | 10,91 | $1,54 \times 10^{-3}$  | $2,41 \times 10^{-2}$ |
| <i>CPA6</i>   | -2,69 | 1,83   | 10,68 | $1,71 \times 10^{-3}$  | $2,59 \times 10^{-2}$ |
| <i>BRINP3</i> | -3,39 | 3,00   | 9,87  | $2,51 \times 10^{-3}$  | $3,38 \times 10^{-2}$ |
| <i>LRRC8E</i> | 1,86  | 1,06   | 9,14  | $3,55 \times 10^{-3}$  | $4,31 \times 10^{-2}$ |

## Continuación tabla suplementaria 72.

|      | logFC | logCPM | F    | p-valor               | FDR                   |
|------|-------|--------|------|-----------------------|-----------------------|
| NFE2 | 2,92  | -1,57  | 8,69 | 4,40x10 <sup>-3</sup> | 4,96x10 <sup>-2</sup> |

**Tabla suplementaria 73. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente COAD-3.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|          | logFC | logCPM | F      | p-valor                | FDR                    |
|----------|-------|--------|--------|------------------------|------------------------|
| CHRNA4   | 8,07  | -1,28  | 205,34 | 5,01x10 <sup>-22</sup> | 1,31x10 <sup>-19</sup> |
| BMP7     | 4,63  | 2,25   | 104,45 | 2,82x10 <sup>-15</sup> | 2,43x10 <sup>-13</sup> |
| ADAMTS20 | 6,76  | -4,27  | 92,96  | 4,63x10 <sup>-14</sup> | 3,29x10 <sup>-12</sup> |
| RNFT2    | 2,94  | 0,14   | 48,82  | 1,64x10 <sup>-9</sup>  | 5,24x10 <sup>-8</sup>  |
| SEZ6L2   | 2,62  | 4,10   | 47,42  | 2,48x10 <sup>-9</sup>  | 7,58x10 <sup>-8</sup>  |
| MYO3A    | 3,92  | -0,23  | 38,80  | 3,60x10 <sup>-8</sup>  | 8,70x10 <sup>-7</sup>  |
| WDR72    | 4,81  | 1,13   | 37,49  | 5,51x10 <sup>-8</sup>  | 1,28x10 <sup>-6</sup>  |
| TRPM3    | 3,26  | -1,32  | 34,53  | 1,48x10 <sup>-7</sup>  | 3,03x10 <sup>-6</sup>  |
| PTPRC    | 2,71  | 5,43   | 27,76  | 1,59x10 <sup>-6</sup>  | 2,47x10 <sup>-5</sup>  |
| CWH43    | -4,34 | 5,01   | 23,34  | 8,32x10 <sup>-6</sup>  | 1,02x10 <sup>-4</sup>  |
| CPNE9    | 3,18  | -2,14  | 21,81  | 1,51x10 <sup>-5</sup>  | 1,71x10 <sup>-4</sup>  |
| FAM162A  | -1,78 | 6,97   | 21,20  | 1,91x10 <sup>-5</sup>  | 2,10x10 <sup>-4</sup>  |
| CBARP    | 2,06  | 0,25   | 20,06  | 3,02x10 <sup>-5</sup>  | 3,10x10 <sup>-4</sup>  |
| ZFR2     | 3,05  | -1,86  | 19,82  | 3,32x10 <sup>-5</sup>  | 3,34x10 <sup>-4</sup>  |
| MUC16    | 3,59  | -2,34  | 18,15  | 6,56x10 <sup>-5</sup>  | 5,96x10 <sup>-4</sup>  |
| NAPEPLD  | -2,21 | 5,59   | 15,89  | 1,69x10 <sup>-4</sup>  | 1,31x10 <sup>-3</sup>  |
| TAL1     | 1,71  | 1,91   | 15,51  | 1,99x10 <sup>-4</sup>  | 1,50x10 <sup>-3</sup>  |
| GPRIN1   | 2,45  | 2,34   | 14,40  | 3,21x10 <sup>-4</sup>  | 2,21x10 <sup>-3</sup>  |
| CELSR1   | 1,95  | 3,53   | 14,20  | 3,51x10 <sup>-4</sup>  | 2,37x10 <sup>-3</sup>  |
| CNNM1    | 1,78  | 0,38   | 13,74  | 4,30x10 <sup>-4</sup>  | 2,81x10 <sup>-3</sup>  |
| ST14     | -1,77 | 9,09   | 13,01  | 5,93x10 <sup>-4</sup>  | 3,64x10 <sup>-3</sup>  |
| SHROOM3  | -1,60 | 7,51   | 12,84  | 6,39x10 <sup>-4</sup>  | 3,87x10 <sup>-3</sup>  |
| PPP1R14D | -2,36 | 5,62   | 12,19  | 8,57x10 <sup>-4</sup>  | 4,96x10 <sup>-3</sup>  |
| MYH10    | 1,93  | 4,23   | 11,29  | 1,30x10 <sup>-3</sup>  | 6,86x10 <sup>-3</sup>  |
| PANK3    | -1,50 | 7,49   | 11,16  | 1,38x10 <sup>-3</sup>  | 7,20x10 <sup>-3</sup>  |
| ACACB    | -1,98 | 6,31   | 11,11  | 1,40x10 <sup>-3</sup>  | 7,31x10 <sup>-3</sup>  |
| SCIN     | -2,88 | 5,85   | 10,29  | 2,06x10 <sup>-3</sup>  | 9,95x10 <sup>-3</sup>  |
| COL22A1  | 3,35  | -0,10  | 9,21   | 3,43x10 <sup>-3</sup>  | 1,50x10 <sup>-2</sup>  |
| STX19    | -2,10 | 3,42   | 8,72   | 4,34x10 <sup>-3</sup>  | 1,81x10 <sup>-2</sup>  |
| FOLH1    | 2,73  | -0,74  | 7,84   | 6,67x10 <sup>-3</sup>  | 2,56x10 <sup>-2</sup>  |
| CDHR2    | -2,84 | 6,70   | 6,62   | 1,23x10 <sup>-2</sup>  | 4,18x10 <sup>-2</sup>  |
| COL28A1  | -2,27 | 3,29   | 6,59   | 1,25x10 <sup>-2</sup>  | 4,22x10 <sup>-2</sup>  |

**Tabla suplementaria 74. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente COAD-5. Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.**

|          | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | F      | p-valor                | FDR                    |
|----------|-------------------|--------------------|--------|------------------------|------------------------|
| NELFCD   | 2,42              | 5,62               | 229,12 | $3,07 \times 10^{-23}$ | $3,63 \times 10^{-20}$ |
| GLS2     | 2,71              | 0,69               | 90,70  | $4,74 \times 10^{-14}$ | $3,44 \times 10^{-12}$ |
| EIF3I    | 1,90              | 6,84               | 66,99  | $1,17 \times 10^{-11}$ | $4,56 \times 10^{-10}$ |
| ZSWIM8   | -2,67             | 6,54               | 61,67  | $4,61 \times 10^{-11}$ | $1,57 \times 10^{-9}$  |
| ITGA11   | 2,90              | 2,34               | 55,86  | $2,21 \times 10^{-10}$ | $6,15 \times 10^{-9}$  |
| NDST2    | -2,00             | 4,94               | 45,47  | $4,46 \times 10^{-9}$  | $8,74 \times 10^{-8}$  |
| DDX18    | 1,54              | 6,04               | 43,32  | $8,63 \times 10^{-9}$  | $1,55 \times 10^{-7}$  |
| DAB2IP   | -1,64             | 7,38               | 39,51  | $2,87 \times 10^{-8}$  | $4,46 \times 10^{-7}$  |
| DNAH17   | -2,04             | 2,07               | 34,93  | $1,29 \times 10^{-7}$  | $1,69 \times 10^{-6}$  |
| EXD3     | -2,13             | 4,80               | 33,37  | $2,19 \times 10^{-7}$  | $2,69 \times 10^{-6}$  |
| RUBCNL   | 2,55              | 4,91               | 32,13  | $3,36 \times 10^{-7}$  | $3,92 \times 10^{-6}$  |
| MUC4     | -4,95             | 8,48               | 30,22  | $6,58 \times 10^{-7}$  | $7,10 \times 10^{-6}$  |
| OGA      | -1,71             | 7,40               | 30,20  | $6,61 \times 10^{-7}$  | $7,12 \times 10^{-6}$  |
| TTN      | -3,57             | 4,37               | 27,51  | $1,74 \times 10^{-6}$  | $1,67 \times 10^{-5}$  |
| CCDC57   | -1,89             | 5,04               | 27,33  | $1,86 \times 10^{-6}$  | $1,77 \times 10^{-5}$  |
| NRAS     | 1,56              | 6,08               | 24,72  | $4,92 \times 10^{-6}$  | $4,17 \times 10^{-5}$  |
| CCDC17   | -3,88             | 1,51               | 23,77  | $7,05 \times 10^{-6}$  | $5,72 \times 10^{-5}$  |
| FRMD4A   | -1,79             | 4,38               | 22,59  | $1,11 \times 10^{-5}$  | $8,57 \times 10^{-5}$  |
| EXOC3L4  | -4,03             | 3,21               | 22,42  | $1,19 \times 10^{-5}$  | $9,05 \times 10^{-5}$  |
| SEZ6     | -8,02             | -1,01              | 22,15  | $1,32 \times 10^{-5}$  | $9,92 \times 10^{-5}$  |
| ADAMTSL1 | -2,50             | 3,76               | 22,09  | $1,35 \times 10^{-5}$  | $1,01 \times 10^{-4}$  |
| STK3     | 1,62              | 3,30               | 20,49  | $2,54 \times 10^{-5}$  | $1,77 \times 10^{-4}$  |
| GABRE    | -2,37             | 4,52               | 18,85  | $4,93 \times 10^{-5}$  | $3,15 \times 10^{-5}$  |
| LMNTD2   | -2,74             | 3,33               | 17,12  | $1,01 \times 10^{-4}$  | $5,78 \times 10^{-4}$  |
| KCNH6    | -5,77             | 0,11               | 16,85  | $1,13 \times 10^{-4}$  | $6,36 \times 10^{-4}$  |
| OBSCN    | -2,33             | 4,90               | 16,84  | $1,13 \times 10^{-4}$  | $6,38 \times 10^{-4}$  |
| SLC27A6  | -5,11             | -0,32              | 16,59  | $1,26 \times 10^{-4}$  | $6,99 \times 10^{-4}$  |
| ITGA8    | -2,49             | 3,84               | 16,51  | $1,30 \times 10^{-4}$  | $7,19 \times 10^{-4}$  |
| CCDC180  | -2,92             | 2,52               | 16,07  | $1,57 \times 10^{-4}$  | $8,47 \times 10^{-4}$  |
| ATN1     | -1,88             | 7,55               | 15,48  | $2,02 \times 10^{-4}$  | $1,05 \times 10^{-3}$  |
| DCDC2B   | -2,03             | 0,49               | 15,15  | $2,32 \times 10^{-4}$  | $1,19 \times 10^{-3}$  |
| FER1L6   | -5,05             | 4,76               | 12,51  | $7,42 \times 10^{-4}$  | $3,20 \times 10^{-3}$  |
| GNMT     | -2,97             | -1,62              | 11,69  | $1,08 \times 10^{-3}$  | $4,38 \times 10^{-3}$  |
| GATA2    | -2,12             | 3,18               | 10,75  | $1,66 \times 10^{-3}$  | $6,32 \times 10^{-3}$  |
| FTCD     | -5,68             | -1,07              | 8,92   | $3,95 \times 10^{-3}$  | $1,31 \times 10^{-2}$  |
| LRRC34   | 1,55              | -0,59              | 8,82   | $4,13 \times 10^{-3}$  | $1,36 \times 10^{-2}$  |
| PEX6     | -2,05             | 5,09               | 7,72   | $7,09 \times 10^{-3}$  | $2,11 \times 10^{-2}$  |
| ABI3BP   | -3,07             | 4,91               | 7,62   | $7,47 \times 10^{-3}$  | $2,20 \times 10^{-2}$  |
| PRDM16   | -2,53             | 1,35               | 7,54   | $7,76 \times 10^{-3}$  | $2,27 \times 10^{-2}$  |
| ASIC4    | -6,39             | -3,20              | 7,42   | $8,22 \times 10^{-3}$  | $2,38 \times 10^{-2}$  |
| ABCA3    | -2,49             | 3,32               | 7,31   | $8,70 \times 10^{-3}$  | $2,49 \times 10^{-2}$  |
| WDR17    | -3,04             | 1,23               | 6,23   | $1,50 \times 10^{-2}$  | $3,90 \times 10^{-2}$  |

## Continuación tabla suplementaria 74.

|                 | logFC | logCPM | F    | p-valor               | FDR                   |
|-----------------|-------|--------|------|-----------------------|-----------------------|
| <i>KIF5C</i>    | -2,94 | 3,17   | 6,11 | 1,60x10 <sup>-2</sup> | 4,09x10 <sup>-2</sup> |
| <i>ARHGAP10</i> | -2,46 | 4,23   | 5,65 | 2,04x10 <sup>-2</sup> | 4,98x10 <sup>-2</sup> |

**Tabla suplementaria 75. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-7.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | logFC | logCPM | F      | p-valor                | FDR                    |
|----------------|-------|--------|--------|------------------------|------------------------|
| <i>CADPS</i>   | 3,96  | 2,06   | 143,68 | 2,75x10 <sup>-18</sup> | 7,85x10 <sup>-16</sup> |
| <i>TBCD</i>    | 1,60  | 5,66   | 68,52  | 7,90x10 <sup>-12</sup> | 4,83x10 <sup>-10</sup> |
| <i>ADRM1</i>   | 1,96  | 5,99   | 63,31  | 2,99x10 <sup>-11</sup> | 1,55x10 <sup>-9</sup>  |
| <i>PODNL1</i>  | 2,65  | 1,65   | 58,11  | 1,19x10 <sup>-10</sup> | 5,22x10 <sup>-9</sup>  |
| <i>TREX2</i>   | 1,80  | 3,28   | 49,37  | 1,39x10 <sup>-9</sup>  | 4,54x10 <sup>-8</sup>  |
| <i>GBA2</i>    | -2,01 | 7,67   | 48,81  | 1,64x10 <sup>-9</sup>  | 5,26x10 <sup>-8</sup>  |
| <i>ABCA8</i>   | -7,10 | 5,05   | 47,28  | 2,59x10 <sup>-9</sup>  | 7,81x10 <sup>-8</sup>  |
| <i>RRS1</i>    | 2,72  | 4,13   | 46,33  | 3,44x10 <sup>-9</sup>  | 1,00x10 <sup>-7</sup>  |
| <i>MYOT</i>    | -7,73 | 1,41   | 43,40  | 8,41x10 <sup>-9</sup>  | 2,18x10 <sup>-7</sup>  |
| <i>VWA2</i>    | 4,46  | 0,96   | 40,58  | 2,03x10 <sup>-8</sup>  | 4,63x10 <sup>-7</sup>  |
| <i>HAUS7</i>   | 1,61  | 3,52   | 38,31  | 4,21x10 <sup>-8</sup>  | 8,64x10 <sup>-7</sup>  |
| <i>VAR51</i>   | 1,68  | 5,97   | 35,66  | 1,01x10 <sup>-7</sup>  | 1,83x10 <sup>-6</sup>  |
| <i>PLXND1</i>  | 1,64  | 5,55   | 26,39  | 2,63x10 <sup>-6</sup>  | 3,11x10 <sup>-5</sup>  |
| <i>CPEB3</i>   | -2,24 | 3,41   | 22,72  | 1,06x10 <sup>-5</sup>  | 1,04x10 <sup>-4</sup>  |
| <i>POLR1B</i>  | 1,53  | 4,22   | 20,99  | 2,08x10 <sup>-5</sup>  | 1,87x10 <sup>-4</sup>  |
| <i>MMRN1</i>   | -6,01 | 3,93   | 17,44  | 8,80x10 <sup>-5</sup>  | 6,34x10 <sup>-4</sup>  |
| <i>ASXL3</i>   | -4,14 | 2,32   | 16,27  | 1,44x10 <sup>-4</sup>  | 9,56x10 <sup>-4</sup>  |
| <i>CHODL</i>   | -4,36 | 0,19   | 15,60  | 1,91x10 <sup>-4</sup>  | 1,22x10 <sup>-3</sup>  |
| <i>ADAM33</i>  | -6,54 | 6,11   | 15,13  | 2,34x10 <sup>-4</sup>  | 1,44x10 <sup>-3</sup>  |
| <i>VPS13A</i>  | -1,91 | 6,60   | 14,73  | 2,79x10 <sup>-4</sup>  | 1,66x10 <sup>-3</sup>  |
| <i>PITPNM3</i> | -1,64 | 5,37   | 14,55  | 3,01x10 <sup>-4</sup>  | 1,78x10 <sup>-3</sup>  |
| <i>ZNF91</i>   | -1,58 | 5,36   | 14,53  | 3,04x10 <sup>-4</sup>  | 1,79x10 <sup>-3</sup>  |
| <i>SULT1A2</i> | -4,37 | 3,98   | 13,97  | 3,87x10 <sup>-4</sup>  | 2,20x10 <sup>-3</sup>  |
| <i>ROBO2</i>   | 2,47  | 0,63   | 13,46  | 4,86x10 <sup>-4</sup>  | 2,67x10 <sup>-3</sup>  |
| <i>RAPGEF3</i> | -2,00 | 4,32   | 12,20  | 8,52x10 <sup>-4</sup>  | 4,23x10 <sup>-3</sup>  |
| <i>LCAT</i>    | -2,19 | 3,12   | 11,25  | 1,31x10 <sup>-3</sup>  | 6,08x10 <sup>-3</sup>  |
| <i>HTRA3</i>   | 1,94  | 3,90   | 11,14  | 1,39x10 <sup>-3</sup>  | 6,33x10 <sup>-3</sup>  |
| <i>TTBK2</i>   | -1,68 | 4,01   | 10,92  | 1,53x10 <sup>-3</sup>  | 6,87x10 <sup>-3</sup>  |
| <i>FAM13A</i>  | -1,58 | 5,87   | 10,72  | 1,68x10 <sup>-3</sup>  | 7,40x10 <sup>-3</sup>  |
| <i>CDH23</i>   | -2,33 | 2,55   | 10,43  | 1,92x10 <sup>-3</sup>  | 8,23x10 <sup>-3</sup>  |
| <i>CPA3</i>    | -2,92 | 4,20   | 8,82   | 4,14x10 <sup>-3</sup>  | 1,55x10 <sup>-2</sup>  |
| <i>NPY</i>     | -8,00 | -0,93  | 7,93   | 6,38x10 <sup>-3</sup>  | 2,20x10 <sup>-2</sup>  |
| <i>SLC26A9</i> | 2,86  | -3,85  | 7,49   | 7,96x10 <sup>-3</sup>  | 2,62x10 <sup>-2</sup>  |
| <i>ZFR2</i>    | -4,03 | -2,10  | 7,26   | 8,92x10 <sup>-3</sup>  | 2,88x10 <sup>-2</sup>  |
| <i>ALK</i>     | -4,50 | -0,97  | 7,25   | 8,97x10 <sup>-3</sup>  | 2,90x10 <sup>-2</sup>  |
| <i>AQP8</i>    | -6,88 | 8,54   | 6,84   | 1,10x10 <sup>-2</sup>  | 3,41x10 <sup>-2</sup>  |

