

PROYECTO FINAL DE INGENIERÍA

BÚSQUEDA DE MARCADORES GLICOINMUNOLÓGICOS PARA PROGNOSIS Y TRATAMIENTO DE CÁNCER COLORRECTAL

Mazzeo, Christian Alejandro - LU 1049260

Licenciatura en Bioinformática

Tutores:

Dra. Mariño, Karina Valeria
Universidad Argentina de la Empresa
Laboratorio de Glicómica Funcional y Molecular
IBYME-CONICET

Dr. Cagnoni, Alejandro Javier
Laboratorio de Glicómica Funcional y Molecular
IBYME-CONICET

Co-Tutor:

Dra. Girotti, María Romina
Universidad Argentina de la Empresa
IBYME-CONICET

8 de noviembre de 2021



UNIVERSIDAD ARGENTINA DE LA EMPRESA
FACULTAD DE INGENIERÍA Y CIENCIAS EXACTAS

Agradecimientos

A la Dra. Mariño y al Dr. Cagnoni, por darme esta inmensa oportunidad, por la paciencia que me han tenido y sobre todas las cosas, por el excelente trato.

Al Dr. Prada por haberme alentado a postularme para esta tesis y además, en conjunto con el Dr. Martínez Tosar, por transmitirme toda su pasión por la biología y biotecnología.

A Naty, por acompañarme y alentarme incondicionalmente en cada momento de mi vida.

A Ada, mi hija perruna, por estar siempre conmigo y esperarme cada día después de clases.

A mis viejos.

Resumen

El cáncer colorrectal (CCR) es una enfermedad de alta incidencia y letalidad, siendo la segunda causa de muerte por enfermedades neoplásicas en Argentina (2018). Su desarrollo y progresión se correlaciona con factores genéticos, ambientales, inmunológicos, dietarios, la microbiota intestinal, y con la preexistencia de enfermedades inflamatorias intestinales (EII). El desarrollo del CCR asociado a colitis (CCRAC) en pacientes con EII presenta similitudes y diferencias respecto del CCR esporádico (CCRE), pero el tratamiento de estos pacientes es complejo, debido a que las terapias para EII se basan en el uso de agentes inmunosupresores durante períodos prolongados. En ese sentido, en este proyecto nos propusimos evaluar la glicobiología del proceso neoplásico en el intestino, comparando estudios de bases de datos de transcriptómica en modelos experimentales de CCRE y CCRAC, y a su vez, si la desregulación de glicogenes en modelos de EII podría ser un factor clave en el desarrollo y la progresión tumoral. Finalmente, abordamos el análisis de bases de datos de pacientes para validar estos resultados.

En este trabajo, hemos establecido en modelos experimentales una firma de 15 glicogenes desregulados, tanto en CCRAC como en CCRE, que permiten postular mecanismos comunes a ambas patologías, mientras que el gen *Plaur*, que codifica al receptor de uroquinasa, aparece desregulado específicamente en el modelo de CCRAC. A su vez, los estudios comparativos entre los modelos experimentales de EII y CCRAC, demostraron que siete de los glicogenes que definen la firma en CCRAC ya se encuentran desregulados en inflamación intestinal, incluyendo *Plaur*, marcando potenciales mecanismos glicobiológicos que conectan la inflamación con el desarrollo neoplásico. Al realizar el mismo análisis en pacientes con los datos de transcriptómica obtenidos de la base de datos de TCGA (COAD), pudimos establecer una firma de 45 glicogenes desregulados para CCR, de los cuales tres (*NOTUM*, *PTN*, *KLB*) establecen una firma en común entre los modelos experimentales y la patología humana.

Dada la relevancia del infiltrado inmune en neoplasias, analizamos el infiltrado inmune de los modelos experimentales de CCRAC, CCRE y EII con ImmuCC, y el de pacientes con cuatro algoritmos de deconvolución distintos: xCell, Cibersort, quanTIsec y EPIC. Para las poblaciones inmunes detectadas por los cuatro algoritmos, los resultados en pacientes fueron similares. Por último, todos los pipelines y el código escrito en R para la obtención de los resultados de esta tesis se encuentran disponibles en la plataforma [GitHub](#).

Abstract

Colorectal cancer (CRC) is a highly incidental and fatal disease, being the second cause of death from neoplastic diseases in Argentina (2018). Its development and progression are correlated with genetic, environmental, immunological, dietary factors and alterations in the intestinal microbiota. Likewise, inflammatory bowel diseases (IBD), have been described as a risk factor for the development of colorectal cancer. The development of colitis-associated CRC (CACRC) in patients with IBD has similarities and differences with respect to sporadic CRC (SCRC), but the treatment of CACRC patients is complex, since IBD therapies are based on the use of immunosuppressive agents. Therefore, in this project we set out to evaluate the glycobiology of the neoplastic process in the intestine, comparing studies of transcriptomic databases in experimental models of SCRC and CACRC, and in turn, if the dysregulation of glycogenes in an experimental model of IBD could influence neoplastic development. Lastly, we analyzed publicly available patient databases in order to validate those results obtained in experimental models.

In this work, we have established a signature of 15 deregulated glycogenes, both in CACRC and SCRC, which allow us to postulate mechanisms common to both pathologies, while the *Plaur* gene, which encodes urokinase receptor, appears specifically deregulated in CACRC. In turn, comparative studies between experimental models of IBD and CACRC, showed that seven of the glycogenes that define the signature in CACRC are already deregulated in intestinal inflammation, including *Plaur*, marking potential glycobiological mechanisms that connect inflammation with neoplastic development. By performing the same analysis in patients with transcriptomic data obtained from the TCGA database (COAD), we were able to establish a signature of 45 deregulated glycogenes for CRC, of which three establish a common signature between experimental models and human pathology, namely NOTUM, PTN, KLB.

Given the relevance of the immune infiltrate in neoplasms, we analyzed the immunophenotype of the experimental models of CACRC, SCRC and IBD with ImmuCC, and in patients with four different deconvolution algorithms: xCell, Cibersort, quanTIseq and EPIC. The immune populations detected by the four algorithms showed similar results with patients' samples. Finally, all the pipelines and the code written in R to obtain the results of this thesis are available on the [GitHub](#) platform.

Índice

Agradecimientos	2
Resumen	3
Abstract	4
1. Introducción	8
1.1 Cáncer colorrectal: patofisiología, factores de riesgo y desafíos actuales	8
1.2 Glicobiología tumoral	12
2. Objetivos	18
2.1 Objetivo general	18
2.2 Objetivos específicos	18
3. Materiales y métodos	19
3.1 Bases de datos para modelos murinos	19
3.2 Bases de datos para muestras humanas	20
3.3 Listado de genes involucrados en glicosilación o reconocimiento	20
3.4 Procesamiento de datos provenientes de microarrays	20
3.5 Procesamiento de datos provenientes de RNA-Seq	20
3.6 Análisis descriptivo	21
3.8 Caracterización del infiltrado inmune	21
3.9 Análisis de sobrevida	25
3.10 Software	26
3.11 Hardware	26
4. Resultados	27
4.1 Análisis de transcriptómica en modelos experimentales de cáncer colorrectal y enfermedades inflamatorias intestinales	27
4.1.1 Desregulaciones en la expresión de glicogenes en modelos experimentales de CCRE (Apc ^{min/+}) y Asociado a Colitis (AOM-DSS)	27