**Tabla suplementaria 76. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-8.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón;  $F$ : estadístico  $F$  del  $F$  test de cuasiverosimilitud;  $p$ -valor:  $p$ -valor nominal derivado del valor  $F$  sin ningún tipo de corrección múltiple;  $\text{FDR}$ :  $p$ -valor con corrección Benjamin-Hochberg.

|                | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | $F$    | $p$ -valor             | $\text{FDR}$           |
|----------------|-------------------|--------------------|--------|------------------------|------------------------|
| <i>STAU1</i>   | 2,33              | 6,64               | 267,93 | $4,99 \times 10^{-25}$ | $5,06 \times 10^{-22}$ |
| <i>ITGA11</i>  | 4,78              | 2,73               | 212,68 | $2,08 \times 10^{-22}$ | $1,30 \times 10^{-19}$ |
| <i>BCL2L1</i>  | 2,37              | 6,22               | 122,81 | $9,24 \times 10^{-17}$ | $1,86 \times 10^{-14}$ |
| <i>PLA2G4B</i> | -3,41             | 4,31               | 80,74  | $4,29 \times 10^{-13}$ | $3,39 \times 10^{-11}$ |
| <i>ATIC</i>    | 1,78              | 5,80               | 67,24  | $1,10 \times 10^{-11}$ | $6,06 \times 10^{-10}$ |
| <i>CDYL2</i>   | 2,49              | 2,63               | 62,24  | $3,96 \times 10^{-11}$ | $1,86 \times 10^{-9}$  |
| <i>ROBO2</i>   | 4,24              | 0,75               | 58,68  | $1,02 \times 10^{-10}$ | $4,34 \times 10^{-9}$  |
| <i>DKC1</i>    | 2,18              | 5,69               | 48,12  | $2,02 \times 10^{-9}$  | $6,03 \times 10^{-8}$  |
| <i>DNAH1</i>   | -3,66             | 5,06               | 42,60  | $1,08 \times 10^{-8}$  | $2,60 \times 10^{-7}$  |
| <i>UGGT2</i>   | 1,68              | 3,85               | 41,82  | $1,38 \times 10^{-8}$  | $3,23 \times 10^{-7}$  |
| <i>NRBP2</i>   | -3,52             | 6,23               | 40,63  | $2,01 \times 10^{-8}$  | $4,45 \times 10^{-7}$  |
| <i>CMSS1</i>   | 2,03              | 2,81               | 40,26  | $2,26 \times 10^{-8}$  | $4,95 \times 10^{-7}$  |
| <i>TOP3B</i>   | -2,04             | 4,57               | 36,98  | $6,54 \times 10^{-8}$  | $1,29 \times 10^{-6}$  |
| <i>IRS2</i>    | 2,63              | 5,32               | 36,71  | $7,15 \times 10^{-8}$  | $1,40 \times 10^{-6}$  |
| <i>PAN2</i>    | -1,91             | 5,95               | 32,40  | $3,07 \times 10^{-7}$  | $4,99 \times 10^{-6}$  |
| <i>CDKL1</i>   | -1,91             | 3,06               | 25,96  | $3,09 \times 10^{-6}$  | $3,67 \times 10^{-5}$  |
| <i>PLEKHH1</i> | -2,21             | 6,44               | 25,56  | $3,58 \times 10^{-6}$  | $4,17 \times 10^{-5}$  |
| <i>PTPRCAP</i> | -3,54             | 4,00               | 21,48  | $1,72 \times 10^{-5}$  | $1,59 \times 10^{-4}$  |
| <i>TPH1</i>    | -7,19             | 4,54               | 21,24  | $1,88 \times 10^{-5}$  | $1,71 \times 10^{-4}$  |
| <i>TNS2</i>    | -2,28             | 6,76               | 19,22  | $4,24 \times 10^{-5}$  | $3,41 \times 10^{-4}$  |
| <i>IL16</i>    | -2,44             | 4,99               | 18,17  | $6,51 \times 10^{-5}$  | $4,91 \times 10^{-4}$  |
| <i>LCN12</i>   | -3,34             | 1,55               | 18,15  | $6,58 \times 10^{-5}$  | $4,95 \times 10^{-4}$  |
| <i>IFFO1</i>   | -2,17             | 4,12               | 17,47  | $8,72 \times 10^{-5}$  | $6,26 \times 10^{-4}$  |
| <i>SEC24D</i>  | 1,56              | 5,73               | 17,06  | $1,03 \times 10^{-4}$  | $7,23 \times 10^{-4}$  |
| <i>CSMD2</i>   | 2,10              | -0,37              | 15,24  | $2,24 \times 10^{-4}$  | $1,37 \times 10^{-3}$  |
| <i>CEP290</i>  | -1,52             | 4,21               | 15,12  | $2,36 \times 10^{-4}$  | $1,43 \times 10^{-3}$  |
| <i>CYP3A7</i>  | 2,76              | -0,83              | 14,12  | $3,64 \times 10^{-4}$  | $2,06 \times 10^{-3}$  |
| <i>EPDR1</i>   | 1,73              | 3,81               | 13,36  | $5,08 \times 10^{-4}$  | $2,71 \times 10^{-3}$  |
| <i>LRRC69</i>  | -2,20             | -0,83              | 12,39  | $7,84 \times 10^{-4}$  | $3,92 \times 10^{-3}$  |
| <i>STAG3</i>   | -2,08             | 2,59               | 11,36  | $1,25 \times 10^{-3}$  | $5,75 \times 10^{-3}$  |
| <i>USH2A</i>   | -3,37             | -1,73              | 10,81  | $1,62 \times 10^{-3}$  | $7,10 \times 10^{-3}$  |
| <i>AEBP1</i>   | 1,75              | 7,07               | 10,80  | $1,62 \times 10^{-3}$  | $7,12 \times 10^{-3}$  |
| <i>SH2D7</i>   | -2,78             | 2,47               | 8,85   | $4,07 \times 10^{-3}$  | $1,51 \times 10^{-2}$  |
| <i>AVPR1B</i>  | -3,34             | -1,58              | 7,57   | $7,64 \times 10^{-3}$  | $2,52 \times 10^{-2}$  |
| <i>DISP2</i>   | -3,63             | 4,34               | 7,47   | $8,04 \times 10^{-3}$  | $2,63 \times 10^{-2}$  |
| <i>ACADL</i>   | -3,45             | -1,68              | 6,64   | $1,22 \times 10^{-2}$  | $3,67 \times 10^{-2}$  |

**Tabla suplementaria 77. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente COAD-9. Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_{2FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_{2CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.**

|          | $\log_{2FC}$ | $\log_{2CPM}$ | F      | p-valor                | FDR                    |
|----------|--------------|---------------|--------|------------------------|------------------------|
| COL11A1  | 7,29         | 1,02          | 318,18 | $4,66 \times 10^{-27}$ | $1,24 \times 10^{-23}$ |
| GLA      | 2,07         | 3,65          | 109,24 | $1,12 \times 10^{-15}$ | $1,62 \times 10^{-13}$ |
| MTMR6    | 2,09         | 4,88          | 91,33  | $4,15 \times 10^{-14}$ | $4,03 \times 10^{-12}$ |
| RAD21    | 1,69         | 7,18          | 83,23  | $2,44 \times 10^{-13}$ | $1,92 \times 10^{-11}$ |
| ZNF92    | 2,13         | 3,02          | 71,70  | $3,62 \times 10^{-12}$ | $2,01 \times 10^{-10}$ |
| CSAD     | -2,98        | 4,71          | 55,17  | $2,67 \times 10^{-10}$ | $8,67 \times 10^{-9}$  |
| CHSY3    | 3,48         | -0,67         | 52,59  | $5,52 \times 10^{-10}$ | $1,65 \times 10^{-8}$  |
| AIFM3    | -4,27        | 6,44          | 40,52  | $2,08 \times 10^{-8}$  | $4,07 \times 10^{-7}$  |
| VPS13D   | -1,61        | 4,49          | 39,08  | $3,29 \times 10^{-8}$  | $6,07 \times 10^{-7}$  |
| TTN      | -3,98        | 4,37          | 32,01  | $3,50 \times 10^{-7}$  | $4,63 \times 10^{-6}$  |
| MASTL    | 1,56         | 3,29          | 29,47  | $8,58 \times 10^{-7}$  | $1,02 \times 10^{-5}$  |
| EFCAB11  | 1,64         | 1,99          | 28,46  | $1,23 \times 10^{-6}$  | $1,40 \times 10^{-5}$  |
| PTPN18   | -1,73        | 7,00          | 28,41  | $1,26 \times 10^{-6}$  | $1,42 \times 10^{-5}$  |
| DPP4     | 2,20         | 4,84          | 26,89  | $2,19 \times 10^{-6}$  | $2,32 \times 10^{-5}$  |
| PKHD1L1  | -6,07        | 2,45          | 23,12  | $9,04 \times 10^{-6}$  | $7,92 \times 10^{-5}$  |
| ANKK1    | -8,75        | -0,95         | 20,76  | $2,28 \times 10^{-5}$  | $1,77 \times 10^{-4}$  |
| NIFK     | 1,53         | 4,63          | 20,52  | $2,51 \times 10^{-5}$  | $1,93 \times 10^{-4}$  |
| AHNAK    | -2,94        | 10,54         | 20,02  | $3,07 \times 10^{-5}$  | $2,28 \times 10^{-4}$  |
| PITPNM3  | -2,01        | 5,36          | 20,00  | $3,09 \times 10^{-5}$  | $2,30 \times 10^{-4}$  |
| CFAP70   | -2,46        | 1,88          | 19,72  | $3,47 \times 10^{-5}$  | $2,54 \times 10^{-4}$  |
| CACNA1F  | -5,45        | 1,65          | 16,62  | $1,24 \times 10^{-4}$  | $7,50 \times 10^{-4}$  |
| ZNF692   | -2,07        | 4,95          | 16,12  | $1,54 \times 10^{-4}$  | $9,00 \times 10^{-4}$  |
| NEIL3    | 2,41         | 0,63          | 15,99  | $1,62 \times 10^{-4}$  | $9,42 \times 10^{-4}$  |
| FAT2     | -3,21        | 0,68          | 15,76  | $1,79 \times 10^{-4}$  | $1,02 \times 10^{-3}$  |
| SGSM1    | -3,71        | 2,67          | 15,75  | $1,80 \times 10^{-4}$  | $1,03 \times 10^{-3}$  |
| BCO2     | -2,88        | 1,33          | 14,98  | $2,51 \times 10^{-4}$  | $1,36 \times 10^{-3}$  |
| PLEC     | -1,61        | 8,98          | 14,93  | $2,55 \times 10^{-4}$  | $1,38 \times 10^{-3}$  |
| PCED1A   | -1,66        | 5,25          | 14,67  | $2,86 \times 10^{-4}$  | $1,51 \times 10^{-3}$  |
| PCDH17   | 2,10         | 1,46          | 14,58  | $2,98 \times 10^{-4}$  | $1,56 \times 10^{-3}$  |
| UTY      | -2,66        | 3,67          | 13,60  | $4,58 \times 10^{-4}$  | $2,25 \times 10^{-3}$  |
| SNAP91   | -8,98        | -0,24         | 13,50  | $4,77 \times 10^{-4}$  | $2,33 \times 10^{-3}$  |
| COL5A1   | 2,00         | 6,61          | 11,82  | $1,01 \times 10^{-3}$  | $4,37 \times 10^{-3}$  |
| ASTN1    | -4,85        | -0,40         | 11,77  | $1,04 \times 10^{-3}$  | $4,45 \times 10^{-3}$  |
| FIGN     | -2,88        | 2,00          | 11,56  | $1,14 \times 10^{-3}$  | $4,82 \times 10^{-3}$  |
| RAET1E   | 2,26         | -2,22         | 10,65  | $1,74 \times 10^{-3}$  | $6,82 \times 10^{-3}$  |
| SERPINA4 | 3,38         | -3,87         | 10,50  | $1,90 \times 10^{-3}$  | $7,35 \times 10^{-3}$  |
| TMEM171  | -2,81        | 5,50          | 10,27  | $2,08 \times 10^{-3}$  | $7,92 \times 10^{-3}$  |
| CD1A     | 3,38         | -2,60         | 10,05  | $2,31 \times 10^{-3}$  | $8,62 \times 10^{-3}$  |
| HMCN2    | -4,84        | 5,47          | 9,97   | $2,39 \times 10^{-3}$  | $8,88 \times 10^{-3}$  |
| OPLAH    | -1,82        | 5,75          | 9,72   | $2,69 \times 10^{-3}$  | $9,78 \times 10^{-3}$  |
| EFHB     | -7,73        | -1,36         | 9,24   | $3,39 \times 10^{-3}$  | $1,19 \times 10^{-2}$  |
| PPP1R12C | -1,60        | 6,90          | 7,59   | $7,56 \times 10^{-3}$  | $2,32 \times 10^{-2}$  |

Continuación tabla suplementaria 77.

|                       | logFC | logCPM | F    | p-valor               | FDR                   |
|-----------------------|-------|--------|------|-----------------------|-----------------------|
| <i>TMEM52B</i>        | 2,51  | -2,91  | 7,44 | 8,15x10 <sup>-3</sup> | 2,46x10 <sup>-2</sup> |
| <i>CDHR4</i>          | -6,82 | -2,54  | 7,02 | 1,00x10 <sup>-2</sup> | 2,92x10 <sup>-2</sup> |
| <i>KLHL33</i>         | -3,58 | -1,88  | 6,94 | 1,05x10 <sup>-2</sup> | 3,02x10 <sup>-2</sup> |
| <i>SVEP1</i>          | -2,21 | 4,75   | 6,34 | 1,42x10 <sup>-2</sup> | 3,87x10 <sup>-2</sup> |
| <i>PTGES3L-AARSD1</i> | -2,34 | 0,70   | 6,01 | 1,69x10 <sup>-2</sup> | 4,44x10 <sup>-2</sup> |
| <i>SLC6A19</i>        | -9,24 | 5,11   | 6,00 | 1,69x10 <sup>-2</sup> | 4,46x10 <sup>-2</sup> |

**Tabla suplementaria 78. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en la paciente COAD-11.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | logFC | logCPM | F     | p-valor                | FDR                    |
|----------------|-------|--------|-------|------------------------|------------------------|
| <i>GABRP</i>   | 5,84  | -0,60  | 97,79 | 1,08x10 <sup>-14</sup> | 3,23x10 <sup>-12</sup> |
| <i>SPARC</i>   | 3,35  | 8,03   | 96,60 | 1,38x10 <sup>-14</sup> | 3,91x10 <sup>-12</sup> |
| <i>TGFBI</i>   | 3,19  | 6,26   | 65,96 | 1,52x10 <sup>-11</sup> | 2,08x10 <sup>-9</sup>  |
| <i>TUBGCP6</i> | -2,15 | 6,25   | 61,00 | 5,50x10 <sup>-11</sup> | 6,09x10 <sup>-9</sup>  |
| <i>PRRT2</i>   | -5,49 | 4,56   | 53,86 | 3,86x10 <sup>-10</sup> | 3,17x10 <sup>-8</sup>  |
| <i>EMILIN2</i> | 1,82  | 3,94   | 43,80 | 7,44x10 <sup>-9</sup>  | 4,02x10 <sup>-7</sup>  |
| <i>FAM162A</i> | -2,42 | 6,96   | 35,88 | 9,40x10 <sup>-8</sup>  | 3,15x10 <sup>-6</sup>  |
| <i>SHROOM4</i> | 2,08  | 2,93   | 35,55 | 1,13x10 <sup>-7</sup>  | 3,64x10 <sup>-6</sup>  |
| <i>MROH7</i>   | -4,68 | 4,34   | 32,82 | 2,65x10 <sup>-7</sup>  | 7,43x10 <sup>-6</sup>  |
| <i>ADCY6</i>   | -1,95 | 7,65   | 32,52 | 2,95x10 <sup>-7</sup>  | 8,10x10 <sup>-6</sup>  |
| <i>SLC7A11</i> | 3,40  | 2,53   | 25,48 | 3,70x10 <sup>-6</sup>  | 6,43x10 <sup>-5</sup>  |
| <i>MBOAT1</i>  | -1,63 | 6,00   | 21,22 | 1,90x10 <sup>-5</sup>  | 2,48x10 <sup>-4</sup>  |
| <i>ZNF345</i>  | -2,47 | 2,65   | 20,91 | 2,15x10 <sup>-5</sup>  | 2,75x10 <sup>-4</sup>  |
| <i>ZNF536</i>  | -3,24 | 1,54   | 19,85 | 3,28x10 <sup>-5</sup>  | 3,85x10 <sup>-4</sup>  |
| <i>VWA5B2</i>  | -4,49 | 1,61   | 19,54 | 3,72x10 <sup>-5</sup>  | 4,26x10 <sup>-4</sup>  |
| <i>ATG4D</i>   | -2,08 | 5,68   | 18,04 | 6,87x10 <sup>-5</sup>  | 7,06x10 <sup>-4</sup>  |
| <i>HOOK2</i>   | -1,82 | 5,91   | 17,90 | 7,27x10 <sup>-5</sup>  | 7,41x10 <sup>-4</sup>  |
| <i>COL14A1</i> | 2,38  | 5,54   | 17,53 | 8,48x10 <sup>-5</sup>  | 8,33x10 <sup>-4</sup>  |
| <i>IGSF9</i>   | -4,16 | 6,09   | 17,48 | 8,65x10 <sup>-5</sup>  | 8,44x10 <sup>-4</sup>  |
| <i>MYO15A</i>  | -2,70 | 0,68   | 17,03 | 1,05x10 <sup>-4</sup>  | 9,83x10 <sup>-4</sup>  |
| <i>RIMBP2</i>  | -3,00 | 2,20   | 16,26 | 1,45x10 <sup>-4</sup>  | 1,29x10 <sup>-3</sup>  |
| <i>CHP2</i>    | -4,18 | 8,06   | 15,90 | 1,69x10 <sup>-4</sup>  | 1,46x10 <sup>-3</sup>  |
| <i>NGEF</i>    | -2,07 | 5,22   | 15,49 | 2,01x10 <sup>-4</sup>  | 1,67x10 <sup>-3</sup>  |
| <i>SPRED1</i>  | 1,67  | 5,35   | 15,40 | 2,08x10 <sup>-4</sup>  | 1,72x10 <sup>-3</sup>  |
| <i>KLHL31</i>  | -3,34 | 0,37   | 14,85 | 2,65x10 <sup>-4</sup>  | 2,08x10 <sup>-3</sup>  |
| <i>KIF18A</i>  | 2,00  | 1,58   | 14,02 | 3,80x10 <sup>-4</sup>  | 2,80x10 <sup>-3</sup>  |
| <i>PGPEP1</i>  | -1,58 | 6,41   | 13,90 | 4,01x10 <sup>-4</sup>  | 2,92x10 <sup>-3</sup>  |
| <i>ULBP3</i>   | 2,38  | -1,72  | 11,67 | 1,09x10 <sup>-3</sup>  | 6,50x10 <sup>-3</sup>  |
| <i>PLEKHG6</i> | -2,29 | 6,43   | 11,20 | 1,35x10 <sup>-3</sup>  | 7,70x10 <sup>-3</sup>  |
| <i>TTN</i>     | -1,88 | 4,37   | 10,10 | 2,25x10 <sup>-3</sup>  | 1,16x10 <sup>-2</sup>  |
| <i>COL4A4</i>  | 1,72  | 2,12   | 9,53  | 2,95x10 <sup>-3</sup>  | 1,42x10 <sup>-2</sup>  |

Continuación tabla suplementaria 78.

|         | logFC | logCPM | F    | p-valor               | FDR                   |
|---------|-------|--------|------|-----------------------|-----------------------|
| NEURL1  | -2,12 | 5,56   | 9,02 | 3,77x10 <sup>-3</sup> | 1,72x10 <sup>-2</sup> |
| LRRC4B  | -1,68 | 1,34   | 7,86 | 6,62x10 <sup>-3</sup> | 2,65x10 <sup>-2</sup> |
| GDNF    | -3,37 | 2,36   | 7,80 | 6,82x10 <sup>-3</sup> | 2,72x10 <sup>-2</sup> |
| NRXN1   | -3,16 | 2,39   | 7,74 | 7,01x10 <sup>-3</sup> | 2,78x10 <sup>-2</sup> |
| CGN     | -1,59 | 7,85   | 7,74 | 7,01x10 <sup>-3</sup> | 2,78x10 <sup>-2</sup> |
| NEXMIF  | -2,66 | 1,44   | 7,45 | 8,10x10 <sup>-3</sup> | 3,10x10 <sup>-2</sup> |
| SLC6A13 | -5,97 | -3,48  | 7,25 | 8,94x10 <sup>-3</sup> | 3,34x10 <sup>-2</sup> |
| TYMP    | 1,62  | 5,06   | 6,93 | 1,05x10 <sup>-3</sup> | 3,80x10 <sup>-2</sup> |
| STAB2   | -2,83 | 1,27   | 6,40 | 1,38x10 <sup>-2</sup> | 4,67x10 <sup>-2</sup> |

**Tabla suplementaria 79. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente READ-8.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|          | logFC | logCPM | F      | p-valor                | FDR                    |
|----------|-------|--------|--------|------------------------|------------------------|
| P4HA3    | 4,22  | 0,82   | 155,09 | 4,74x10 <sup>-19</sup> | 3,40x10 <sup>-16</sup> |
| COL18A1  | 3,64  | 6,79   | 121,29 | 1,21x10 <sup>-16</sup> | 5,28x10 <sup>-14</sup> |
| KIAA1549 | 2,42  | 2,63   | 70,73  | 4,60x10 <sup>-12</sup> | 6,01x10 <sup>-10</sup> |
| ADAMTS2  | 2,97  | 4,33   | 43,12  | 9,18x10 <sup>-9</sup>  | 4,38x10 <sup>-7</sup>  |
| PAN2     | -1,97 | 5,95   | 34,17  | 1,67x10 <sup>-7</sup>  | 5,14x10 <sup>-6</sup>  |
| MTHFD1L  | 2,79  | 2,93   | 28,60  | 1,17x10 <sup>-7</sup>  | 2,59x10 <sup>-5</sup>  |
| SMPD3    | -2,90 | 6,47   | 27,37  | 1,83x10 <sup>-6</sup>  | 3,70x10 <sup>-5</sup>  |
| CA2      | -5,10 | 9,77   | 24,90  | 4,59x10 <sup>-6</sup>  | 7,88x10 <sup>-5</sup>  |
| RELN     | 2,83  | 1,92   | 24,00  | 6,46x10 <sup>-6</sup>  | 1,04x10 <sup>-4</sup>  |
| NPHP3    | -2,06 | 5,78   | 23,96  | 6,57x10 <sup>-6</sup>  | 1,06x10 <sup>-4</sup>  |
| PCGF3    | -1,51 | 6,56   | 20,15  | 2,92x10 <sup>-5</sup>  | 3,60x10 <sup>-4</sup>  |
| CES2     | -2,97 | 9,18   | 18,83  | 4,97x10 <sup>-5</sup>  | 5,55x10 <sup>-4</sup>  |
| CGREF1   | 2,40  | 2,56   | 18,59  | 5,49x10 <sup>-5</sup>  | 6,01x10 <sup>-4</sup>  |
| IL34     | 2,41  | 1,53   | 17,16  | 9,90x10 <sup>-5</sup>  | 9,76x10 <sup>-4</sup>  |
| SCARA5   | -3,22 | 5,96   | 15,98  | 1,63x10 <sup>-4</sup>  | 1,47x10 <sup>-3</sup>  |
| HSD17B13 | -4,02 | 0,16   | 15,51  | 1,99x10 <sup>-4</sup>  | 1,73x10 <sup>-3</sup>  |
| TTN      | -2,27 | 4,37   | 13,75  | 4,28x10 <sup>-4</sup>  | 3,18x10 <sup>-3</sup>  |
| RRP12    | 1,68  | 4,43   | 13,42  | 4,94x10 <sup>-4</sup>  | 3,59x10 <sup>-3</sup>  |
| SEC16B   | -2,04 | 4,28   | 13,28  | 5,26x10 <sup>-4</sup>  | 3,75x10 <sup>-3</sup>  |
| CRISPLD1 | 3,18  | 0,38   | 12,64  | 7,00x10 <sup>-4</sup>  | 4,73x10 <sup>-3</sup>  |
| HEPHL1   | -7,44 | -1,70  | 11,90  | 9,80x10 <sup>-4</sup>  | 6,22x10 <sup>-3</sup>  |
| PTGDR2   | -4,22 | 4,15   | 11,24  | 1,33x10 <sup>-3</sup>  | 7,88x10 <sup>-3</sup>  |
| MMP10    | 4,39  | -0,58  | 11,24  | 1,33x10 <sup>-3</sup>  | 7,88x10 <sup>-3</sup>  |
| VWA5A    | -1,78 | 6,36   | 10,97  | 1,50x10 <sup>-3</sup>  | 8,67x10 <sup>-3</sup>  |
| RIPOR3   | -1,52 | 3,91   | 10,61  | 1,77x10 <sup>-3</sup>  | 9,89x10 <sup>-3</sup>  |
| UGT1A7   | -5,21 | -1,47  | 10,39  | 1,97x10 <sup>-3</sup>  | 1,08x10 <sup>-2</sup>  |
| STPG3    | -4,27 | 0,05   | 9,62   | 2,82x10 <sup>-3</sup>  | 1,43x10 <sup>-2</sup>  |
| GNG13    | -5,57 | -0,02  | 9,55   | 2,91x10 <sup>-3</sup>  | 1,47x10 <sup>-2</sup>  |
| UGT1A4   | -3,03 | 6,15   | 9,53   | 2,94x10 <sup>-3</sup>  | 1,48x10 <sup>-2</sup>  |
| UGT1A6   | -3,01 | 6,61   | 9,31   | 3,27x10 <sup>-3</sup>  | 1,61x10 <sup>-2</sup>  |



**Continuación tabla suplementaria 79.**

|                 | logFC | logCPM | F    | p-valor               | FDR                   |
|-----------------|-------|--------|------|-----------------------|-----------------------|
| <i>UGT1A10</i>  | -2,92 | 6,95   | 9,23 | 3,40x10 <sup>-3</sup> | 1,65x10 <sup>-2</sup> |
| <i>ADAMTS17</i> | -1,84 | 2,10   | 8,75 | 4,29x10 <sup>-3</sup> | 1,99x10 <sup>-2</sup> |
| <i>IQCN</i>     | -2,54 | 2,13   | 8,33 | 5,26x10 <sup>-3</sup> | 2,32x10 <sup>-2</sup> |
| <i>FER1L6</i>   | 2,06  | 4,80   | 8,17 | 5,69x10 <sup>-3</sup> | 2,47x10 <sup>-2</sup> |
| <i>GP2</i>      | -3,82 | 2,74   | 7,81 | 6,77x10 <sup>-3</sup> | 2,83x10 <sup>-2</sup> |
| <i>NOXO1</i>    | -1,92 | 3,87   | 6,58 | 1,26x10 <sup>-2</sup> | 4,49x10 <sup>-2</sup> |

**Tabla suplementaria 80. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 103. Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.**

|                | logFC | logCPM | F      | p-valor                | FDR                    |
|----------------|-------|--------|--------|------------------------|------------------------|
| <i>KRT80</i>   | 5,94  | -0,28  | 101,22 | 5,37x10 <sup>-15</sup> | 3,00x10 <sup>-12</sup> |
| <i>CMSS1</i>   | 2,88  | 2,90   | 90,09  | 5,40x10 <sup>-14</sup> | 1,98x10 <sup>-11</sup> |
| <i>COL8A1</i>  | 5,19  | 2,90   | 50,74  | 9,39x10 <sup>-10</sup> | 8,83x10 <sup>-8</sup>  |
| <i>PLCB4</i>   | 2,43  | 5,78   | 47,88  | 2,17x10 <sup>-9</sup>  | 1,85x10 <sup>-7</sup>  |
| <i>NCOA7</i>   | 2,83  | 5,24   | 30,84  | 5,28x10 <sup>-7</sup>  | 2,01x10 <sup>-5</sup>  |
| <i>ENDOD1</i>  | -3,06 | 6,92   | 29,99  | 7,13x10 <sup>-7</sup>  | 2,58x10 <sup>-5</sup>  |
| <i>DNAH2</i>   | 2,38  | -0,03  | 28,15  | 1,38x10 <sup>-6</sup>  | 4,52x10 <sup>-5</sup>  |
| <i>OSM</i>     | 4,38  | -0,84  | 22,86  | 1,00x10 <sup>-5</sup>  | 2,32x10 <sup>-4</sup>  |
| <i>SIDT2</i>   | -1,61 | 6,27   | 21,53  | 1,68x10 <sup>-5</sup>  | 3,54x10 <sup>-4</sup>  |
| <i>CLCA1</i>   | -6,82 | 10,39  | 21,50  | 1,70x10 <sup>-5</sup>  | 3,57x10 <sup>-4</sup>  |
| <i>CARD11</i>  | 2,56  | 2,94   | 20,30  | 2,75x10 <sup>-5</sup>  | 5,25x10 <sup>-4</sup>  |
| <i>ASIC3</i>   | -3,89 | 1,62   | 18,17  | 6,52x10 <sup>-5</sup>  | 1,04x10 <sup>-3</sup>  |
| <i>ZNF469</i>  | 2,11  | 0,95   | 17,21  | 9,70x10 <sup>-5</sup>  | 1,43x10 <sup>-3</sup>  |
| <i>ZNF491</i>  | -2,51 | 0,97   | 16,19  | 1,49x10 <sup>-4</sup>  | 1,99x10 <sup>-3</sup>  |
| <i>AKAP9</i>   | -1,74 | 7,48   | 15,83  | 1,74x10 <sup>-4</sup>  | 2,23x10 <sup>-3</sup>  |
| <i>NPY4R</i>   | -4,96 | 3,22   | 15,66  | 1,87x10 <sup>-4</sup>  | 2,36x10 <sup>-3</sup>  |
| <i>BRCA2</i>   | 2,34  | 2,11   | 15,66  | 1,87x10 <sup>-4</sup>  | 2,36x10 <sup>-3</sup>  |
| <i>RSPO4</i>   | 3,36  | -1,70  | 15,30  | 2,18x10 <sup>-4</sup>  | 2,66x10 <sup>-3</sup>  |
| <i>LRRN2</i>   | -4,53 | 4,71   | 14,45  | 3,14x10 <sup>-4</sup>  | 3,52x10 <sup>-3</sup>  |
| <i>PLEKHA4</i> | 2,22  | 3,64   | 13,94  | 3,94x10 <sup>-4</sup>  | 4,20x10 <sup>-3</sup>  |
| <i>SEC14L6</i> | -5,21 | -1,37  | 12,75  | 6,68x10 <sup>-4</sup>  | 6,32x10 <sup>-3</sup>  |
| <i>PBLD</i>    | -3,18 | 6,06   | 12,65  | 6,98x10 <sup>-4</sup>  | 6,54x10 <sup>-3</sup>  |
| <i>TTC39A</i>  | -2,15 | 5,83   | 11,47  | 1,19x10 <sup>-3</sup>  | 9,77x10 <sup>-3</sup>  |
| <i>LRRC75B</i> | -1,52 | 3,48   | 10,23  | 2,12x10 <sup>-3</sup>  | 1,51x10 <sup>-2</sup>  |
| <i>IGSF9</i>   | -2,86 | 6,09   | 9,93   | 2,44x10 <sup>-3</sup>  | 1,69x10 <sup>-2</sup>  |
| <i>CD1C</i>    | -1,53 | 3,16   | 9,25   | 3,36x10 <sup>-3</sup>  | 2,15x10 <sup>-2</sup>  |
| <i>DNAI1</i>   | -5,33 | -2,85  | 7,11   | 9,61x10 <sup>-3</sup>  | 4,60x10 <sup>-2</sup>  |