4.1.1.1	Alteraciones en el transcriptoma de colon en un modelo experimental de CCRAC (AOM-DSS)	29
4.1.1.2	Alteraciones en el transcriptoma de colon en un modelo experimental de CCRE (<i>APC^{min/+}</i>)	31
4.1.1.3	Caracterización del infiltrado inmune	34
4.1.2	Desregulaciones en la expresión de glicogenes en un modelo de CCRAC (AOM-DSS) analizado según localización anatómica y cronicidad de la inflamación intestinal	35
4.1.2.1	Alteraciones en el transcriptoma de colon en modelo experimental de CCRAC (AOM-DSS) Distal Temprano	37
4.1.2.2	Alteraciones en el transcriptoma de colon en modelo experimental de CCRAC (AOM-DSS) Distal Tardío	38
4.1.2.3	Caracterización del infiltrado inmune en modelo experimental de CCRAC (AOM-DSS) Distal	41
4.1.3	Comparación en la expresión de glicogenes en el modelo de inflamación intestinal por dextrano sulfato sódico (DSS) y en CCRAC (AOM-DSS)	44
4.1.3.1	Alteraciones en el transcriptoma de colon en modelo experimental de CCRAC (AOM-DSS)	45
4.1.3.2	Alteraciones en glicogenes en un modelo experimental de enfermedades inflamatorias intestinales por dextran sulfato sódico (DSS)	47
4.1.3.3	Caracterización del infiltrado inmune	50
4.1.4	Búsqueda de una firma robusta de glicogenes en modelos experimentales de CCRAC	50
4.2	Análisis de la expresión de glicogenes en biopsias de pacientes con CCR	54
4.2.1	Análisis descriptivo	55
4.2.2	Caracterización del infiltrado inmune	58
4.2.3	Comparación del infiltrado inmune según distintos algoritmos	65

4.2.4	Comparación del infiltrado inmune en un modelo experimental de CCRAC y pacientes de CCR	68
4.2.5	Alteraciones en el glicotranscriptoma de CCR en pacientes	69
4.3	Comparación entre los modelos experimentales y pacientes con CCR	71
4.3.1	Firma de glicogenes desregulados compartidas entre modelos experimentales y pacientes con CCR	71
4.3.2	Curva de supervivencia para la firma de glicogenes en pacientes con CCR	72
5.	Discusión	74
5.1	Análisis de los algoritmos para caracterización del infiltrado inmune	74
5.2	Glicofirmas en cáncer colorrectal	76
5.3	<i>Plaur</i> : ¿La conexión glicobiológica entre EII y CCRAC?	79
	Bibliografía	81
	Anexo A: Listado de pacientes con cáncer colorrectal de la base de datos TCGA (COAD)	96
	Anexo B: Listado de glicogenes en modelos experimentales murinos y pacientes de CCR.	106
	Anexo C: Resultados de la comparación por grupos del infiltrado inmune en modelos experimentales murinos y pacientes de CCR.	125
	Anexo D: Caracterización del infiltrado tumoral en pacientes con CCR utilizando xCell para los grupos: tumor vs. control, estadio vs. control, MSI vs. control y dMMR vs. control.	132
	Anexo E: Conversión de variable cuantitativa a categórica.	140
	Anexo F: Curva de supervivencia en pacientes con CCR filtrado por estadio I y II para los genes <i>NOTUM</i> , <i>PTN</i> y <i>KLB</i> .	142

1. Introducción

1.1 Cáncer colorrectal: patofisiología, factores de riesgo y desafíos actuales

El cáncer colorrectal (CCR) es el tercer tipo de cáncer más frecuente a nivel mundial y el segundo en términos de mortalidad (Bray *et al.*, 2018). En 2018, se diagnosticaron casi 1.850.000 nuevos casos de CCR y cerca de 900.000 personas fallecieron por esta enfermedad globalmente. En Argentina se informaron cerca de 16.000 casos y 8.700 muertes, siendo la segunda causa de muerte por cáncer, detrás del cáncer de pulmón (Bray *et al.*, 2018).

Esta patología generalmente comienza como un pólipo adenomatoso benigno (un crecimiento anormal de las células de la pared interna del intestino grueso), que luego puede progresar hacia un cáncer invasivo y finalmente metastásico. Los diferentes estadios reflejan la acumulación de cambios genéticos que ocurren durante la progresión tumoral, afectando la funcionalidad de proteínas de distintas vías de señalización que regulan la proliferación celular, supervivencia, resistencia, migración e invasión de células tumorales. Las mutaciones genéticas más frecuentemente encontradas en CCR son APC (del inglés, *adenomatous polyposis coli*, 85%), KRAS (del inglés *Kirsten rat sarcoma viral oncogene*, 35-45%), TGFβR (receptor de TGFβ, 30%) y el gen supresor tumoral P53 (35-55%) (Markowitz y Bertagnolli, 2009; De Freitas Junior y Morgado-Díaz, 2016).

Si bien no se conocen las causas determinantes del CCR, su desarrollo y progresión se correlacionan con factores genéticos, ambientales, inmunológicos, dietarios y relacionados con la microbiota intestinal (Markowitz y Bertagnolli, 2009). El CCR puede presentarse de tres maneras: esporádica, hereditaria o familiar.

- La enfermedad esporádica, en donde no hay antecedentes familiares, representa aproximadamente el 70% de los casos totales de CCR. Generalmente comienza a partir de los 50 años y se encuentra relacionado con factores ambientales y dietarios (Lin *et al.*, 2009).
- Los pacientes con cáncer de colon hereditario representan menos del 10% de los casos totales. Éstos se subdividen en función de la presencia masiva de pólipos en el colon, que incluyen la poliposis adenomatosa familiar, la poliposis asociada a *MUTYH* (gen homólogo a *mutY* de *E. coli*) y los síndromes de poliposis hamartomatosos, o bien la ausencia de poliposis, representada por el cáncer colorrectal hereditario no asociado a poliposis (en inglés, HNPCC) o Síndrome de Lynch.

- El tercer tipo de CCR se conoce como familiar, y puede representar hasta el 25% de los casos totales. Se presenta en pacientes con antecedente familiar de CCR pero que no es consistente con los síndromes hereditarios.

Asimismo, los pacientes con enfermedades inflamatorias intestinales (EII), como la enfermedad de Crohn y Colitis Ulcerosa, presentan un mayor riesgo de desarrollo de cáncer colorrectal (denominado cáncer colorrectal asociado a colitis, CCRAC), ya que las mismas conducen a anomalías de la estructura y función del tracto gastrointestinal y donde la inflamación crónica favorece el desarrollo neoplásico (Beaugerie e Itzkowitz, 2015). La terapia principal para el tratamiento de EII se basa en el uso de agentes inmunosupresores para disminuir la inflamación de la mucosa intestinal. Sin embargo, evidencias recientes sugieren que el uso de estas terapias por períodos prolongados podría promover la carcinogénesis (Axelrad *et al.*, 2016). El desarrollo de CCRAC en pacientes con EII comparte mecanismos moleculares y alteraciones genéticas con el CCR esporádico (CCRE), incluyendo mutaciones en APC, activación del oncogén KRAS, activación de COX-2 (ciclooxigenasa-2) y pérdida de funcionalidad de P53. Sin embargo, la patogénesis del CCRAC se diferencia del carcinoma colorrectal esporádico con respecto al tiempo y la secuencia de estas alteraciones, tal como se presenta en la Figura 1, dando lugar a un proceso de carcinogénesis más acelerado y multifocal (Axelrad *et al.*, 2016).