**Tabla suplementaria 81. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 109.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | F     | p-valor                | FDR                   |
|----------------|-------------------|--------------------|-------|------------------------|-----------------------|
| <i>SLCO1B3</i> | 7,90              | -0,71              | 56,43 | $1,89 \times 10^{-10}$ | $1,80 \times 10^{-8}$ |
| <i>HCAR3</i>   | 5,42              | -1,71              | 43,45 | $8,30 \times 10^{-9}$  | $5,09 \times 10^{-7}$ |
| <i>PLCE1</i>   | -2,42             | 6,83               | 35,22 | $1,17 \times 10^{-7}$  | $5,23 \times 10^{-6}$ |
| <i>PACC1</i>   | 1,91              | 1,61               | 31,64 | $4,00 \times 10^{-7}$  | $1,50 \times 10^{-5}$ |
| <i>GBA2</i>    | -1,53             | 7,67               | 30,51 | $5,94 \times 10^{-7}$  | $2,12 \times 10^{-5}$ |
| <i>NECTIN3</i> | -2,28             | 5,08               | 30,01 | $7,08 \times 10^{-7}$  | $2,47 \times 10^{-5}$ |
| <i>TRIM46</i>  | 2,40              | 0,16               | 27,98 | $1,47 \times 10^{-6}$  | $4,59 \times 10^{-5}$ |
| <i>PTGDR</i>   | -4,55             | 3,99               | 23,52 | $7,76 \times 10^{-6}$  | $1,91 \times 10^{-4}$ |
| <i>CALHM3</i>  | 4,88              | -4,38              | 23,29 | $9,09 \times 10^{-6}$  | $2,18 \times 10^{-4}$ |
| <i>VWCE</i>    | 2,88              | -0,51              | 22,23 | $1,28 \times 10^{-5}$  | $2,88 \times 10^{-4}$ |
| <i>ANO7</i>    | -3,75             | 6,18               | 19,91 | $3,20 \times 10^{-5}$  | $6,07 \times 10^{-4}$ |
| <i>MYCBPAP</i> | -5,41             | 0,20               | 17,53 | $8,49 \times 10^{-5}$  | $1,33 \times 10^{-3}$ |
| <i>RIMKLA</i>  | -3,21             | 4,54               | 17,34 | $9,17 \times 10^{-5}$  | $1,42 \times 10^{-3}$ |
| <i>CNGB3</i>   | 3,54              | -3,81              | 16,36 | $1,39 \times 10^{-4}$  | $2,00 \times 10^{-3}$ |
| <i>GAD1</i>    | 3,23              | -2,64              | 15,85 | $1,72 \times 10^{-4}$  | $2,36 \times 10^{-3}$ |
| <i>BACE2</i>   | 1,74              | 5,75               | 15,77 | $1,78 \times 10^{-4}$  | $2,43 \times 10^{-3}$ |
| <i>KIF26A</i>  | -2,93             | 3,49               | 14,71 | $2,82 \times 10^{-4}$  | $3,52 \times 10^{-3}$ |
| <i>DOCK5</i>   | -1,60             | 6,25               | 14,51 | $3,06 \times 10^{-4}$  | $3,75 \times 10^{-3}$ |
| <i>COL12A1</i> | 3,12              | 5,94               | 14,39 | $3,24 \times 10^{-4}$  | $3,92 \times 10^{-3}$ |
| <i>OTOP2</i>   | -8,26             | 5,29               | 14,28 | $3,39 \times 10^{-4}$  | $4,08 \times 10^{-3}$ |
| <i>IGFBP1</i>  | 4,16              | -3,11              | 14,18 | $3,54 \times 10^{-4}$  | $4,20 \times 10^{-3}$ |
| <i>LYAR</i>    | 1,51              | 3,41               | 12,78 | $6,59 \times 10^{-4}$  | $6,77 \times 10^{-3}$ |
| <i>ABCA10</i>  | -2,41             | 3,03               | 12,42 | $7,75 \times 10^{-4}$  | $7,69 \times 10^{-3}$ |
| <i>DOP1B</i>   | -1,50             | 6,24               | 12,05 | $9,15 \times 10^{-4}$  | $8,74 \times 10^{-3}$ |
| <i>AGFG2</i>   | -1,82             | 6,11               | 11,33 | $1,27 \times 10^{-3}$  | $1,12 \times 10^{-2}$ |
| <i>PCLO</i>    | -2,52             | 3,90               | 10,42 | $1,94 \times 10^{-3}$  | $1,54 \times 10^{-2}$ |
| <i>ANLN</i>    | 2,14              | 3,51               | 10,26 | $2,09 \times 10^{-3}$  | $1,63 \times 10^{-2}$ |
| <i>PROB1</i>   | -2,10             | 2,72               | 9,82  | $2,56 \times 10^{-3}$  | $1,89 \times 10^{-2}$ |
| <i>KDM5D</i>   | -2,03             | 5,55               | 9,79  | $2,61 \times 10^{-3}$  | $1,92 \times 10^{-2}$ |
| <i>CCDC178</i> | -5,60             | -2,42              | 9,76  | $2,64 \times 10^{-3}$  | $1,93 \times 10^{-2}$ |
| <i>HOXA1</i>   | -1,74             | 1,08               | 9,68  | $2,74 \times 10^{-3}$  | $1,98 \times 10^{-2}$ |
| <i>DLL3</i>    | 2,22              | -2,66              | 9,58  | $2,88 \times 10^{-3}$  | $2,06 \times 10^{-2}$ |
| <i>CDH23</i>   | -2,22             | 2,55               | 9,42  | $3,10 \times 10^{-3}$  | $2,18 \times 10^{-2}$ |
| <i>SIGLEC1</i> | -2,09             | 4,91               | 9,01  | $3,78 \times 10^{-3}$  | $2,51 \times 10^{-2}$ |
| <i>APCDD1L</i> | -7,85             | -0,50              | 8,62  | $4,55 \times 10^{-3}$  | $2,90 \times 10^{-2}$ |
| <i>AHNAK</i>   | -1,69             | 10,54              | 8,00  | $6,19 \times 10^{-3}$  | $3,63 \times 10^{-2}$ |
| <i>SLC14A1</i> | -2,03             | 1,34               | 7,96  | $6,29 \times 10^{-3}$  | $3,67 \times 10^{-2}$ |
| <i>XKR6</i>    | -1,51             | 0,61               | 7,46  | $8,09 \times 10^{-3}$  | $4,43 \times 10^{-2}$ |

**Tabla suplementaria 82. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 402.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón;  $F$ : estadístico  $F$  del  $F$  test de cuasiverosimilitud;  $p$ -valor:  $p$ -valor nominal derivado del valor  $F$  sin ningún tipo de corrección múltiple;  $\text{FDR}$ :  $p$ -valor con corrección Benjamin-Hochberg.

|                | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | $F$   | $p$ -valor            | $\text{FDR}$          |
|----------------|-------------------|--------------------|-------|-----------------------|-----------------------|
| <i>SBF1</i>    | -1,61             | 6,87               | 48,34 | $1,89 \times 10^{-9}$ | $7,14 \times 10^{-8}$ |
| <i>POLR1C</i>  | 2,23              | 4,11               | 45,82 | $4,02 \times 10^{-9}$ | $1,37 \times 10^{-7}$ |
| <i>XKRX</i>    | 3,88              | 0,65               | 40,94 | $1,82 \times 10^{-8}$ | $4,88 \times 10^{-7}$ |
| <i>XPOT</i>    | 1,90              | 5,49               | 39,45 | $2,93 \times 10^{-8}$ | $7,35 \times 10^{-7}$ |
| <i>DDN</i>     | 5,42              | -1,81              | 37,55 | $5,42 \times 10^{-8}$ | $1,24 \times 10^{-6}$ |
| <i>ZNF749</i>  | 2,46              | 1,72               | 32,63 | $2,83 \times 10^{-7}$ | $5,16 \times 10^{-6}$ |
| <i>MRPS12</i>  | 1,94              | 4,74               | 31,36 | $4,40 \times 10^{-7}$ | $7,57 \times 10^{-6}$ |
| <i>AHNAK</i>   | -3,92             | 10,53              | 29,04 | $1,00 \times 10^{-6}$ | $1,55 \times 10^{-5}$ |
| <i>TPCN1</i>   | -1,69             | 6,82               | 27,04 | $2,07 \times 10^{-6}$ | $2,87 \times 10^{-5}$ |
| <i>KIF13A</i>  | -2,15             | 6,00               | 25,73 | $3,37 \times 10^{-6}$ | $4,38 \times 10^{-5}$ |
| <i>P4HA2</i>   | -2,44             | 5,38               | 23,29 | $8,48 \times 10^{-6}$ | $9,60 \times 10^{-5}$ |
| <i>SYNE2</i>   | -2,33             | 7,58               | 23,24 | $8,66 \times 10^{-6}$ | $9,79 \times 10^{-5}$ |
| <i>CCN5</i>    | -7,26             | 3,22               | 21,16 | $1,95 \times 10^{-5}$ | $1,95 \times 10^{-5}$ |
| <i>TRPV6</i>   | -6,28             | -1,03              | 18,48 | $5,73 \times 10^{-5}$ | $4,90 \times 10^{-4}$ |
| <i>DNAJA1</i>  | 2,42              | 6,91               | 17,34 | $9,18 \times 10^{-5}$ | $7,28 \times 10^{-4}$ |
| <i>SCN11A</i>  | -5,62             | 0,25               | 16,13 | $1,53 \times 10^{-4}$ | $1,13 \times 10^{-3}$ |
| <i>CBX2</i>    | 2,69              | 0,27               | 15,51 | $1,99 \times 10^{-4}$ | $1,42 \times 10^{-3}$ |
| <i>WNK2</i>    | -1,60             | 7,40               | 15,22 | $2,25 \times 10^{-4}$ | $1,57 \times 10^{-3}$ |
| <i>ABI3BP</i>  | -4,90             | 4,91               | 14,32 | $3,32 \times 10^{-4}$ | $2,16 \times 10^{-3}$ |
| <i>ZNF256</i>  | -1,75             | 1,77               | 14,18 | $3,55 \times 10^{-4}$ | $2,28 \times 10^{-3}$ |
| <i>TRPM5</i>   | -4,09             | 2,96               | 12,20 | $8,53 \times 10^{-4}$ | $4,71 \times 10^{-3}$ |
| <i>ANK3</i>    | -2,41             | 6,03               | 10,68 | $1,71 \times 10^{-3}$ | $8,28 \times 10^{-3}$ |
| <i>NBEAL1</i>  | -2,06             | 5,10               | 10,53 | $1,84 \times 10^{-3}$ | $8,80 \times 10^{-3}$ |
| <i>SLC9A9</i>  | -1,89             | 3,15               | 10,39 | $1,96 \times 10^{-3}$ | $9,24 \times 10^{-3}$ |
| <i>PDE2A</i>   | -2,78             | 4,49               | 10,05 | $2,30 \times 10^{-3}$ | $1,05 \times 10^{-2}$ |
| <i>CHAC1</i>   | 2,30              | 0,86               | 8,93  | $3,92 \times 10^{-3}$ | $1,60 \times 10^{-2}$ |
| <i>WDR64</i>   | -6,04             | -2,97              | 8,67  | $4,44 \times 10^{-3}$ | $1,76 \times 10^{-2}$ |
| <i>ACADL</i>   | -4,57             | -1,69              | 8,29  | $5,34 \times 10^{-3}$ | $2,05 \times 10^{-2}$ |
| <i>PRDM16</i>  | -2,60             | 1,34               | 7,24  | $9,02 \times 10^{-3}$ | $3,12 \times 10^{-2}$ |
| <i>FRMD3</i>   | -1,98             | 4,41               | 7,19  | $9,23 \times 10^{-3}$ | $3,18 \times 10^{-2}$ |
| <i>RECQL4</i>  | 1,51              | 3,95               | 6,97  | $1,03 \times 10^{-2}$ | $3,47 \times 10^{-2}$ |
| <i>CCDC88A</i> | -1,59             | 3,99               | 6,65  | $1,22 \times 10^{-2}$ | $3,94 \times 10^{-2}$ |
| <i>CCDC110</i> | -1,79             | 0,39               | 6,63  | $1,22 \times 10^{-2}$ | $3,96 \times 10^{-2}$ |
| <i>GRIN3A</i>  | -2,90             | 0,90               | 6,56  | $1,27 \times 10^{-2}$ | $4,08 \times 10^{-2}$ |
| <i>CDHR2</i>   | -2,92             | 6,69               | 6,53  | $1,29 \times 10^{-2}$ | $4,12 \times 10^{-2}$ |
| <i>ZNF98</i>   | -6,06             | -3,65              | 6,49  | $1,31 \times 10^{-2}$ | $4,18 \times 10^{-2}$ |
| <i>SV2B</i>    | -2,92             | 0,84               | 6,11  | $1,60 \times 10^{-2}$ | $4,88 \times 10^{-2}$ |

**Tabla suplementaria 83. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente 501.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y  $\text{FDR} < 5\%$  que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes.  $\log_2\text{FC}$ : cambio en la expresión del gen expresada en logaritmo en base 2;  $\log_2\text{CPM}$ : promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                 | $\log_2\text{FC}$ | $\log_2\text{CPM}$ | F     | p-valor                | FDR                    |
|-----------------|-------------------|--------------------|-------|------------------------|------------------------|
| <i>SIPA1L2</i>  | -2,75             | 6,38               | 87,70 | $9,06 \times 10^{-14}$ | $1,32 \times 10^{-11}$ |
| <i>METTL26</i>  | 2,34              | 4,51               | 72,81 | $2,76 \times 10^{-12}$ | $2,74 \times 10^{-10}$ |
| <i>RPUSD4</i>   | 1,68              | 4,36               | 57,37 | $1,46 \times 10^{-10}$ | $8,30 \times 10^{-9}$  |
| <i>UBA2</i>     | 1,78              | 5,82               | 56,90 | $1,66 \times 10^{-10}$ | $9,25 \times 10^{-9}$  |
| <i>POLA1</i>    | 2,24              | 2,84               | 55,09 | $2,73 \times 10^{-10}$ | $1,42 \times 10^{-8}$  |
| <i>COMMD6</i>   | 1,82              | 5,48               | 41,56 | $1,50 \times 10^{-8}$  | $4,33 \times 10^{-7}$  |
| <i>SETMAR</i>   | 1,51              | 3,46               | 40,95 | $1,81 \times 10^{-8}$  | $5,13 \times 10^{-7}$  |
| <i>GPT</i>      | -5,59             | 6,51               | 39,94 | $2,50 \times 10^{-8}$  | $6,69 \times 10^{-7}$  |
| <i>TEP1</i>     | -2,26             | 6,68               | 37,44 | $5,62 \times 10^{-8}$  | $1,35 \times 10^{-6}$  |
| <i>TMEM208</i>  | 1,74              | 4,53               | 36,70 | $7,16 \times 10^{-8}$  | $1,66 \times 10^{-6}$  |
| <i>B3GNTL1</i>  | 1,57              | 2,69               | 35,10 | $1,22 \times 10^{-7}$  | $2,59 \times 10^{-6}$  |
| <i>CSTF2</i>    | 1,81              | 4,23               | 34,00 | $1,77 \times 10^{-7}$  | $3,56 \times 10^{-6}$  |
| <i>PADI2</i>    | -4,75             | 8,53               | 27,52 | $1,74 \times 10^{-6}$  | $2,52 \times 10^{-5}$  |
| <i>DGKA</i>     | -2,20             | 6,06               | 26,89 | $2,19 \times 10^{-6}$  | $3,08 \times 10^{-5}$  |
| <i>SPTB</i>     | -2,37             | 2,99               | 26,65 | $2,39 \times 10^{-6}$  | $3,33 \times 10^{-5}$  |
| <i>SSTR1</i>    | -2,62             | 3,66               | 25,28 | $3,99 \times 10^{-6}$  | $5,10 \times 10^{-5}$  |
| <i>HAGHL</i>    | 1,57              | 3,10               | 23,63 | $7,43 \times 10^{-6}$  | $8,68 \times 10^{-5}$  |
| <i>NOP14</i>    | 1,59              | 5,24               | 23,36 | $8,27 \times 10^{-6}$  | $9,51 \times 10^{-5}$  |
| <i>TNXB</i>     | -6,26             | 5,49               | 21,28 | $1,86 \times 10^{-5}$  | $1,90 \times 10^{-4}$  |
| <i>PTPRF</i>    | -1,88             | 9,01               | 20,98 | $2,09 \times 10^{-5}$  | $2,10 \times 10^{-4}$  |
| <i>MYCBPAP</i>  | -7,80             | 0,19               | 20,43 | $2,60 \times 10^{-5}$  | $2,50 \times 10^{-4}$  |
| <i>MYH3</i>     | -2,53             | 2,38               | 19,63 | $3,59 \times 10^{-5}$  | $3,28 \times 10^{-4}$  |
| <i>ADGRV1</i>   | -2,57             | 1,67               | 18,40 | $5,92 \times 10^{-5}$  | $4,98 \times 10^{-4}$  |
| <i>DAAM2</i>    | -2,51             | 5,48               | 17,17 | $9,87 \times 10^{-5}$  | $7,68 \times 10^{-4}$  |
| <i>SEMA3D</i>   | -4,18             | 1,57               | 16,26 | $1,45 \times 10^{-4}$  | $1,06 \times 10^{-3}$  |
| <i>CAPN9</i>    | -3,97             | 4,97               | 15,32 | $2,16 \times 10^{-4}$  | $1,47 \times 10^{-3}$  |
| <i>MICAL3</i>   | -1,53             | 6,84               | 15,21 | $2,26 \times 10^{-4}$  | $1,53 \times 10^{-3}$  |
| <i>LRP1</i>     | -1,91             | 8,27               | 14,99 | $2,49 \times 10^{-4}$  | $1,65 \times 10^{-3}$  |
| <i>SHROOM3</i>  | -1,76             | 7,50               | 14,19 | $3,52 \times 10^{-4}$  | $2,20 \times 10^{-3}$  |
| <i>CD1D</i>     | -1,89             | 3,16               | 12,54 | $7,33 \times 10^{-4}$  | $4,01 \times 10^{-3}$  |
| <i>ROS1</i>     | -9,84             | 0,77               | 12,52 | $7,39 \times 10^{-4}$  | $4,03 \times 10^{-3}$  |
| <i>RANBP17</i>  | 1,57              | 1,76               | 12,21 | $8,51 \times 10^{-4}$  | $4,55 \times 10^{-3}$  |
| <i>FBLN2</i>    | -3,25             | 5,30               | 10,43 | $1,93 \times 10^{-3}$  | $8,87 \times 10^{-3}$  |
| <i>SLC36A1</i>  | -2,67             | 6,29               | 9,71  | $2,71 \times 10^{-3}$  | $1,16 \times 10^{-2}$  |
| <i>COL19A1</i>  | -4,57             | -0,64              | 9,40  | $3,14 \times 10^{-3}$  | $1,31 \times 10^{-2}$  |
| <i>ABCC4</i>    | 1,59              | 4,99               | 9,20  | $3,45 \times 10^{-3}$  | $1,41 \times 10^{-2}$  |
| <i>WDR4</i>     | 1,60              | 2,91               | 8,94  | $3,91 \times 10^{-3}$  | $1,56 \times 10^{-2}$  |
| <i>DNASE1L2</i> | 2,54              | 0,23               | 8,77  | $4,23 \times 10^{-3}$  | $1,66 \times 10^{-2}$  |
| <i>CYP4F12</i>  | -1,60             | 6,12               | 8,73  | $4,31 \times 10^{-3}$  | $1,68 \times 10^{-2}$  |
| <i>RSP02</i>    | -4,58             | 1,81               | 8,50  | $4,83 \times 10^{-3}$  | $1,84 \times 10^{-2}$  |
| <i>SYTL3</i>    | -2,14             | 2,54               | 8,44  | $4,98 \times 10^{-3}$  | $1,88 \times 10^{-2}$  |
| <i>RASSF4</i>   | -1,51             | 5,26               | 8,27  | $5,41 \times 10^{-3}$  | $2,01 \times 10^{-2}$  |

Continuación tabla suplementaria 83.

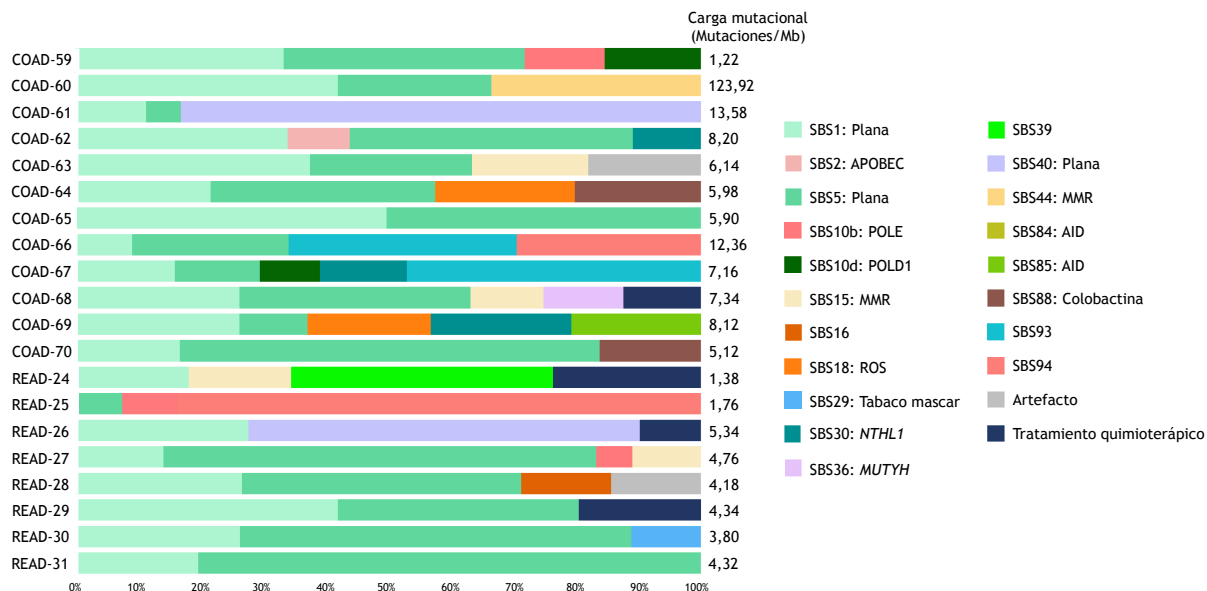
|                | logFC | logCPM | F    | p-valor               | FDR                   |
|----------------|-------|--------|------|-----------------------|-----------------------|
| <i>SHC2</i>    | -2,67 | 3,78   | 8,25 | 5,47x10 <sup>-3</sup> | 2,03x10 <sup>-2</sup> |
| <i>ZNF831</i>  | -3,35 | 1,20   | 8,17 | 5,67x10 <sup>-3</sup> | 2,09x10 <sup>-2</sup> |
| <i>KRT4</i>    | 3,35  | 1,94   | 7,88 | 6,55x10 <sup>-3</sup> | 2,34x10 <sup>-2</sup> |
| <i>NLGN1</i>   | -5,95 | 0,57   | 7,72 | 7,08x10 <sup>-3</sup> | 2,50x10 <sup>-2</sup> |
| <i>APC2</i>    | -1,95 | 1,31   | 7,29 | 8,80x10 <sup>-3</sup> | 2,96x10 <sup>-2</sup> |
| <i>EEF2K</i>   | -1,53 | 5,44   | 7,08 | 9,78x10 <sup>-3</sup> | 3,20x10 <sup>-2</sup> |
| <i>CMKLR1</i>  | -1,74 | 3,67   | 6,85 | 1,10x10 <sup>-2</sup> | 3,51x10 <sup>-2</sup> |
| <i>FAM221B</i> | -3,27 | -1,32  | 6,54 | 1,28x10 <sup>-2</sup> | 3,95x10 <sup>-2</sup> |
| <i>PKD2</i>    | -1,99 | 4,78   | 6,18 | 1,55x10 <sup>-2</sup> | 4,58x10 <sup>-2</sup> |

**Tabla suplementaria 84. Listado de genes procedentes de la aproximación *All vs one* que presentan variantes en línea germinal con frecuencia poblacional rara en el paciente READ-4.** Listado de genes procedentes de la aproximación *All vs one* con  $|\log_2 \text{fold change}| \geq 1,5$  y FDR <5% que muestran alteraciones en línea germinal que cumplen los criterios de filtrado de variantes. logFC: cambio en la expresión del gen expresada en logaritmo en base 2; logCPM: promedio de expresión de todas las muestras para el gen en cuestión expresado en valores logarítmicos en base 2 de contajes por millón; F: estadístico F del F test de cuasiverosimilitud; p-valor: p-valor nominal derivado del valor F sin ningún tipo de corrección múltiple; FDR: p-valor con corrección Benjamin-Hochberg.

|                | logFC | logCPM | F      | p-valor                | FDR                    |
|----------------|-------|--------|--------|------------------------|------------------------|
| <i>CFAP57</i>  | 3,07  | 0,58   | 177,72 | 1,82x10 <sup>-20</sup> | 3,59x10 <sup>-18</sup> |
| <i>CAPZA2</i>  | 1,76  | 6,69   | 93,23  | 2,77x10 <sup>-14</sup> | 1,22x10 <sup>-12</sup> |
| <i>TAF1C</i>   | -2,50 | 5,94   | 72,22  | 3,19x10 <sup>-12</sup> | 9,01x10 <sup>-11</sup> |
| <i>NDUFA13</i> | 1,83  | 6,68   | 70,77  | 4,54x10 <sup>-12</sup> | 1,24x10 <sup>-10</sup> |
| <i>DSG4</i>    | 5,31  | -1,64  | 62,75  | 3,46x10 <sup>-11</sup> | 7,74x10 <sup>-10</sup> |
| <i>CCDC113</i> | 2,67  | 1,78   | 53,15  | 4,71x10 <sup>-10</sup> | 8,28x10 <sup>-9</sup>  |
| <i>TFB2M</i>   | 1,67  | 3,93   | 49,28  | 1,44x10 <sup>-9</sup>  | 2,28x10 <sup>-8</sup>  |
| <i>MPP1</i>    | 1,78  | 4,59   | 47,76  | 2,25x10 <sup>-9</sup>  | 3,46x10 <sup>-8</sup>  |
| <i>RIPOR3</i>  | -3,76 | 3,90   | 44,53  | 5,94x10 <sup>-9</sup>  | 8,27x10 <sup>-8</sup>  |
| <i>SLC4A4</i>  | -5,59 | 7,68   | 36,49  | 7,68x10 <sup>-8</sup>  | 8,31x10 <sup>-7</sup>  |
| <i>YARS1</i>   | 1,88  | 5,73   | 32,97  | 2,51x10 <sup>-7</sup>  | 2,42x10 <sup>-6</sup>  |
| <i>KCTD1</i>   | -3,54 | 2,78   | 32,61  | 2,85x10 <sup>-7</sup>  | 2,71x10 <sup>-6</sup>  |
| <i>PLCE1</i>   | -2,19 | 6,83   | 30,35  | 6,28x10 <sup>-7</sup>  | 5,50x10 <sup>-6</sup>  |
| <i>THEM6</i>   | 1,64  | 4,79   | 29,53  | 8,39x10 <sup>-7</sup>  | 7,09x10 <sup>-6</sup>  |
| <i>UBA7</i>    | -2,54 | 6,21   | 26,96  | 2,13x10 <sup>-6</sup>  | 1,63x10 <sup>-5</sup>  |
| <i>TTN</i>     | -3,36 | 4,36   | 25,03  | 4,37x10 <sup>-6</sup>  | 3,12x10 <sup>-5</sup>  |
| <i>LRP3</i>    | 1,69  | 4,41   | 21,40  | 1,77x10 <sup>-5</sup>  | 1,09x10 <sup>-4</sup>  |
| <i>COL11A1</i> | 2,87  | -0,71  | 21,24  | 1,89x10 <sup>-5</sup>  | 1,15x10 <sup>-4</sup>  |
| <i>RRP12</i>   | 2,00  | 4,44   | 20,40  | 2,63x10 <sup>-5</sup>  | 1,55x10 <sup>-4</sup>  |
| <i>TTL6</i>    | -4,09 | 4,10   | 20,12  | 2,95x10 <sup>-5</sup>  | 1,71x10 <sup>-4</sup>  |
| <i>BRCA1</i>   | 2,00  | 3,43   | 19,67  | 3,53x10 <sup>-5</sup>  | 2,01x10 <sup>-4</sup>  |
| <i>TCTE1</i>   | -5,48 | -1,27  | 19,00  | 4,63x10 <sup>-5</sup>  | 2,56x10 <sup>-4</sup>  |
| <i>ALS2CL</i>  | -1,84 | 6,37   | 17,85  | 7,42x10 <sup>-5</sup>  | 3,90x10 <sup>-4</sup>  |
| <i>CHRNE</i>   | -1,94 | 3,40   | 16,10  | 1,55x10 <sup>-4</sup>  | 7,45x10 <sup>-4</sup>  |
| <i>CHRD</i>    | -2,33 | 3,05   | 15,62  | 1,90x10 <sup>-4</sup>  | 8,92x10 <sup>-4</sup>  |
| <i>CCDC180</i> | -2,86 | 2,51   | 15,49  | 2,00x10 <sup>-4</sup>  | 9,36x10 <sup>-4</sup>  |
| <i>ZNF596</i>  | -1,58 | 1,95   | 14,55  | 3,02x10 <sup>-4</sup>  | 1,33x10 <sup>-3</sup>  |
| <i>CHST6</i>   | -4,56 | 0,52   | 14,48  | 3,10x10 <sup>-4</sup>  | 1,37x10 <sup>-3</sup>  |
| <i>SLC27A6</i> | -4,52 | -0,32  | 14,38  | 3,25x10 <sup>-4</sup>  | 1,42x10 <sup>-3</sup>  |

Continuación tabla suplementaria 84.

|                | logFC | logCPM | F     | p-valor               | FDR                   |
|----------------|-------|--------|-------|-----------------------|-----------------------|
| <i>CHL1</i>    | -3,27 | 3,93   | 13,53 | 4,71x10 <sup>-4</sup> | 1,96x10 <sup>-3</sup> |
| <i>UPK3B</i>   | -4,22 | 0,37   | 12,63 | 7,05x10 <sup>-4</sup> | 2,78x10 <sup>-3</sup> |
| <i>CMYA5</i>   | -4,04 | 1,96   | 12,28 | 8,23x10 <sup>-4</sup> | 3,18x10 <sup>-3</sup> |
| <i>COL6A6</i>  | -6,57 | -2,71  | 11,32 | 1,27x10 <sup>-3</sup> | 4,65x10 <sup>-3</sup> |
| <i>FLRT1</i>   | -3,46 | 1,86   | 11,00 | 1,48x10 <sup>-3</sup> | 5,28x10 <sup>-3</sup> |
| <i>BLM</i>     | 1,71  | 2,44   | 10,67 | 1,72x10 <sup>-3</sup> | 6,01x10 <sup>-3</sup> |
| <i>ZNF423</i>  | -2,36 | 2,86   | 10,56 | 1,82x10 <sup>-3</sup> | 6,30x10 <sup>-3</sup> |
| <i>CYP21A2</i> | -8,79 | -1,00  | 10,47 | 1,89x10 <sup>-3</sup> | 6,51x10 <sup>-3</sup> |
| <i>STEAP4</i>  | -3,40 | 3,61   | 9,98  | 2,38x10 <sup>-3</sup> | 7,92x10 <sup>-3</sup> |
| <i>SLC7A14</i> | -6,40 | -0,04  | 9,93  | 2,44x10 <sup>-3</sup> | 8,10x10 <sup>-3</sup> |
| <i>GPR146</i>  | -2,16 | 2,54   | 9,42  | 3,10x10 <sup>-3</sup> | 9,95x10 <sup>-3</sup> |
| <i>SOSTDC1</i> | -2,62 | 2,77   | 9,38  | 3,16x10 <sup>-3</sup> | 1,01x10 <sup>-2</sup> |
| <i>DOCK8</i>   | -1,69 | 4,56   | 8,97  | 3,85x10 <sup>-3</sup> | 1,20x10 <sup>-2</sup> |
| <i>CGN</i>     | -1,66 | 7,85   | 8,27  | 5,42x10 <sup>-3</sup> | 1,61x10 <sup>-2</sup> |
| <i>NACAD</i>   | -3,66 | 2,27   | 8,01  | 6,15x10 <sup>-3</sup> | 1,79x10 <sup>-2</sup> |
| <i>EFS</i>     | -2,67 | 3,31   | 7,54  | 7,75x10 <sup>-3</sup> | 2,16x10 <sup>-2</sup> |
| <i>XAF1</i>    | -2,35 | 4,88   | 6,84  | 1,10x10 <sup>-2</sup> | 2,90x10 <sup>-2</sup> |
| <i>VWA3A</i>   | -4,23 | -1,39  | 6,80  | 1,13x10 <sup>-2</sup> | 2,96x10 <sup>-2</sup> |
| <i>HPR</i>     | -7,52 | -2,05  | 6,25  | 1,49x10 <sup>-2</sup> | 3,72x10 <sup>-2</sup> |
| <i>ADARB1</i>  | -2,99 | 4,66   | 6,12  | 1,59x10 <sup>-2</sup> | 3,93x10 <sup>-2</sup> |
| <i>NFASC</i>   | -2,16 | 5,08   | 6,07  | 1,63x10 <sup>-2</sup> | 4,02x10 <sup>-2</sup> |
| <i>UNC5D</i>   | -2,93 | -0,02  | 5,93  | 1,76x10 <sup>-2</sup> | 4,26x10 <sup>-2</sup> |