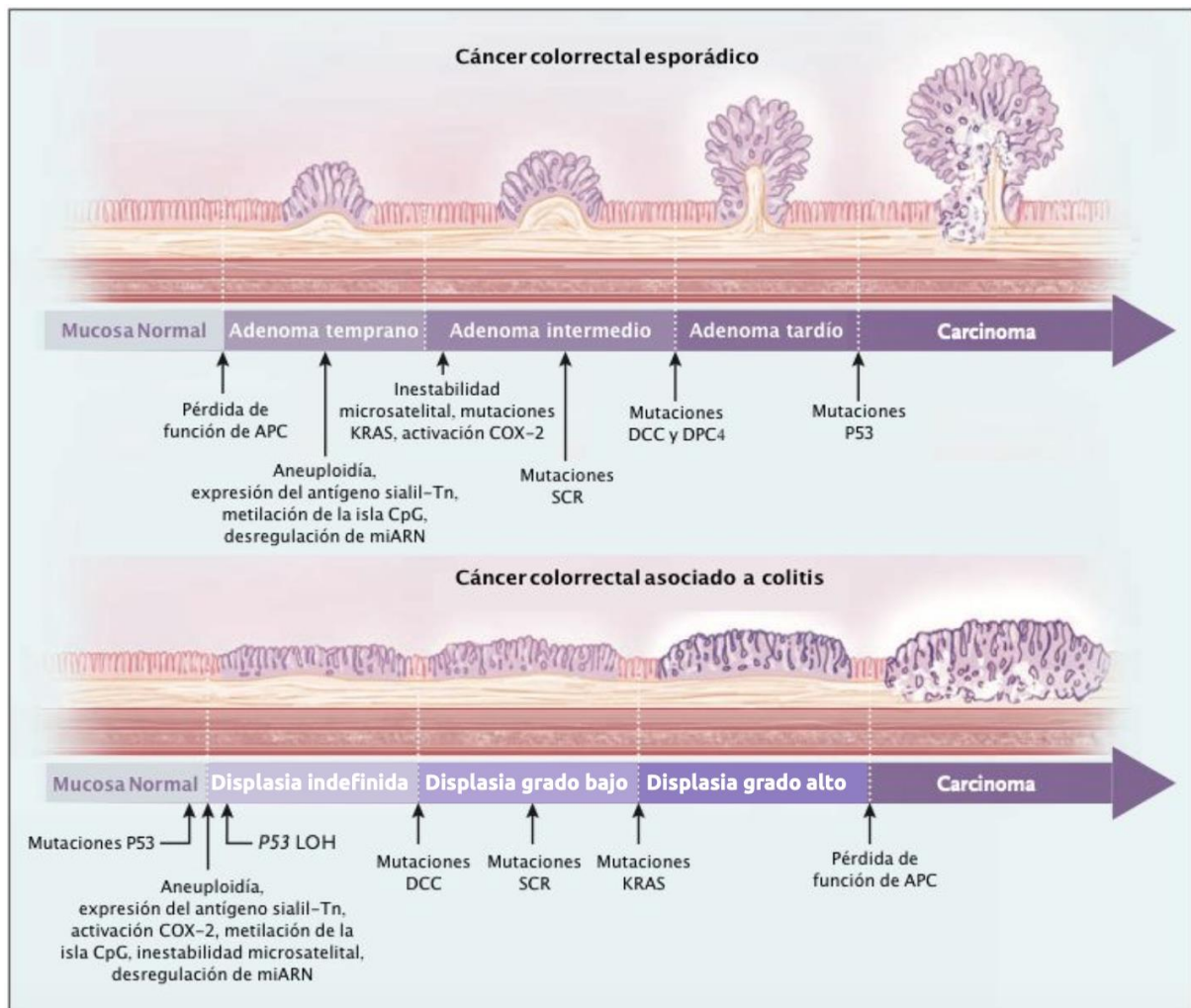


Figura 1: Comparación de mecanismos moleculares y cambios genéticos involucrados en el desarrollo de cáncer colorrectal esporádico y cáncer colorrectal asociado a colitis. Adaptado de Beaugerie e Itzkowitz (2015).

Una de las características distintivas del CCRAC (cuando se compara con el CCRE) es la mutación temprana de P53, presente en el 85% de los casos. Notablemente, P53 ya se encuentra alterada en mucosas inflamadas de pacientes con colitis ulcerosa que no habían desarrollado carcinomas, indicando el papel esencial de la inflamación. Asimismo, la pérdida de función de APC, un evento temprano en CCR esporádico, es mucho menos frecuente y suele ser tardío en CCRAC (Beaugerie e Itzkowitz, 2015). Actualmente se está dedicando mucho esfuerzo a la exploración del papel de la inflamación crónica en el desarrollo del CCR y el rol que ejercen las terapias inmunosupresoras utilizadas para el tratamiento de EII en la promoción del CCR. A pesar de ello, los conocimientos actuales sobre la interacción entre EII y CCR

siguen siendo escasos y los mecanismos moleculares que determinan el desarrollo y evolución de este tipo de CCR aún no han sido completamente elucidados.

Cuando el diagnóstico se realiza de manera temprana, el cáncer colorrectal suele ser tratable e incluso curable con terapias estándar como la remoción quirúrgica, la quimioterapia o la radioterapia (Markowitz *et al.*, 2002). Si el tumor se encuentra confinado dentro de las paredes del colon (Grado I y II) es operable, pero si no es tratado a tiempo éste se expande a regiones de nodos linfáticos (Grado III) hasta que finalmente se produce metástasis (Grado IV) (Markowitz y Bertagnolli 2009). Si bien las estrategias terapéuticas planteadas pueden resultar eficientes en estadios tempranos, también muestran desventajas importantes. Los agentes quimioterapéuticos, como el oxaliplatino, han resultado eficientes para el tratamiento del CCR, pero aproximadamente el 50% de los pacientes no responden y desarrollan serios efectos secundarios en el sistema nervioso. La quimioradioterapia, utilizada para el tratamiento de cáncer rectal, no presenta beneficios para el 40% de los pacientes con estadios avanzados. A diferencia de los estadios tempranos, los tumores de Grado IV resultan generalmente incurables, aún a pesar de los más recientes avances que han mejorado considerablemente el grado de supervivencia de estos pacientes. Las mismas incluyen terapias dirigidas contra EGFR (cetuximab, panitumumab) y terapias antiangiogénicas como bevacizumab (anti-VEGF), ziv-aflibercept, y ramucirumab (anti-VEGF receptor 2) (Ohhara *et al.*, 2016).