**Figura suplementaria 1. Patrón de firmas mutacionales de nucleótido único (COSMIC v3.3<sup>19</sup>) y carga mutacional tumoral de los 20 tumores procedentes de la base de datos TCGA (cohortes COAD y READ) que forman parte de la fase de réplica. El tumor del paciente COAD-60 muestra un elevado porcentaje de firma SBS44 y elevada carga mutacional tumoral, ambos compatibles con defectos en el sistema de reparación de errores de emparejamiento de bases o MMR (*MisMatch Repair*). Los patrones de firma y/o carga mutacional tumoral de los CCR de los pacientes COAD-61, COAD-63, COAD-64 y COAD-69 no confirman el fenotipo tumoral recogido en la base de datos en el que se indican defectos en el sistema de reparación MMR. Firmas mutacionales presentes: firmas planas (SBS1, SBS5, SBS40), actividad de familia APOBEC de citidina desaminasas (SBS2), mutación en ED polimerasa  $\epsilon$  (SBS10b) y polimerasa  $\delta 1$  (SBS10d), defectos en sistema MMR (SBS15, SBS44), daño por especies reactivas del oxígeno (SBS18), tabaco de mascar (SBS29), defectos en BER por alteraciones bialélicas en *NTHL1* (SBS30) y *MUTYH* (SBS36), actividad desaminasa de citidina (SBS84, SBS85), exposición a colibactina (SBS88) y etiología desconocida (SBS16, SBS39, SBS93, SBS94).**

**Tabla suplementaria 85. Patrones de firma mutacional de los CCR de los 20 pacientes procedentes de la base de datos TCGA (cohortes COAD y READ) que forman parte de la fase de réplica obtenidas empleando SigProfilerAssignment<sup>94-96</sup> y tomando como referencia la base de datos COSMIC v3.3<sup>19</sup>.**

|   | COAD-59 | COAD-60 | COAD-61 | COAD-62 | COAD-63 | COAD-64 | COAD-65 |
|---|---------|---------|---------|---------|---------|---------|---------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 33,33%  | 41,98%  | 11,35%  | 34,01%  | 37,56%  | 21,72%  | 49,74%  |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |         |         |         | 9,86%   |         |         |         |
| SBS5: Origen desconocido (firma plana)  | 38,46%  | 24,55%  | 5,58%   | 45,24%  | 25,85%  | 35,86%  | 50,26%  |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          | 12,82%  |         |         |         |         |         |         |
| SBS10d: Defectos en la lectura de prueba de POLD1   | 15,38%  |         |         |         |         |         |         |
| SBS15: Sistema de reparación MMR defectuoso   |         |         |         |         | 18,54%  |         |         |
| SBS16: Origen desconocido   |         |         |         |         |         |         |         |
| SBS18: Daño por especies reactivas del oxígeno  |         |         |         |         |         | 22,22%  |         |
| SBS29: Tabaco de mascar   |         |         |         |         |         |         |         |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |         |         |         | 10,88%  |         |         |         |
| SBS36: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>MUTYH</i> |         |         |         |         |         |         |         |
| SBS39: Origen desconocido   |         |         |         |         |         |         |         |
| SBS40: Origen desconocido (firma plana)   |         |         | 83,08%  |         |         |         |         |
| SBS44: Sistema de reparación MMR defectuoso   |         | 33,47%  |         |         |         |         |         |
| SBS84: Actividad desaminasa de citidina inducida por activación (AID)                             |         |         |         |         |         |         |         |
| SBS85: Efectos indirectos de desaminasa de citidina inducida por actividad (AID)                  |         |         |         |         |         |         |         |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |         |         |         |         |         | 20,20%  |         |
| SBS93: Origen desconocido   |         |         |         |         |         |         |         |
| SBS94: Origen desconocido   |         |         |         |         |         |         |         |
| Posible artefacto de secuenciación  |         |         |         |         | 18,05%  |         |         |
| Tratamiento quimioterápico  |         |         |         |         |         |         |         |



## Continuación tabla suplementaria 85.

|   | COAD-66 | COAD-67 | COAD-68 | COAD-69 | COAD-70 | READ-24 |
|---|---------|---------|---------|---------|---------|---------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 9,17%   | 16,01%  | 26,32%  | 26,32%  | 16,76%  | 18,18%  |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |         |         |         |         |         |         |
| SBS5: Origen desconocido (firma plana)  | 25,00%  | 13,52%  | 36,84%  | 10,88%  | 67,05%  |         |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |         |         |         |         |         |         |
| SBS10d: Defectos en la lectura de prueba de POLD1   |         | 9,61%   |         |         |         |         |
| SBS15: Sistema de reparación MMR defectuoso   |         |         | 11,65%  |         |         | 16,36%  |
| SBS16: Origen desconocido   |         |         |         |         |         |         |
| SBS18: Daño por especies reactivas del oxígeno  |         |         |         | 19,65%  |         |         |
| SBS29: Tabaco de mascar   |         |         |         |         |         |         |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |         | 13,88%  |         | 22,46%  |         |         |
| SBS36: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>MUTYH</i> |         |         | 12,78%  |         |         |         |
| SBS39: Origen desconocido   |         |         |         |         |         | 41,82%  |
| SBS40: Origen desconocido (firma plana)   |         |         |         |         |         |         |
| SBS44: Sistema de reparación MMR defectuoso   |         |         |         |         |         |         |
| SBS84: Actividad desaminasa de citidina inducida por activación (AID)                             |         |         |         |         |         |         |
| SBS85: Efectos indirectos de desaminasa de citidina inducida por actividad (AID)                  |         |         |         | 20,70%  |         |         |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |         |         |         |         | 16,18%  |         |
| SBS93: Origen desconocido   | 36,39%  | 46,98%  |         |         |         |         |
| SBS94: Origen desconocido   | 29,44%  |         |         |         |         |         |
| Posible artefacto de secuenciación  |         |         |         |         |         |         |
| Tratamiento quimioterápico  |         |         | 12,41%  |         |         | 23,64%  |

Continuación tabla suplementaria 85.

|   | READ-25 | READ-26 | READ-27 | READ-28 | READ-29 | COAD-30 |
|---|---------|---------|---------|---------|---------|---------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   |         | 27,72%  | 14,19%  | 26,71%  | 42,03%  | 26,39%  |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |         |         |         |         |         |         |
| SBS5: Origen desconocido (firma plana)  | 7,59%   |         | 69,03%  | 44,52%  | 38,41%  | 62,50%  |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          | 8,86%   |         | 5,81%   |         |         |         |
| SBS10d: Defectos en la lectura de prueba de POLD1   |         |         |         |         |         |         |
| SBS15: Sistema de reparación MMR defectuoso   |         |         | 10,97%  |         |         |         |
| SBS16: Origen desconocido   |         |         |         | 14,38%  |         |         |
| SBS18: Daño por especies reactivas del oxígeno  |         |         |         |         |         |         |
| SBS29: Tabaco de mascar   |         |         |         |         |         | 11,11%  |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |         |         |         |         |         |         |
| SBS36: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>MUTYH</i> |         |         |         |         |         |         |
| SBS39: Origen desconocido   |         |         |         |         |         |         |
| SBS40: Origen desconocido (firma plana)   |         | 62,50%  |         |         |         |         |
| SBS44: Sistema de reparación MMR defectuoso   |         |         |         |         |         |         |
| SBS84: Actividad desaminasa de citidina inducida por activación (AID)                             |         |         |         |         |         |         |
| SBS85: Efectos indirectos de desaminasa de citidina inducida por actividad (AID)                  |         |         |         |         |         |         |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |         |         |         |         |         |         |
| SBS93: Origen desconocido   |         |         |         |         |         |         |
| SBS94: Origen desconocido   | 83,54%  |         |         |         |         |         |
| Posible artefacto de secuenciación  |         |         |         | 14,38%  |         |         |
| Tratamiento quimioterápico  |         | 9,78%   |         |         | 19,57%  |         |

Continuación tabla suplementaria 85.

|   | COAD-31 |
|---|---------|
| SBS1: Desaminación espontánea de 5-metilcitosinas (firma plana)                                   | 19,69%  |
| SBS2: Actividad de familia APOBEC de citidina desaminasas   |         |
| SBS5: Origen desconocido (firma plana)  | 80,31%  |
| SBS10b: Mutaciones en el dominio exonucleasa de la polimerasa $\epsilon$                          |         |
| SBS10d: Defectos en la lectura de prueba de POLD1   |         |
| SBS15: Sistema de reparación MMR defectuoso   |         |
| SBS16: Origen desconocido   |         |
| SBS18: Daño por especies reactivas del oxígeno  |         |
| SBS29: Tabaco de mascar   |         |
| SBS30: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>NTHL1</i> |         |
| SBS36: Sistema de reparación por escisión de bases defectuoso debido a mutaciones en <i>MUTYH</i> |         |
| SBS39: Origen desconocido   |         |
| SBS40: Origen desconocido (firma plana)   |         |
| SBS44: Sistema de reparación MMR defectuoso   |         |
| SBS84: Actividad desaminasa de citidina inducida por activación (AID)                             |         |
| SBS85: Efectos indirectos de desaminasa de citidina inducida por actividad (AID)                  |         |
| SBS88: Exposición a colibactina (bacterias <i>E. coli</i> con islas de patogenicidad pks)         |         |
| SBS93: Origen desconocido   |         |
| SBS94: Origen desconocido   |         |
| Posible artefacto de secuenciación  |         |
| Tratamiento quimioterápico  |         |

# ANEXOS



# 11 ANEXOS

## Anexo I. Acuerdo de licencia de uso de BMJ Publishing Group Limited para la modificación de la figura 1 de Genetic architecture of colorectal cancer, Peters U, Bien S, Zubair N, 64(10), 1623-36. Gut. 2015<sup>37</sup>.

### License Details

This Agreement between Anael Lopez-Novo ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

[Print](#) [Copy](#)

|                                 |   |
|---------------------------------|---|
| License Number                  | 5576080852807   |
| License date                    | Jun 25, 2023  |
| Licensed Content Publisher      | BMJ Publishing Group Ltd.   |
| Licensed Content Publication    | Gut   |
| Licensed Content Title          | Genetic architecture of colorectal cancer   |
| Licensed Content Author         | Ulrike Peters,Stephanie Bien,Niha Zubair  |
| Licensed Content Date           | Oct 1, 2015   |
| Licensed Content Volume         | 64  |
| Licensed Content Issue          | 10  |
| Type of Use                     | Dissertation/Thesis   |
| Requestor Type                  | Individual Account  |
| Format                          | Print and electronic  |
| Portion Used                    | Figure/table/extract  |
| Number of translations          | 1   |
| Number of figure/table/extracts | 1   |
| Will you be translating?        | Yes   |
| Title                           | Mrs   |
| Institution name                | Universidade de Santiago de Compostela  |
| Expected presentation date      | Sep 2023  |
| Portions                        | Figure 1  |
| Specific Languages              | Spanish   |
| Circulation                     | 50  |
| Requestor Location              | Anael Lopez-Novo<br>Calle del Horreo, 80. Piso 3<br><br>Santiago de Compostela, 15701<br>Spain<br>Attn: Anael Lopez-Novo<br>GB674738491 |
| Publisher Tax ID                |   |
| Total                           | <b>0.00 EUR</b>   |

[BACK](#)

Copyright © 2023 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Data Security and Privacy](#). For California Residents. [Terms and Conditions](#).  
Comments? We would like to hear from you. E-mail us at [customer-care@copyright.com](mailto:customer-care@copyright.com)

BMJ

**Genetic architecture of colorectal cancer**

Author: Ulrike Peters,Stephanie Bien,Niha Zubair  
 Publication: Gut  
 Publisher: BMJ Publishing Group Ltd.  
 Date: Oct 1, 2015  
 Copyright © 2015, BMJ Publishing Group Ltd. & British Society of Gastroenterology

**Order Completed**

Thank you for your order.

This Agreement between Anael Lopez-Novó ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

|                       |               |                                   |
|-----------------------|---------------|-----------------------------------|
| <b>License Number</b> | 5576080852807 | <a href="#">Printable Details</a> |
| <b>License date</b>   | Jun 25, 2023  |                                   |

**Licensed Content**

|                                     |   |
|-------------------------------------|---|
| <b>Licensed Content Publisher</b>   | BMJ Publishing Group Ltd.                 |
| <b>Licensed Content Publication</b> | Gut                                       |
| <b>Licensed Content Title</b>       | Genetic architecture of colorectal cancer |
| <b>Licensed Content Author</b>      | Ulrike Peters,Stephanie Bien,Niha Zubair  |
| <b>Licensed Content Date</b>        | Oct 1, 2015                               |
| <b>Licensed Content Volume</b>      | 64  |
| <b>Licensed Content issue</b>       | 10  |

**About Your Work**

|                                   |  |
|-----------------------------------|--|
| <b>Title</b>                      | Mrs                                    |
| <b>Institution name</b>           | Universidade de Santiago de Compostela |
| <b>Expected presentation date</b> | Sep 2023                               |

**Requester Location**

|                           |  |
|---------------------------|--|
| <b>Requester Location</b> | Anael Lopez-Novó<br>Calle del Herreo, 80. Piso 3<br>Santiago de Compostela, 15701<br>Spain<br>Attn: Anael Lopez-Novó |
|---------------------------|--|

**Price**

|              |          |
|--------------|----------|
| <b>Total</b> | 0.00 EUR |
|--------------|----------|

**Order Details**

|  |                      |
|--|----------------------|
| <b>Type of Use</b>                     | Dissertation/Thesis  |
| <b>Requestor Type</b>                  | Individual           |
| <b>Format</b>                          | Print and electronic |
| <b>Portion Used</b>                    | Figure/table/extract |
| <b>Number of languages</b>             | 1                    |
| <b>Number of figure/table/extracts</b> | 1                    |
| <b>Will you be translating?</b>        | Yes                  |

**Additional Data**

|                           |          |
|---------------------------|----------|
| <b>Portions</b>           | Figure 1 |
| <b>Specific Languages</b> | Spanish  |
| <b>Circulation</b>        | 50       |

**Tax Details**

|                         |             |
|-------------------------|-------------|
| <b>Publisher Tax ID</b> | GB674738491 |
|-------------------------|-------------|

CLOSE WINDOW

ORDER MORE

**Total: 0.00 EUR**

BMJ PUBLISHING GROUP LTD. LICENSE  
TERMS AND CONDITIONS

Jun 25, 2023

---

---

This Agreement between Anael Lopez-Novo ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

|                              |   |
|------------------------------|---|
| License Number               | 5576080852807                             |
| License date                 | Jun 25, 2023                              |
| Licensed Content Publisher   | BMJ Publishing Group Ltd.                 |
| Licensed Content Publication | Gut                                       |
| Licensed Content Title       | Genetic architecture of colorectal cancer |
| Licensed Content Author      | Ulrike Peters,Stephanie Bien,Niha Zubair  |
| Licensed Content Date        | Oct 1, 2015                               |
| Licensed Content Volume      | 64  |
| Licensed Content Issue       | 10  |
| Type of Use                  | Dissertation/Thesis                       |
| Requestor type               | Individual                                |
| Format                       | Print and electronic                      |
| Portion                      | Figure/table/extract                      |



# ANAEL LÓPEZ NOVO

|                                      |  |
|--------------------------------------|--|
| Number of figure/table/extracts      | 1  |
| Description of figure/table/extracts | Figure 1   |
| Will you be translating?             | Yes  |
| Number of languages                  | 1  |
| Languages                            | Spanish  |
| Circulation/distribution             | 50   |
| Title                                | Mrs  |
| Institution name                     | Universidade de Santiago de Compostela   |
| Expected presentation date           | Sep 2023   |
| Portions                             | Figure 1   |
| Specific Languages                   | Spanish  |
| Requestor Location                   | Anael Lopez-Novos<br>Calle del Horreo, 80. Piso 3<br><br>Santiago de Compostela, 15701<br>Spain<br>Attn: Anael Lopez-Novos |
| Publisher Tax ID                     | GB674738491  |
| Total                                | 0.00 EUR   |
| Terms and Conditions                 |  |

### BMJ Terms and Conditions for Permissions

When you submit your order you are subject to the terms and conditions set out below. You will also have agreed to the Copyright Clearance Center's ("CCC") terms and conditions regarding billing and payment <https://s100.copyright.com/App/PaymentTermsAndConditions.jsp>. CCC are acting as BMJ Publishing Group Limited's ("BMJs") agent.