En las últimas dos décadas, nuevas terapias basadas en la estimulación de la respuesta inmune antitumoral, conocidas como inmunoterapias, han ofrecido estrategias prometedoras alternativas a la quimioterapia o radioterapia tradicionales (Sharma y Allison, 2015; Fritz *et al.*, 2019). Estas terapias, tales como el uso de vacunas terapéuticas, la transferencia adoptiva de linfocitos T y la inhibición de puntos de chequeo, ya han demostrado resultados muy positivos y respuestas duraderas en neoplasias que presentan una alta carga de mutaciones, tales como melanoma, cáncer de pulmón, riñón, hígado, cabeza y cuello (Liu y Guo, 2018). Sin embargo, la inmunoterapia no ha mostrado hasta el momento respuestas prometedoras en cáncer colorrectal, excepto para aquellos pacientes con deficiencias en el sistema de reparación de errores de emparejamiento del ADN (dMMR) (Ganesh *et al.*, 2019; Lynch y Murphy, 2016; Kim *et al.*, 2019). En ese sentido, se ha descrito en individuos con CCR inserciones o pérdida de secuencias repetidas cortas en el ADN (microsatélites), fenómeno que se conoce como inestabilidad microsatelital (MSI). La MSI es responsable de la acumulación de mutaciones en diferentes genes asociados con cáncer y que son cruciales en la iniciación y progresión del

tumor. En consecuencia, la FDA aprobó en 2017 el uso de anticuerpos monoclonales anti-PD-1 (pembrolizumab y nivolumab) como terapia de segunda línea para pacientes con cáncer de colon metastásico y con inestabilidad microsatelital alta (MSI-H), lo que representa a menos del 5% del total de pacientes con CCR, y en 2020 como tratamiento de primera línea para los pacientes con CCR metastásico y MSI-H o dMMR. Recientemente, y al igual que para otras neoplasias, con el objetivo de lograr terapias más específicas se ha subclasificado molecularmente los tipos de tumor colorrectal, incluyendo: clasificación genómica (CMS 1-4) (Guinney *et al.*, 2015), inmune (C1-C6) (Thorsson *et al.*, 2018), por mutaciones (Braf, RAS), o clínicas (el sistema TNM, distal-proximal o izquierdo-derecho) (Dienstmann *et al.*, 2017). Actualmente, los ensayos clínicos en CCR están focalizados en encontrar nuevos subtipos de pacientes que puedan responder favorablemente al uso de inmunoterapias (Abdul-Latifa *et al.*, 2020).

La experiencia clínica demuestra que no hay una estrategia universal para el tratamiento de pacientes de CCR, dado que la enfermedad es multifactorial y que las características propias de cada paciente determinan el éxito de un determinado tratamiento. De esta manera, la medicina de precisión busca determinar cuál es la terapia más efectiva para cada caso individual y, asimismo, evitar que pacientes que no van a responder a una determinada terapia sufran innecesariamente los efectos tóxicos secundarios de la misma (Altunel *et al.*, 2020).

El desafío actual en investigación en CCR es lograr un sistema eficiente de estratificación de pacientes, que aporte no sólo al mejor diagnóstico y prevención, sino también al diseño, desarrollo y aplicación de nuevas terapias (Hull *et al.*, 2020). Caracterizar de forma detallada los cambios moleculares que ocurren en los distintos tipos de CCR puede ser clave en este proceso. En este contexto, la glicoinmunología tumoral es un campo de investigación emergente, focalizado en descifrar el papel de las interacciones entre carbohidratos y lectinas y su papel en la regulación de la respuesta del sistema inmune (Cagnoni *et al.*, 2016; Zhou *et al.*, 2018).

1.2 Glicobiología tumoral

La glicosilación es una de las modificaciones post-traduccionales más frecuentes, y se estima que entre el 0,5% y el 1% del genoma humano está dedicado a enzimas involucradas en la biosíntesis y degradación de glicoconjugados (Varki *et al.*, 2017).

El proceso de biosíntesis de los glicanos se diferencia del que genera los ácidos nucleicos y las proteínas debido a que los mismos se biosintetizan en una secuencia de reacciones enzimáticas que involucran glicosiltransferasas y glicosidasas (Figura 2). Estas reacciones son secuenciales y en ciertos casos competitivas, ya que carecen de un molde a seguir. Como resultado, los glicoconjugados obtenidos son naturalmente heterogéneos (generando “glicofomas” específicas) y proveen dinamismo en diversos procesos celulares de comunicación (Varki et al., 2017).

A medida que las tecnologías permitieron una mejor caracterización del glicoma celular (conjunto de estructuras glicosídicas presentes en una célula), se demostró que el mismo no es sólo variado, sino además, versátil (Varki *et al.*, 2017): las estructuras presentes en el glicoma responden a variables tan diversas como el estado nutricional de la célula, el estadio de desarrollo o en el caso de las células inmunes, su grado de activación y/o diferenciación (Rabinovich y Croci, 2012; Zhou *et al.*, 2018). La versatilidad y dinamismo del glicoma también se refleja en situaciones fisiopatológicas, donde las estructuras del perfil de glicanos, se ven alteradas por diversos procesos asociados a inflamación o malignización. Esta glicosilación aberrante también se ha descrito en glicoproteínas de suero (Mariño *et al.*, 2011).

El glicoma celular ejerce un papel central como mediador de diferentes procesos fisiopatológicos, incluyendo la modulación de la respuesta inmune. La interpretación de la información codificada por el glicoma se procesa biológicamente a partir de la interacción con proteínas con afinidad por carbohidratos, denominadas lectinas, que lo traducen en respuestas fisiológicas (Vasta *et al.*, 2012; Rabinovich y Croci, 2012). En particular, la alteración del glicoma en células tumorales y glicoproteínas circulantes en suero se ha descrito como una característica distintiva de las enfermedades neoplásicas (Mariño *et al.*, 2011; Venkitachalam *et al.*, 2016; Adamczyk *et al.*, 2012).