Subject to the terms set out herein, BMJ hereby grants to you (the Licensee) a non-exclusive, non-transferable licence to re-use material as detailed in your request for this/those purpose(s) only and in accordance with the following conditions:

- 1) **Scope of Licence:** Use of the Licensed Material(s) is restricted to the ways specified by you during the order process and any additional use(s) outside of those specified in that request, require a further grant of permission.
- 2) **Acknowledgement:** In all cases, due acknowledgement to the original publication with permission from BMJ should be stated adjacent to the reproduced Licensed Material. The format of such acknowledgement should read as follows:  
"Reproduced from [publication title, author(s), volume number, page numbers, copyright notice year] with permission from BMJ Publishing Group Ltd."
- 3) **Third Party Material:** BMJ acknowledges to the best of its knowledge, it has the rights to licence your reuse of the Licensed Material, subject always to the caveat that images/diagrams, tables and other illustrative material included within, which have a separate copyright notice, are presumed as excluded from the licence. Therefore, you should ensure that the Licensed Material you are requesting is original to BMJ and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested in any way indicates that it was reprinted or adapted by BMJ with permission from another source, then you should seek permission from that source directly to re-use the Licensed Material, as this is outside of the licence granted herein.
- 4) **Altering/Modifying Material:** The text of any material for which a licence is granted may not be altered in any way without the prior express permission of BMJ. If adaptation of the material has been approved via [bmj.permissions@bmj.com](mailto:bmj.permissions@bmj.com) you must include the disclaimer: "Adapted by permission from BMJ Publishing Group Limited. [publication title, author, volume number, page numbers, copyright notice year]"
- 5) **Reservation of Rights:** BMJ reserves all rights not specifically granted in the combination of (i) the licence details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment Terms and Conditions.
- 6) **Timing of Use:** First use of the Licensed Material must take place within 12 months of the grant of permission.
- 7) **Creation of Contract and Termination:** Once you have submitted an order via RightsLink and this is received by CCC, and subject to you completing accurate details of your proposed use, this is when a binding contract is in effect and our acceptance occurs. As you are ordering rights from a periodical, to the fullest extent permitted by law, you will have no right to cancel the contract from this point other than for BMJ's material breach or fraudulent misrepresentation or as otherwise permitted under a statutory right. Payment must be made in accordance with CCC's Billing and Payment Terms and conditions. In the event that you breach any material condition of these terms and condition or any of CCC's Billing and Payment Terms and Conditions, the license is automatically terminated upon written notice from BMJ or CCC or as otherwise provided for in CCC's Billing and Payment Terms and Conditions, where these apply. Continued use of materials where a licence has been terminated, as well as any use of the Licensed Materials beyond the scope of an unrevoked

licence, may constitute intellectual property rights infringement and BMJ reserves the right to take any and all action to protect its intellectual property rights in the Licensed Materials.

8) **Warranties:** BMJ makes no express or implied representations or warranties with respect to the Licensed Material and to the fullest extent permitted by law this is provided on an "as is" basis. For the avoidance of doubt BMJ does not warrant that the Licensed Material is accurate or fit for any particular purpose.

9) **Limitation of Liability:** To the fullest extent permitted by law, BMJ disclaims all liability for any indirect, consequential or incidental damages (including without limitation, damages for loss of profits, information or interruption) arising out of the use or inability to use the Licensed Material or the inability to obtain additional rights to use the Licensed Material. To the fullest extent permitted by law, the maximum aggregate liability of BMJ for any claims, costs, proceedings and demands for direct losses caused by BMJ's breaches of its obligations herein shall be limited to twice the amount paid by you to CCC for the licence granted herein.

10) **Indemnity:** You hereby indemnify and hold harmless BMJ and their respective officers, directors, employees and agents, from and against any and all claims, costs, proceeding or demands arising out of your unauthorised use of the Licensed Material.

11) **No Transfer of License:** This licence is personal to you, and may not be assigned or transferred by you without prior written consent from BMJ or its authorised agent(s). BMJ may assign or transfer any of its rights and obligations under this Agreement, upon written notice to you.

12) **No Amendment Except in Writing:** This licence may not be amended except in a writing signed by both parties (or, in the case of BMJ, by CCC on BMJ's behalf).

13) **Objection to Contrary terms:** BMJ hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment Terms and Conditions. These terms and conditions, together with CCC's Billing and Payment Terms and Conditions (which to the extent they are consistent are incorporated herein), comprise the entire agreement between you and BMJ (and CCC) and the Licensee concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment Terms and Conditions, these terms and conditions shall control.

14) **Revocation:** BMJ or CCC may, within 30 days of issuance of this licence, deny the permissions described in this licence at their sole discretion, for any reason or no reason, with a full refund payable to you should you have not been able to exercise your rights in full. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice from BMJ or CCC will not, to the fullest extent permitted by law alter or invalidate the denial. For the fullest extent permitted by law in no event will BMJ or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to BMJ and/or CCC for denied permissions.

15) **Restrictions to the license:**

15.1) **Promotion:** BMJ will not give permission to reproduce in full or in part any Licensed Material for use in the promotion of the following:

- a) non-medical products that are harmful or potentially harmful to health
- b) medical products that do not have a product license granted by the Medicines and Healthcare products Regulatory Agency (MHRA) or its international equivalents. Marketing of the product may start only after data sheets have been released to members of the medical profession and must conform to the marketing authorization contained in the product license.

16) **Translation:** This permission is granted for non-exclusive world English language rights only unless explicitly stated in your licence. If translation rights are granted, a professional translator should be employed and it must be a true reproduction, accurately conveying the original meaning and of the same quality.

17) **STM Permissions Guidelines:** For content reuse in journals that qualify for permission under the STM Permissions Guidelines (which may be updated from time to time) the terms and conditions of the Guidelines supersede those in this licence. <https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/>

18) **General:** Neither party shall be liable for failure, default or delay in performing its obligations under this Licence, caused by a Force Majeure event which shall include any act of God, war, or threatened war, act or threatened act of terrorism, riot, strike, lockout, individual action, fire, flood, drought, tempest or other event beyond the reasonable control of either party.

18.1) In the event that any provision of this Agreement is held to be invalid, the remainder of the provisions shall continue in full force and effect.

18.2) There shall be no right whatsoever for any third party to enforce the terms and conditions of this Agreement. The Parties hereby expressly wish to exclude the operation of the Contracts (Rights of Third Parties) Act 1999 and any other legislation which has this effect and is binding on this agreement.

18.3) To the fullest extent permitted by law, this Licence will be governed by the laws of England and shall be governed and construed in accordance with the laws of England. Any action arising out of or relating to this agreement shall be brought in courts situated in England save where it is necessary for BMJ for enforcement to bring proceedings to bring an action in an alternative jurisdiction.

V1.1

## Anexo II. Documento de información al participante en investigación biomédica facilitado a los pacientes participantes en el proyecto TOGETHER.



### DOCUMENTO DE INFORMACIÓN PARTICIPANTE EN INVESTIGACIÓN BIOMÉDICA

**TÍTULO LÍNEA DE INVESTIGACIÓN:** GENÉTICA DEL CÁNCER. PROYECTO TOGETHER  
**IDENTIDAD RESPONSABLE:** FUNDACIÓN PÚBLICA GALEGA DE MEDICINA XENÓMICA  
**GRUPO DE INVESTIGACIÓN:** GRUPO DE MEDICINA XENÓMICA

#### FINALIDAD DE LA PARTICIPACIÓN.-

Toda investigación biomédica requiere recoger datos y muestras biológicas de personas afectadas por la patología a estudiar y de personas no afectadas por dicha enfermedad, para analizarlas y obtener conclusiones para conocer mejor y avanzar en el diagnóstico y/o tratamiento de las enfermedades que se van a investigar. De hecho, muchos de los avances científicos obtenidos en los últimos años en medicina son fruto de este tipo de estudios, en los que se emplean muestras biológicas humanas que o bien son remanentes de muestras que han sido obtenidas para el diagnóstico/control de las enfermedades, o bien se solicitan *ex profeso* para una investigación concreta. Por este motivo, **le invitamos a colaborar en la Línea de investigación en GENÉTICA DEL CÁNCER.**

El objetivo principal de esta **Línea de investigación** es conocer las bases genéticas de la predisposición al cáncer; sus bases moleculares; las Interacciones Gen X Ambiente; y los aspectos relacionados con la Medicina Personalizada y la estratificación de pacientes y enfermedad, con el fin de establecer las estrategias para la prevención, diagnóstico y seguimiento clínico del cáncer. La finalidad principal de esta investigación es progresar en el conocimiento de la patología a estudiar y en su prevención, diagnóstico, pronóstico y tratamiento.

**Debe saber que la realización de los proyectos de investigación en los que se utilicen muestras humanas, será previamente autorizada por el Comité de Ética competente.**

Toda la información que le facilitamos en este documento y la actividad del Grupo de investigación y el Biobanco están reguladas por la Ley Orgánica 15/1999, de 13 de Diciembre, de Protección de Datos de Carácter Personal (LOPD), la Ley 14/2007, de 3 de Julio de Investigación Biomédica (LIBM), el Real Decreto 1716/2011, de 11 de noviembre.

Su **participación es totalmente voluntaria**. Si firma el consentimiento informado, confirmará que desea participar. Puede negarse a participar o retirar su consentimiento en cualquier momento posterior a la firma sin tener que explicar los motivos. Su **no-participación o retirada posterior del consentimiento no afectará en modo alguno a su atención sanitaria presente o futura**. Antes de firmar el consentimiento puede preguntarnos cualquier duda que le surja o consultar con terceras personas.

#### MUESTRAS BIOLÓGICAS E INFORMACIÓN ASOCIADA. RIESGOS.-

MUESTRAS BIOLÓGICAS. Sangre, otros fluidos, células o tejido obtenidas durante su atención sanitaria y que ya no es necesario utilizarlas para la finalidad por la que se obtuvieron y que le son solicitadas para ser incluidas en esta línea de investigación.

El responsable guardará y dispondrá de estas muestras para realizar los estudios de investigación biomédica. Las muestras y la información asociada a las mismas se almacenarán en las áreas establecidas para tal fin, dentro de las instalaciones de la Fundación Pública Galega de Medicina Xenómica, bajo la responsabilidad de Ángel Carracedo Álvarez.

La donación de estas muestras no impedirá que usted o su familia puedan usarlas, cuando sea necesario por motivos de salud, siempre que estén disponibles y no hayan sido anonimadas.

INFORMACIÓN CLÍNICA que, junto con los resultados de los estudios realizados en las muestras, nos permita extraer conclusiones útiles para el manejo de las enfermedades. Para el buen desarrollo de los estudios de investigación, es necesario obtener datos relativos al donante de las muestras, por lo que necesitaremos acceder a su historia clínica y/o datos genómicos (si los hubiera) para recabar la información que será conservada junto a la muestra. (*Más información en el apartado de Confidencialidad.*)

En caso de ser necesaria alguna información o muestra adicional, y siempre que usted nos lo autorice en la hoja de consentimiento, la institución sanitaria se podría poner en contacto con usted para solicitarle nuevamente su colaboración.



### **CONDICIONES DE LA DONACIÓN.-**

La donación y utilización de muestras biológicas humanas será gratuita, por lo que usted no obtendrá ni ahora ni en el futuro ningún beneficio económico por la misma. Usted renuncia a cualquier derecho de naturaleza económica, patrimonial o potestativa sobre los resultados o potenciales beneficios económicos que puedan derivarse de las investigaciones que se realicen con la muestra que cede.

Tampoco obtendrá ningún otro beneficio directo para su salud como resultado de su donación. Sin embargo, los conocimientos obtenidos gracias a los estudios llevados a cabo a partir de su muestra y de muchas otras pueden ayudar al avance médico y, por ello, a otras personas.

La titularidad de los resultados de la investigación corresponderá al investigador y a la Institución donde se realice la investigación.

### **CONFIDENCIALIDAD.-**

Toda la información referida a usted será considerada confidencial y tratada en base a las indicaciones establecidas en la normativa indicada al inicio de este documento. Tal información sólo estará disponible para el personal autorizado, el cual tiene el **deber legal de guardar secreto**. Sus muestras y sus datos clínicos asociados a las mismas, pasarán a formar parte del fichero de datos "Investigación biomédica e biobancos", responsabilidad de la **Consellería de Sanidade** y del **Servizo Galego de Saúde**.

**Para garantizar la confidencialidad de su identidad (asegurar que la información de su muestra no se relaciona con su identidad), su muestra sólo irá identificada, desde su entrada en la investigación, con un código.** Sólo este código, y nunca su identidad aparecerá en el material con el que se trabaje. La relación entre su código y su identidad, quedará custodiada por el personal autorizado, en una base de datos que cumple todos los requisitos legales exigidos. De esta manera podemos asegurar que cualquier información que se obtenga a partir de sus muestras, permanezca **confidencial** pero pueda ser asociada a sus datos, por si fuera información de interés clínico.

La cesión de las muestras y los datos asociados a las mismas, a los investigadores colaboradores con esta línea de investigación, se realizará de manera codificada (se mantendrá por el investigador responsable en este centro el vínculo que relaciona su identidad con la muestra cedida). A todos ellos se les exigirá que nos garanticen que trabajarán con el mismo nivel de protección de datos exigido por la normativa española.

Por otro lado, es posible que los resultados de las investigaciones sean publicados en la literatura científica, pero entendiendo estos resultados como los obtenidos de la totalidad de las muestras, no los resultados individuales. Si este fuera el caso, su identidad permanecerá completamente confidencial y nunca formará parte de ninguna publicación.

Con su aceptación a participar con esta línea de investigación en genética del cáncer, usted accede a que esta información pueda ser transferida en las mencionadas condiciones. Usted puede no autorizarnos a realizar las cesiones aquí indicadas, indicando su decisión en el apartado correspondiente de la hoja del consentimiento informado.

Puede ejercitar los **DERECHOS ACCESO, RECTIFICACIÓN, CANCELACIÓN Y/O OPOSICIÓN (DERECHOS ARCO)** que le reconoce la normativa española, dirigiendo su solicitud por escrito junto a una copia de su DNI (con el fin de garantizar que el ejercicio de estos derechos es realizado por la persona autorizada para ello) a:

- **ENTREGA EN PERSONA**, en la Fundación Pública Galega de Medicina Genómica a la atención de Ángel Carracedo Álvarez
- **CORREO POSTAL**: Fundación Pública Galega de Medicina Genómica, CHUS, Edif Consultas, planta -2 Trav Choupana s/n, 15706 Santiago de Compostela
- **CORREO ELECTRÓNICO**: angel.carracedo@usc.es

Si le surge cualquier duda sobre su participación en esta investigación, usted puede contactar con los investigadores de la Línea de investigación en genética del cáncer, en el **Teléfono: 981951491**. En todo momento se pondrán los medios necesarios para facilitarle la información más adecuada.

Si usted decide **REVOCAR SU CONSENTIMIENTO**, deberá hacernos llegar, a la atención del **Director Ejecutivo** de la Fundación Pública Galega de Medicina Xenómica, D. Ángel Carracedo Álvarez, la hoja de revocación totalmente cubierta y firmada que aparece al final de este documento.





Usted debe saber que la revocación tendrá efectos a partir del momento en que se formalice y no afectará a los datos resultantes de las investigaciones que ya se hayan llevado a cabo hasta ese momento.

## **RESTRICCIONES DE USO DE LA MUESTRA.-**

Usted puede indicarnos si quiere establecer algún tipo de restricción sobre sus muestras y datos, en relación a su posible uso en determinados proyectos de investigación o en cuanto a determinadas cesiones. Para ello dispone de un apartado específico en la hoja de firma del consentimiento informado.

## **INFORMACIÓN SOBRE RESULTADOS DE LA INVESTIGACIÓN.-**

La evaluación de los resultados se hará sólo por grupos (por ejemplo, hombres / mujeres, grupos de edad, diagnósticos, etc.) y no de forma individual. Debe comprender que los resultados de valor que se pudieran obtener provenirían del estudio de múltiples muestras, y en ningún caso de la suya exclusivamente.

Las implicaciones médicas de los resultados de las distintas pruebas, si es que los hay, sólo serán conocidas cuando se haya completado la investigación.

El responsable de la investigación tendrá a disposición de los participantes la información sobre los proyectos de investigación en los que se utilicen las muestras y datos.

En determinadas circunstancias el Comité de Ética competente podrá decidir si es necesario contactar con el participante para facilitarle información de manera individualizada.

Las diferentes investigaciones en las que se utilicen sus muestras y datos pueden requerir la realización de estudios de biología celular y genéticos, y a partir de ellos se puede obtener información que puede ser relevante para su salud o la de su familia. En los estudios genéticos se puede descubrir información no buscada, lo que se denominan hallazgos inesperados. Cuando esto se produzca, los resultados obtenidos serán validados y analizados por profesionales para determinar si según el conocimiento médico actual tienen una aplicabilidad clínica o tratamiento que aconseje su comunicación a las personas afectadas.

**Usted debe saber que tiene derecho a conocer, o no, la información obtenida con el análisis de sus muestras.**

**En el caso de que usted decida no ser informado, la ley establece que cuando la información obtenida sea necesaria para evitar un grave perjuicio para la salud de sus familiares, un Comité de expertos estudiará el caso y deberá decidir entre la conveniencia o no de informar a los afectados o a sus representantes legales.**

## **DESTINO FINAL DE LA MUESTRA**

El remanente de las muestras será almacenado en la Colección de muestras, creada en la Fundación Pública Gallega de Medicina Genómica por el Dr. Ángel Carracedo e inscrita con el código C.0002864 en el registro nacional do ISCIII, para la Línea de investigación en genética del cáncer con el fin de poder utilizar estas muestras y sus datos asociados en los distintos proyectos de investigación en los que participe el responsable de la misma, D. Ángel Carracedo y cualquier otro personal adscrito o colaborador de los Grupos de Investigación participantes; siempre y cuando, dichos estudios se enmarquen en la finalidad de esta línea de investigación. Deben saber que a realización de todos estos proyectos de investigación en los que se utilicen estas muestras, será previamente autorizada por el Comité de Ética competente.

En una segunda fase, cuando así lo determine la persona responsable de la investigación con la que Ustedes accedieron a colaborar con las muestras y datos, y siempre que así lo autoricen en la hoja de Firma del consentimiento informado, las muestras sobrantes junto a los datos asociados serán depositadas en el Biobanco do Complexo Hospitalario Universitario de Santiago.

Los Biobancos son bancos de almacenamiento de muestras de origen humano para su utilización en investigaciones nacionales o internacionales dentro del campo de la biomedicina. Su funcionamiento se centra en gestionar, bajo criterios de seguridad, calidad y eficiencia; la recepción, procesamiento, almacenamiento y posterior cesión de muestras a los investigadores solicitantes, para que utilicen las mismas en sus proyectos de investigación; siempre y cuando, éstos cumplan todos los requisitos éticos y legales exigibles para este tipo de prácticas.



El Biobanco guardará y dispondrá de estas muestras sobrantes para realizar los estudios de investigación biomédica. Las muestras y la información asociada a las mismas se almacenarán en las áreas establecidas para tal fin, dentro de las instalaciones del Complejo Hospitalario Universitario de Santiago, bajo la responsabilidad de la Dirección del Biobanco do Complejo Hospitalario Universitario de Santiago.

La cesión de las muestras y los datos asociados a las mismas, a los investigadores solicitantes de las muestras, comunitarios (países miembros de la Unión Europea) o extracomunitarios (no miembros), se efectuará con carácter general de manera anónima o disociada, es decir, sólo se cederá la muestra asociada a datos genéricos, sin que se pueda identificar por medios razonables su identidad. En las investigaciones en las que se considere necesario para el desarrollo de las mismas, a propuesta del investigador responsable y previa autorización del comité de ética competente, podrán enviarse sus muestras codificadas (se mantendrá por el biobanco el vínculo que relaciona su identidad con la muestra cedida). A todos ellos se les exigirá que nos garanticen que trabajarán con el mismo nivel de protección de datos exigido por la normativa española.

El Biobanco do Complejo Hospitalario Universitario de Santiago asume como propio todo lo descrito en los apartados previos de este documento en relación a la confidencialidad, los posibles hallazgos, etc.

El biobanco podrá repercutir únicamente los costes de obtención, procesado y envío de las muestras a los investigadores/instituciones que las solicitaron.

#### **BIOBANCO DO COMPLEXO HOSPITALARIO UNIVERSITARIO DE SANTIAGO.**

**TITULAR DIRECCIÓN CIENTÍFICA:** Máximo Fraga.

**TELÉFONO DE CONTACTO:** 981 955 148 **CORREO ELECTRÓNICO:** Maximo.Francisco.Fraga.Rodriguez@sergas.es

#### **CIERRE DEL BIOBANCO.**

De producirse un eventual cierre del biobanco o revocación de la autorización para su constitución y funcionamiento, la información sobre el destino de las muestras estará a su disposición en la página web del Biobanco y/o Centro Hospitalario [www.idisantiago.es/biobanco](http://www.idisantiago.es/biobanco) y en el Registro Nacional de Biobancos para Investigación Biomédica del Instituto de Salud Carlos III (ISC III), con página web [www.isciii.es](http://www.isciii.es), con el fin de que pueda manifestar su conformidad o disconformidad con el destino previsto para las muestras.

***Muchas gracias por su colaboración.***





### CONSENTIMIENTO INFORMADO DEL PARTICIPANTE MAYOR DE EDAD

(Espacio para colocar la etiqueta con los datos del donante muestra)

Yo..... (Nombre y apellidos del participante manuscritos) he leído la hoja de información y pude hacer todas las preguntas que consideré necesarias, y acepto participar en la Línea de investigación en genética del cáncer con mis muestras y datos para que se utilicen en las condiciones que me han informado.

Marcar en caso de que se trate de una muestra CONTROL

**RESTRICCIONES DE USO (usted podrá indicarnos cruzando la correspondiente casilla las restricciones de uso que considere. En caso de no marcar ninguna casilla entenderemos que NO manifiesta ninguna condición de uso).**

- NO autorizo a utilizar mis muestras y datos en investigaciones que .....
- NO autorizo a que consulten y obtengan información de mi historia clínica.
- NO acepto que contacten conmigo cuando sea necesario.
- NO quiero ser informado de los resultados de las investigaciones que sean de interés para mi salud.
- NO autorizo la cesión de mis muestras y datos de manera codificada (disociada).
- NO autorizo la cesión de mis muestras y datos fuera de la Unión Europea.
- NO autorizo el depósito en el **Biobanco do Complexo Hospitalario Universitario de Santiago** de mis muestras y datos en las condiciones indicadas en la hoja de información

Si acepta que se contacte con usted cuando sea necesario, por favor indique su Teléfono / E-mail de contacto:

.....

Firma del participante  
Identidad del participante: .....

Firma de la persona que informa  
Identidad de la persona que informa: .....

Firma del testigo  
Confirmo el consentimiento verbal del participante, el cual lo autoriza a firmar en su nombre  
Identidad del testigo: ..... DNI: .....  
Solo se firmará en aquellos casos en los que el participante no pueda leer y/o escribir y delegue la firma en el testigo.

Firma del representante legal  
Identidad representante legal: ..... DNI: .....  
Solo se firmará en aquellos casos en los que el participante esté incapacitado legalmente.

En ....., a ..... de ..... de .....

*Le agradecemos su desinteresada colaboración con el avance de la ciencia y la medicina.*



### REVOCACIÓN DEL CONSENTIMIENTO

Yo, D./Dña: ..... con DNI .....

Revoco / anulo el consentimiento prestado en fecha: .....  
 Y no deseo proseguir la donación voluntaria realizada al ..... (nombre del Centro Sanitario), que doy con esta fecha por finalizada.

- SOLICITO ELIMINACIÓN SOLO DE LAS MUESTRAS.
- SOLICITO ELIMINACIÓN SOLO DE MIS DATOS PERSONALES.

- La muestra quedará anonimizada irreversiblemente y podrá ser utilizada en proyectos de investigación.
- SOLICITO ELIMINACIÓN TOTAL DE MIS DATOS Y MUESTRAS.

Fdo.:

En ..... a ..... de ..... de 20.....

Yo, D./Dña: ..... con DNI .....

En calidad de: (señale la opción aplicable a su caso)

- Testigo que firma en lugar
- Representante legal

de D./Dña: ..... con DNI .....

Revoco / anulo el consentimiento prestado en fecha: .....  
 Y no deseo proseguir la donación voluntaria realizada al ..... (nombre del Centro Sanitario), que doy con esta fecha por finalizada.

- SOLICITO ELIMINACIÓN SOLO DE LAS MUESTRAS.
- SOLICITO ELIMINACIÓN SOLO DE MIS DATOS PERSONALES.

- La muestra quedará anonimizada irreversiblemente y podrá ser utilizada en proyectos de investigación.
- SOLICITO ELIMINACIÓN TOTAL DE MIS DATOS Y MUESTRAS.

Fdo.:

En ..... a ..... de ..... de 20.....

## Anexo III. Aprobación del proyecto “Integración de datos ómicos germinales y tumorales para la identificación de nuevos genes de susceptibilidad hereditaria a cáncer colorrectal (TOGETHER) (PI17/00509)” por el Comité Ético de Investigación de Galicia.



Secretaría Técnica  
Comité Autonómico de Ética da Investigación de Galicia  
Secretaría Xeral, Consellería de Sanidade  
Edificio Administrativo San Lázaro  
15703 SANTIAGO DE COMPOSTELA  
Tel: 881548425, Correo-e: ceic@sergas.es



### DICTAMEN DEL COMITÉ DE ÉTICA DE LA INVESTIGACIÓN DE SANTIAGO-LUGO

Guillermo José Prada Ramallal, Secretario del Comité de Ética de la Investigación de Santiago-Lugo,

#### CERTIFICA:

Que este Comité evaluó en su reunión del día 20 de julio de 2017 el estudio:

**Título:** Integración de datos ómicos germinales y tumorales para la identificación de nuevos genes de susceptibilidad hereditaria a cáncer colorrectal (TOGETHER)  
**Promotor:** Clara Ruiz Ponte  
**Tipo de estudio:** Outros  
**Versión:**  
**Código del Promotor:**  
**Código de Registro:** 2017/355

Y, tomando en consideración las siguientes cuestiones:

- La pertinencia del estudio, teniendo en cuenta el conocimiento disponible, así como los requisitos legales aplicables, y en particular la Ley 14/2007, de investigación biomédica, el Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humana, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica, la ORDEN SAS/3470/2009, de 16 de diciembre, por la que se publican las Directrices sobre estudios Postautorización de Tipo Observacional para medicamentos de uso humano, y la Circular nº 07/2004, de investigaciones clínicas con productos sanitarios.
- La idoneidad del protocolo en relación con los objetivos del estudio, justificación de los riesgos y molestias previsibles para el sujeto, así como los beneficios esperados.
- Los principios éticos de Declaración de Helsinki vigente.
- Los Procedimientos Normalizados de Trabajo del Comité.

Emite un dictamen **FAVORABLE** para la realización del estudio por el/la investigador/a del centro:

| Centros                                       | Investigadores Principales |
|---|----------------------------|
| Fundación Pública Galega de Medicina Xenómica | Clara Ruiz Ponte           |
|   |                            |

En Santiago de Compostela, a 26 de julio 2017.

El Secretario del Comité Territorial de Ética de la Investigación de Santiago Lugo,



Emetido digitalmente por: PRADA RAMALLAL  
WILHELMO JOSSE - 44841744R  
Fecha: 2017.07.30 23:29:56 +0200

Guillermo José Prada Ramallal



XUNTA DE GALICIA  
CONSELLERÍA DE SANIDADE  
Secretaría Xeral Técnica

Secretaría Técnica  
Comité Autonómico de Ética da Investigación de Galicia  
Secretaría Xeral, Consellería de Sanidade  
Edificio Administrativo San Lázaro  
15703 SANTIAGO DE COMPOSTELA  
Tel: 881540425. Correo-e: [ceic@sergas.es](mailto:ceic@sergas.es)



Guillermo José Prada Ramallal, Secretario del Comité de Ética de la Investigación de Santiago-Lugo,

**HACE CONSTAR QUE:**

1.- El Comité Territorial de Ética de la Investigación de Santiago-Lugo cumple tanto en su composición como en sus PNTs los requisitos legales vigentes (RD 1090/2015 de ensayos clínicos, y la Ley 14/2007 de Investigación Biomédica).

2.- La composición actual del Comité Territorial de Ética de la Investigación de Santiago-Lugo es:

- **Juan Manuel Vázquez Lago (Presidente)**. Médico especialista en Medicina Preventiva y Salud Pública. Área de Gestión Integrada de Santiago.
- **Pilar Rodríguez Ledo (Vicepresidenta)**. Médico especialista en Medicina Familiar y Comunitaria. Área de Gestión Integrada de Lugo.
- **Guillermo José Prada Ramallal (Secretario)**. Médico especialista en Farmacología Clínica. Área de Gestión Integrada de Santiago. Fundación Ramón Domínguez.
- **Lorenzo Armenteros del Olmo (Vicesecretario)**. Médico especialista en Medicina Familiar y Comunitaria. Área de Gestión Integrada de Lugo.
- **Francisco Campos Pérez**. Biólogo. Instituto de Investigación Sanitaria de Santiago de Compostela.
- **Rosana Castelo Domínguez**. Farmacéutica de Atención Primaria. Área de Gestión Integrada de Santiago.
- **Ricardo García Martínez**. Licenciado en Derecho. Área de Gestión Integrada de Lugo.
- **Jaime Gulín Dávila**. Farmacéutico especialista en Farmacia Hospitalaria. Área de Gestión Integrada de Lugo.
- **Victor Herrán Carreira**. Paciente. ADIL-Asociación de Diabéticos Lucense.
- **María Jesús Lamas Díaz**. Farmacéutica especialista en Farmacia Hospitalaria. Área de Gestión Integrada de Santiago.
- **Carlos Rodríguez Moreno**. Médico especialista en Farmacología Clínica. Área de Gestión Integrada de Santiago.
- **Rafael Carlos Vidal Pérez**. Médico especialista en Cardiología. Área de Gestión Integrada de Lugo.
- **María Jesús Wandosell Picatoste**. Enfermera. Área de Gestión Integrada de Santiago.

Para que conste donde proceda, y a petición del promotor/investigador, en Santiago de Compostela, a 26 de julio de 2017.

El Secretario del Comité Territorial de Ética de la Investigación de Santiago Lugo,



Emisado digitalmente por PRADA RAMALLAL  
GUILLERMO JOSE - 44841744R  
Fecha: 2017.07.30 23:29:59 +0200'

Guillermo José Prada Ramallal



XUNTA DE GALICIA  
CONSELLERÍA DE SANIDADE  
Secretaría Xeral Técnica

Secretaría Técnica  
Comité Autonómico de Ética da Investigación de Galicia  
Secretaría Xeral, Consellería de Sanidade  
Edificio Administrativo San Lázaro  
15703 SANTIAGO DE COMPOSTELA  
Tel: 881549425. Correo-e: [oeic@sergas.es](mailto:oeic@sergas.es)



Se emite un informe FAVORABLE CONDICIONADO a que se tengan en cuenta las siguientes cuestiones:

**1. Aspectos metodológicos:**

Deben aportarse los compromisos de los investigadores que van a hacer el reclutamiento y recogida de muestras (endoscopias y cirugías) en los distintos centros participantes.

**2. Modelos de documentos:**

Para sucesivas solicitudes de evaluación, se sugiere utilizar los siguientes modelos recomendados por el Comité:

1.- Protocolo de estudio: Puede encontrar y descargar el formulario para presentación de protocolos de investigación elaborado y aprobado por el CEI-SL en nuestra página web: [http://acis.sergas.es/DXerais/118/20170426\\_Formulario%20Modelo%20Protocolo-CEIC.pdf](http://acis.sergas.es/DXerais/118/20170426_Formulario%20Modelo%20Protocolo-CEIC.pdf)

2.- Hoja de Información al Participante y Consentimiento informado: Puede encontrar este documento en nuestra página web: <http://acis.sergas.es/Paxinas/web.aspx?tipo=paxtab&idLista=3&idContido=83&migtab=83&idTax=15534&idioma=es>

Firmado digitalmente por: PRADA RAMALLAL  
GUILLERMO JOSE - 44641744R  
Fecha: 2017.07.30 23:30:02 +0200

El cáncer colorrectal es una de las neoplasias con mayor incidencia y mortalidad a nivel mundial. La heredabilidad de esta enfermedad se estima en hasta un 35%; sin embargo, las estrategias empleadas hasta el momento solo han conseguido explicar un pequeño porcentaje de esta. En este trabajo de Tesis Doctoral se integran datos de ADN, ARN y epigenoma, a nivel germinal y tumoral, de pacientes enriquecidos genéticamente por un diagnóstico temprano de cáncer colorrectal. La identificación de nuevos genes candidatos de susceptibilidad a esta neoplasia contribuirá a explicar parte del riesgo heredado a desarrollar la enfermedad, aumentando el conocimiento sobre las bases hereditarias de cáncer colorrectal.