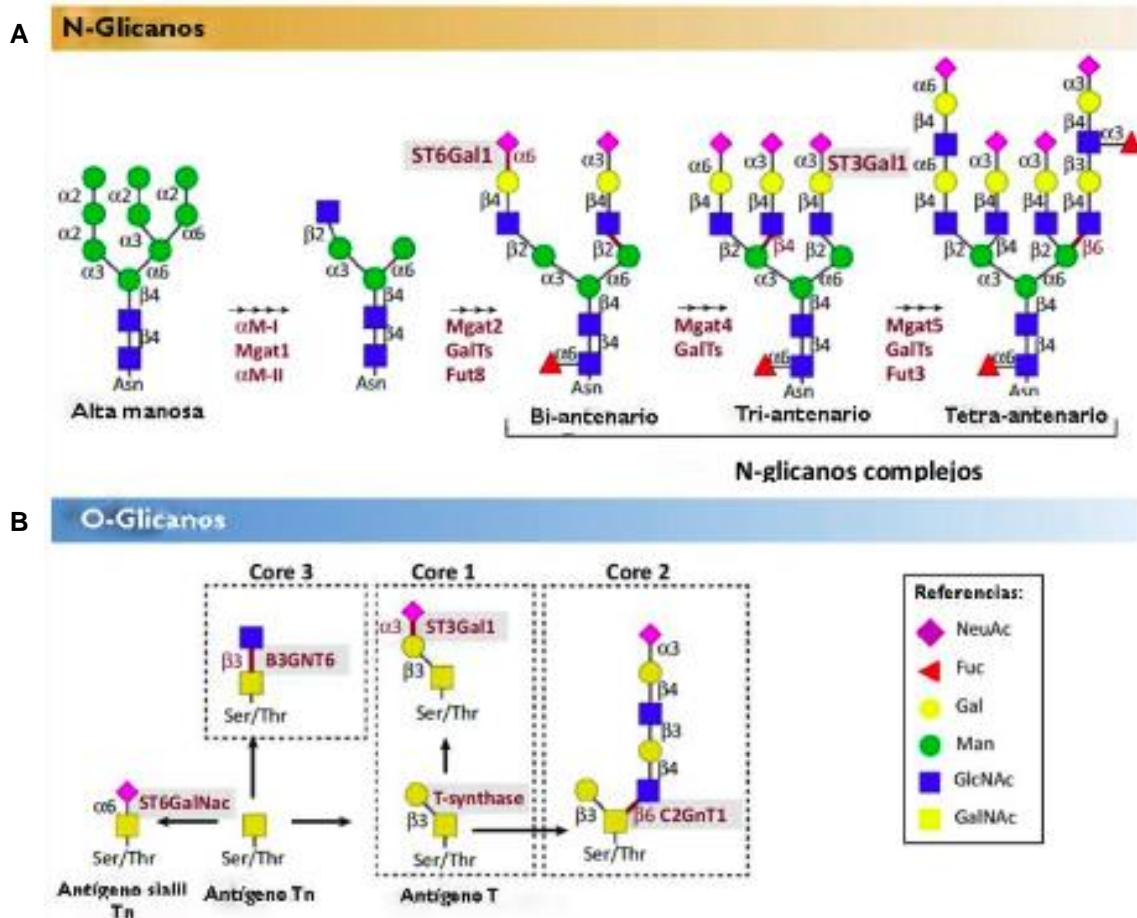


Figura 2: Esquema representativo de algunas vías de glicosilación involucradas en la síntesis de N-glicanos y O-glicanos en glicoproteínas. **A.** N-glicanos. Durante el proceso de síntesis de N-glicanos complejos, la actividad secuencial de las enzimas *N*-acetilglucosaminiltransferasas (MGAT) generan las antenas en N-glicanos complejos que al elongarse forman disacáridos LacNAc. Por otro lado, otra enzima relevante es la $\alpha(2-6)$ sialiltransferasa 1 (ST6GAL1), la cual incorpora ácido siálico en posición $\alpha(2-6)$. **B.** O-glicosilación. Posterior a la acción de las ppGalNAc transferasas, la C1GalT1 (T-sintasa o core-1 sintasa), elonga este residuo y es esencial para la biosíntesis de core-1 O-glicanos; la $\alpha(2-3)$ sialiltransferasa (ST3GAL1) es la potencial encargada de agregar ácido siálico en core-1 O-glicanos, y alternativamente, la $\beta(1-6)$ -*N*-acetilglucosaminiltransferasa 1 (GCNT1 ó C2GNT1) genera core-2 O-glicanos. Adaptado de Toscano *et al.*, 2018. Simbología de glicanos basada en Neelamegham *et al.*, 2019.

La glicosilación aberrante en células tumorales se ha asociado a la ocurrencia y progresión de tumores (Pinho y Reis, 2015), y se presenta como la alteración en la frecuencia de estructuras glicosídicas naturalmente presentes en la célula no malignizada o bien como la

aparición de nuevos glicanos. Estos cambios pueden encontrarse tanto en N- y O-glicanos en glicoproteínas, como en glicoesfingolípidos y glicosaminoglicanos, que presentan estructuras alteradas denominadas antígenos tumorales asociados a carbohidratos (TACA, del inglés: *Tumor-Associated Carbohydrate Antigens*) (Cagnoni *et al.*, 2016, Figura 3). Esta glicosilación aberrante asiste al desarrollo tumoral y es un factor clave que favorece la migración, invasión y metástasis por diversos mecanismos que incluyen la interacciones con lectinas endógenas como galectinas, siglecs y lectinas tipo C, las que pueden modular la respuesta inmune antitumoral y de ese modo, influir sobre el progreso de la neoplasia (Cagnoni *et al.*, 2016).

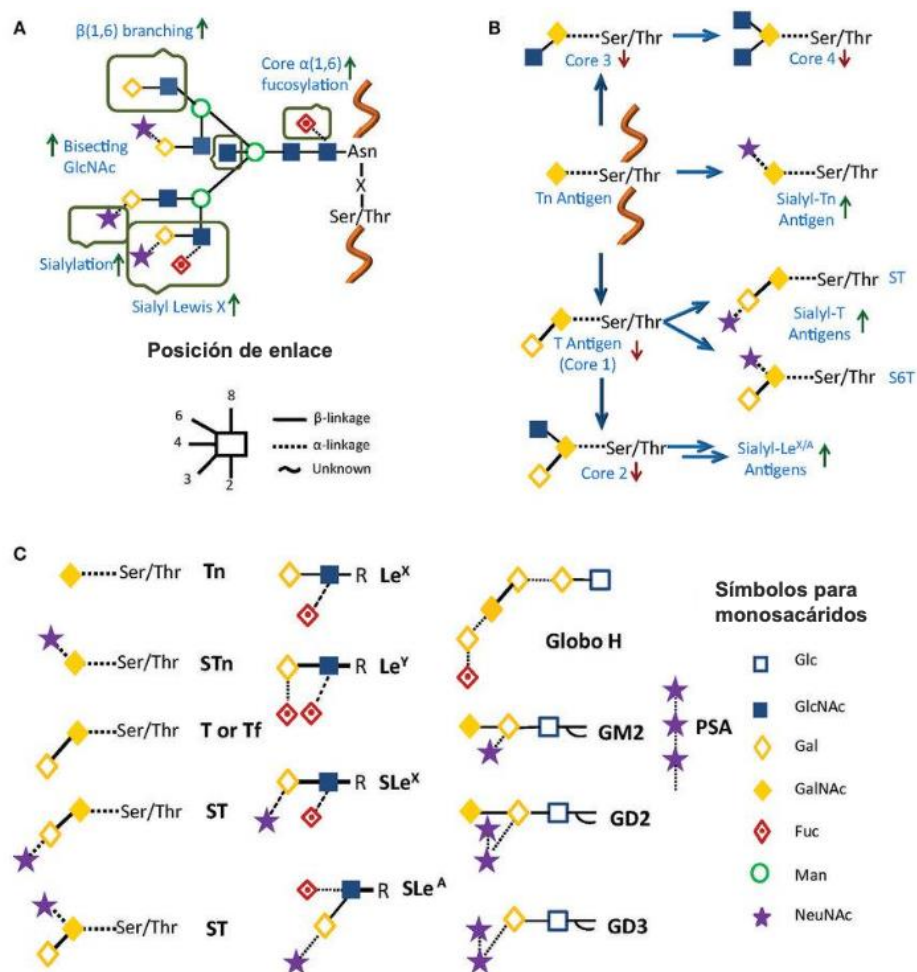


Figura 3: Alteraciones comunes observadas en cáncer en **A.** N-glicanos y **B.** O-glicanos, donde las flechas verdes y rojas representan expresión aumentada y disminuida de estructuras de glicanos, respectivamente. **C.** Principales antígenos de carbohidratos asociados a tumores (TACA). Adaptado de Cagnoni *et al.*, 2016. Los glicanos se representan por combinación de la simbología de Oxford (Harvey *et al.*, 2009) y la del Consortium for Functional Glycomics (Neelamegham *et al.*, 2019).

En particular, en cáncer colorrectal se ha demostrado que los cambios en la expresión de glicosiltransferasas relacionadas con la biosíntesis de N- y O-glicanos pueden modular la adhesión, señalización, pluripotencia, migración e invasividad de células tumorales (Venkitachalam *et al.*, 2017; De Freitas Junior y Morgado-Díaz, 2016). Zhang y colaboradores reportaron una relación entre la disminución en la expresión de la sialiltransferasa ST6GAL1 y un aumento en el poder metastásico de tumores colorrectales de grado avanzado. Esta disminución favorecería el aumento de estructuras $\alpha(2,3)$ -sialiladas, como la estructura antigénica sLeX, que promueve la adhesión de células tumorales al endotelio (Zhang *et al.*, 2017). Por otro lado, Lee y colaboradores describieron que ST6GAL1 induce adhesión y migración en una línea celular de colon malignizada (SW480), así como resistencia a la apoptosis promovida por agentes quimioterapéuticos (Park *et al.*, 2012). Recientemente, se describió el papel inmunomodulador de Galectina-1 (una lectina que reconoce específicamente glicoconjugados con unidades de N-acetilglucosamina) en cáncer colorrectal, a partir de la regulación de linfocitos T reguladores CD8⁺ (Cagnoni *et al.*, 2021).

Por otro lado, Chik y colaboradores (2014) demostraron que el glicoma de las líneas celulares de CCR presenta diferencias con el obtenido de biopsias de pacientes, y esto justifica la necesidad de mayor cantidad de estudios con muestras de pacientes. A modo de ejemplo, un trabajo reciente secuenció 430 glicogenes en líneas celulares de CCR derivadas de pacientes, así como en tejido tumoral, e identificó 12 de estos genes significativamente mutados en CCR, entre ellos, mutaciones en B3GNT2 (una GlcNAc-transferasa responsable de la biosíntesis de polilactosaminas), B4GALT2 (una galactosiltransferasa) y ST6GALNAC2, una $\alpha(2,6)$ -sialiltransferasa. Los autores caracterizaron bioquímicamente las glicosiltransferasas mutadas demostrando que la actividad modificada resultante alteraba el potencial migratorio de las células tumorales, y la expresión de ciertas glicoproteínas. Estos hallazgos sugieren que mutaciones en glicogenes pueden contribuir a la patogénesis de CCR (Venkitachalam *et al.*, 2017). Más allá de un glicoma celular aberrante, la alteración específica de la glicosilación en receptores también juega un papel clave; a modo de ejemplo, el perfil de N-glicosilación de E-cadherina (una de las proteínas clave en adhesión célula-célula epiteliales) se ve afectado durante el desarrollo neoplásico por: a) el aumento en la ramificación $\beta(1-6)$ de N-glicanos complejos por incremento de la expresión de MGAT5 (N-acetilglucosaminil-transferasa V); b) la disminución de GlcNAc bisectante catalizada por MGAT3 (N-acetilglucosaminil-transferasa III); c) el aumento de fucosilación de core mediada por la fucosiltransferasa FUT8 (Figura 3.A).

Estas alteraciones afectan la formación de uniones adherentes célula-célula estables, favoreciendo un fenotipo maligno de las células epiteliales y la migración de células tumorales (De Freitas Junior *et al.*, 2011). La disminución de expresión de MGAT3 y el aumento de expresión de MGAT5 también afectan la N-glicosilación de la integrina $\alpha 3\beta 1$, favoreciendo la migración celular (De Freitas Junior y Morgado-Díaz, 2016).

Estos antecedentes muestran la relevancia de la glicosilación aberrante como una característica distintiva del CCR. Los estudios iniciales de glicobiología en cáncer colorrectal se basaban en el estudio específico de glicosiltransferasas, o alternativamente de lectinas inmunomoduladoras mediante modelos experimentales en animales *knock out* en alguna de estas proteínas, limitando una visión global de la glicobiología del tumor y su microambiente. Con el advenimiento de la bioinformática y la disponibilidad de realizar análisis transcriptómicos, se facilitó considerablemente una visión más integrada y sistemática; sin embargo, los análisis bioinformáticos aplicados a la totalidad de los genes generalmente dejan de lado alteraciones en glicogenes debido a la baja expresión intrínseca de muchos de los genes involucrados en la biosíntesis y degradación de glicoconjugados. Además, los estudios de expresión no concuerdan con el glicoma aberrante detectado de manera estructural o fenotípica.

En este trabajo nos propusimos, a partir de una lista de 740 glicogenes seleccionados que incluyen no solo glicosiltransferasas, glicosidasas y lectinas, sino también glicoproteínas de relevancia biológica (Nairn *et al.*, 2008), estudiar específicamente la relevancia de la glicobiología en el desarrollo de cáncer colorrectal en modelos experimentales y en biopsias de pacientes. Con un abordaje integral, nos propusimos definir firmas de glicogenes que nos permitieran comprender las similitudes y diferencias de CCRAC y CCRE, y también, estudiar la posibilidad de una desregulación temprana del eje glicobiológico en EII que pueda influenciar el desarrollo de neoplasia.

2. Objetivos

2.1 Objetivo general

El presente proyecto de tesis se focalizó en el estudio de cambios en los niveles de expresión de genes específicos involucrados en la biosíntesis y degradación de glicoconjugados, así como también de las lectinas responsables de su reconocimiento y consecuente respuesta inmunológica durante la transición de tejido sano a neoplásico, tanto en modelos animales de cáncer colorrectal esporádico (CCRE) como de cáncer colorrectal asociado a colitis (CCRAC). Se realizó el mismo estudio con biopsias de pacientes de cáncer colorrectal. Asimismo, y dada la relevancia del eje lectinas-glicanos en la respuesta inmune, se caracterizó la composición celular del infiltrado tumoral. Estos estudios se orientan a la potencial identificación de biomarcadores que permitan establecer una firma genética, molecular e inmunológica asociada al eje lectina-glicanos en CCRE/CCRAC con el objetivo de favorecer la detección temprana, la estratificación de pacientes y/o el pronóstico de CCR.

2.2 Objetivos específicos

1. Estudiar el perfil transcriptómico de muestras tumorales en modelos experimentales de CCRAC (modelo de carcinogénesis por azoximetano y dextrano sulfato sódico, AOM-DSS), CCRE (modelo de carcinogénesis en animales *Apc^{min/+}*) y de inflamación intestinal (modelo de inducción por dextrano sulfato sódico, DSS). Mediante el uso de datos de transcriptómica de bases de datos públicas se analizará el nivel de expresión de glicogenes (glicoproteínas de interés, glicosiltransferasas, glicosidasas asociadas a la síntesis y/o hidrólisis de glicoconjugados y lectinas involucradas en su reconocimiento).
2. Comparar las firmas así obtenidas en muestras de pacientes con CCR, validando así la relevancia clínica de la desregulación de glicogenes en modelos experimentales.
3. Analizar el infiltrado inmune de tejidos tumorales en pacientes y modelos animales utilizando distintos métodos de deconvolución.

3. Materiales y métodos

3.1 Bases de datos para modelos murinos

Se utilizaron los datos de transcriptómica de modelos murinos de cáncer colorrectal esporádico (CCRE), cáncer colorrectal asociado a colitis (CCRAC) y enfermedades inflamatorias intestinales (EII) que se obtuvieron desde Gene Expression Omnibus (GEO). En total se descargaron 3 datasets, descritos en la Tabla I.

TABLA I: Descripción de los sets de datos de modelos murinos analizados.

GSE	43338	64658	57533
Cepa y sexo	C57BL/6J Machos	C57BL/6J Hembras	C57BL/6J Machos
Modelos experimentales inflamatorios y/o neoplásicos	<p>CCRAC AOM (10 mg/kg) + 3 ciclos de DSS (2,5% p/v, durante 7 días y 14 recuperación).</p> <p>CCRE <i>Apc^{min/+}</i></p>	<p>CCRAC Temprano AOM (10 mg/kg) + 2 ciclos de DSS (2,5% p/v, durante 7 días y 14 recuperación)</p> <p>CCRAC Tardío AOM (10 mg/kg) + 4 ciclos de DSS (2,5% p/v, durante 7 días y 14 recuperación)</p>	<p>EII 3 ciclos de DSS 2%, durante 7 días y 14 recuperación</p> <p>CCRAC AOM (12,5 mg/kg) + 3 ciclos de DSS 2%, durante 7 días y 14 recuperación).</p>
Toma de muestra	Distal	Proximal y Distal	Distal
Grupos	<ul style="list-style-type: none"> • CCRAC Control (n = 2, c/u pool de 5) • CCRAC (n = 4) • CCRE Control (n = 3, c/u pool de 4) CCRE (n = 5) 	<ul style="list-style-type: none"> • CCRAC Control Proximal (n = 6) • CCRAC Proximal Temprano (n = 3) • CCRAC Proximal Tardío (n = 4) • CCRAC Control Distal (n = 6) • CCRAC Distal Temprano (n = 3) CCRAC Distal Tardío (n = 6) 	<ul style="list-style-type: none"> • CCRAC Control (n = 3) • EII (n = 4) CCRAC (n = 3)
Técnica	Microarray Mouse Expression 430A Array de Affymetrix	Microarray Mouse Expression 430 2.0 Array de Affymetrix	RNA-seq Illumina HiSeq 2000 Pair end
Publicación de referencia	Neufert <i>et al.</i> , 2013.	Abu-Remaileh <i>et al.</i> , 2015.	Gkouskou <i>et al.</i> , 2016.

3.2 Bases de datos para muestras humanas

Los datos clínicos y de transcriptómica de cáncer colorrectal en pacientes se descargaron desde la base de datos de *The Cancer Genome Atlas* (TCGA, The Cancer Genome Atlas Research Network, Weinstein *et al.*, 2013), alojada en *Genomic Data Commons* (<https://gdc.cancer.gov/>, septiembre 2018), Proyecto TCGA-COAD, sitio primario: Colon.

Se descargaron 521 archivos htseq-counts y 521 archivos HTSeq-FPKM y sus respectivos casos clínicos, correspondiente a 461 pacientes. Se removieron muestras que no fueran de tumor primario o control y casos clínicos con mas de una muestra de tumor primario. Por último, se eliminaron los casos clínicos que no tuvieran la información sobre estadio tumoral, inestabilidad microsatelital o *Loss Expression Mismatch Repair Proteins*, obteniendo un total de 302 casos y 41 controles. El detalle de las muestras se encuentra en el Anexo A.

3.3 Listado de genes involucrados en glicosilación o reconocimiento

Se partió de una lista de 740 glicogenes murinos (Nairn *et al.*, 2008) y se buscaron los genes ortólogos en humanos utilizando la herramienta BioMart (Durinck *et al.*, 2005; Durinck *et al.*, 2009), totalizando una lista de 674 genes humanos, la cual se encuentra en el Anexo B.

3.4 Procesamiento de datos provenientes de microarrays

Para los microarrays, la descarga de archivos CEL se realizó con GEOquery (Davis y Meltzer, 2007) y se leyeron con AFFY (Gautier *et al.*, 2004). Para que las muestras sean comparables en cada proyecto se procedió a la normalización de la intensidad del microarray utilizando CGRMA (Wu *et al.*, 2020). Se procedió a filtrar las sondas con intensidad $\leq 2,5$. Para asegurar la calidad se graficó un boxplot de los valores de intensidad por cada muestra y un histograma de densidad vs. $\log(\text{intensidad})$ antes y después de la normalización. Además, se graficó la degradación de RNA. Para las anotaciones se utilizó la base de datos *mouse430a.db* (Carlson, 2016a) o *mouse4302.db* (Carlson, 2016b) según el tipo de microarray.

3.5 Procesamiento de datos provenientes de RNA-Seq

La descarga y descompresión de los archivos SRA se realizó con fast-dump de SRA Toolkit (Leinonen *et al.*, 2011) y el análisis se realizó con FastQC (Andrews, 2010). Se utilizó la anotación `Mus_musculus.GRCm38.94.gtf` y el genoma de referencia

Mus_musculus.GRCm38.dna.primary_assembly.fa que se descargaron del *Ensembl Genome Browser* -Ensembl- (Hubbard *et al.*, 2002) (www.ensembl.org). El mapeo de las lecturas al genoma se realizó con STAR (Dobin *et al.*, 2013) con los parámetros --outSAMtype BAM Unsorted. El ordenamiento de lecturas se realizó con Samtools (Li *et al.*, 2009). Para totalizar la expresión de cada gen se utilizó HTSeq (Anders *et al.*, 2015; Zanini *et al.*, 2021). Para el aseguramiento de la calidad se realizó un gráfico de Distancia de Cook y otro de la media los valores normalizados Mean of normalised counts. Los nombres de los genes se descargaron desde Ensembl utilizando BioMart (Durinck *et al.*, 2005; Durinck *et al.*, 2009).

3.6 Análisis descriptivo

Para ver cómo se agrupaban las muestras se utilizó K-mean con los valores iteration = 100, nstart = 20, y el optimal number igual a la cantidad de grupos predefinido, y también se calcularon la cantidad de grupos óptimos mediante el método de Gap Statistic, Elbow (Zhang *et al.*, 2020) y Sihouette. Para su visualización, se realizó un análisis de componentes principales (PCA) con los valores centrados y luego un T-distributed Stochastic Neighbor Embedding (tSNE) en 2 o 3 dimensiones con una perplexity de 30 y 500 iteraciones (van der Maaten y Hinton, 2008).

3.7 Expresión diferencial y downstream analysis

Para el análisis de expresión diferencial de genes entre los distintos grupos se utilizó Limma (Ritchie *et al.*, 2015) para microarrays y DESeq2 (Love *et al.*, 2014), tanto para RNAseq como para TCGA. Todos los Heatmaps se realizaron con pheatmap (Kolde, 2019) y los volcano plot con EnhancedVolcano (Blighe *et al.*, 2020).

Se tomó como valor significativo un p-valor ajustado $\leq 0,05$ para todas las pruebas realizadas. En la expresión diferencial se utilizó como línea de corte a los genes que además de tener un p-valor significativo tenían una tasa de cambio (*fold change*) ≥ 2 .

3.8 Caracterización del infiltrado inmune

Para caracterizar el infiltrado inmune en el microambiente tumoral se utilizaron diferentes paquetes de análisis disponibles al momento de este trabajo de Tesis. Se describen las poblaciones inmunes detectadas por cada método de deconvolución utilizado en las Tablas II y III. Se utilizó ImmuCC (Chen *et al.*, 2017; Chen *et al.*, 2018) en muestras murinas y

Cibersort (Newman *et al.*, 2015), xCell (Aran *et al.*, 2017), quanTIsec (Finotello *et al.*, 2019) y EPIC (Racle *et al.*, 2017) en muestras humanas. En ratones las muestras se volvieron a procesar siguiendo el flujo recomendado para cada plataforma y dado que ImmuCC no soporta el procesamiento de microarray Affymetrix 430A, no se pudo caracterizar el infiltrado inmune de GSE43338. Para los resultados obtenidos con Cibersort sólo las muestras con un p-valor significativo ($p < 0,05$) fueron consideradas para la caracterización del infiltrado inmune. La configuración utilizada para cada algoritmo se encuentra detallada en la Tabla IV.

Con los resultados del infiltrado tumoral se verificó el supuesto de normalidad de cada variable utilizando `mshapiro_test` de `rstatix` (Kassambara, 2020b), y en caso de no cumplirlo, se procedió con el análisis no paramétrico. Las comparaciones entre grupos se hicieron con Kruskal-Wallis (Kruskal y Wallis, 1952), mientras que las comparaciones de pares se realizaron con la prueba de Wilcoxon (Wilcoxon, 1945). Asimismo, se utilizó False Discovery Rate (Robertson *et al.*, 2019) como método de corrección por comparaciones múltiples. Los resultados de normalidad y de comparaciones entre grupos se encuentran en el Anexo C. Los histogramas y boxplots se realizaron con `ggpubr` (Kassambara, 2020).

TABLA II: Comparación de poblaciones inmunes detectadas por ImmuCC utilizado en modelos experimentales murinos de CCRAC.

Población		RNA-Seq	Microarray
Linfocitos	B	B	B de memoria
			B virgen
	Plasmática		Plasmática
	T CD4 ⁺	T CD4 ⁺	T CD4 ⁺ de memoria
			T CD4 ⁺ virgen
			T CD4 ⁺ folicular
	T CD8 ⁺	T CD8 ⁺	T CD8 ⁺ activado
			T CD8 ⁺ virgen
			T CD8 ⁺ de memoria
	T Reguladores		T Reguladores
	T Cooperadores		T cooperador 1
			T cooperador 2
		T cooperador 17	

Población		RNA-Seq	Microarray
Linfocitos	T Gamma Delta		T Gamma Delta
	NK (asesinos naturales)	NK naturales	NK en reposo
NK activadas			
Leucocitos	Eosinófilos	Eosinófilos	Eosinófilos
	Dendríticas	Dendríticas	Dendríticas activadas
			Dendríticas inmaduras
	Macrófagos	Macrófagos	Macrófagos M0
			Macrófagos M1
			Macrófagos M2
	Mastocitos	Mastocitos	Mastocitos
Monocitos	Monocitos	Monocitos	
Neutrófilos	Neutrófilos	Neutrófilos	

TABLA III: Comparación de poblaciones inmunes detectadas por cada método de deconvolución utilizado en pacientes con cáncer colorrectal.

Población		xCell	Cibersort	quanTIsec	EPIC
Linfocitos	B	B		B	B
		B ingenuo (<i>naïve</i>)	B ingenuo (<i>naïve</i>)		
		B de memoria	B de memoria		
		B de memoria conmutada de clase			
		B pro			
	Plasmática	Plasmática	Plasmática		
	T CD4 ⁺	T CD4 ⁺		T CD4 ⁺	T CD4 ⁺
		T CD4 ⁺ ingenuo	T CD4 ⁺ ingenuo		
		T CD4 ⁺ de memoria			
			T CD4 ⁺ de memoria descansando		
		T CD4 ⁺ de memoria activado			

Población		xCell	Cibersort	quanTIsec	EPIC	
Linfocitos	T CD4 ⁺	T CD4 ⁺ de memoria central		T CD4 ⁺	T CD4 ⁺	
		T CD4 ⁺ de memoria efectora				
	T CD8 ⁺	T CD8 ⁺	T CD8 ⁺	T CD8 ⁺	T CD8 ⁺	T CD8 ⁺
		T CD8 ⁺ virgen				
		T CD8 ⁺ de memoria central				
		T CD8 ⁺ de memoria efectora				
	T Reguladores	T regulador	T regulador	T regulador		
			T regulador ayudante folicular			
	T Cooperadores	T cooperador 1				
		T cooperador 2				
	T Gamma Delta	T Gamma Delta	T Gamma Delta			
	Asesinos naturales	T Asesinos naturales				
		Asesinos naturales			Asesinos naturales	Asesinos naturales
		Asesinos naturales descansando				
		Asesinos naturales activados				
Leucocitos	Basófilos	Basófilos				
	Dendríticas	Dendríticas	Dendríticas descansando	Dendríticas		
		Dendríticas activadas	Dendríticas activadas			
		Dendríticas convencionales				
		Dendríticas inmaduras				
		Dendríticas plasmacitoides				
	Eosinófilos	Eosinófilos	Eosinófilos			
	Macrófagos	Macrófagos			Macrófagos	
		Macrófagos M0				

Población		xCell	Cibersort	quanTIsec	EPIC
Leucocitos	Macrófagos	Macrófagos M1	Macrófagos M1	Macrófagos M1	Macrófagos
		Macrófagos M2	Macrófagos M2	Macrófagos M2	
	Mastocitos	Mastocitos	Mastocitos descansando		
			Mastocitos activados		
	Monocitos	Monocitos	Monocitos	Monocitos	
	Neutrófilos	Neutrófilos	Neutrófilos	Neutrófilos	

TABLA IV: Configuración de métodos de deconvolución elegidos para el proyecto.

Paquete	Configuración
ImmuCC	Método = Support Vector Regression (SVR) Plataforma = Microarray para GSE64658 y RNASeq para GSE57533
Cibersort	Gene Signature = LM22 (22 tipos de células inmunes) Permutation = 100 Quantile Normalization = FALSE (Recommended For RNA-Seq) Absolute = FALSE
quanTIseq	--pipelinestart=decon --tumor=TRUE
xCell	rnaseq = TRUE
Epic	Valores por defecto

3.9 Análisis de sobrevida

El análisis de sobrevida de los pacientes se realizó con el paquete Survival, utilizando el estimador Kaplan-Meier para construir cada curva. Además, se utilizó la prueba de Log-Rank para comparar las distintas curvas de sobrevida. Para las variables continuas (ej. expresión de un gen), se utilizó surv_cutpoint de survminer para encontrar el punto óptimo de corte y transformarlas a una variable dicotómica, asignándoles un valor alto o bajo.

3.10 Software

Para la descarga y procesamiento de los archivos de RNA-seq se escribió un pipeline en BASH. El resto de los pipelines se realizaron en el Lenguaje R versión 3.6.1 y RStudio 1.2.x. Todos los paquetes utilizados se encuentran disponibles o bien en el repositorio CRAN, o bien en BioConductor (Huber *et al.*, 2015).

Todo el código escrito para este proyecto se encuentra disponible en la plataforma [GitHub](https://github.com/chrismaazzeo/Tesis_Marcadores_Glicoinmunologicos) (https://github.com/chrismaazzeo/Tesis_Marcadores_Glicoinmunologicos).

3.11 Hardware

Los pipelines relacionados con RNA-seq se ejecutaron en un servidor con Core i7 con 16GB RAM, disco de 2TB y sistema operativo Linux Ubuntu. El procesamiento de los microarray, comparaciones entre grupos y gráficos se ejecutaron en una Mac mini Intel Core i7 con 8GB de RAM, disco SSD de 250GB y sistema operativo Mac OS *El Capitan*.

4. Resultados

4.1 Análisis de transcriptómica en modelos experimentales de cáncer colorrectal y enfermedades inflamatorias intestinales

4.1.1 *Desregulaciones en la expresión de glicogenes en modelos experimentales de CCRE (Apc^{min/+}) y Asociado a Colitis (AOM-DSS)*

Con el objetivo de estudiar las alteraciones en la glicobiología intestinal para estos dos modelos experimentales de cáncer colorrectal, en una primera instancia se diseñó un flujo de trabajo (*pipeline*) para procesar y analizar las muestras del experimento, el cual se presenta en la Figura 4.

El *pipeline* está conformado por seis grupos de tareas:

1. Preprocesamiento: agrupa las tareas de descarga de los archivos de transcriptómica, junto con sus datos adicionales (identificación de cada muestra, datos clínicos en humanos, etc.) y el procesamiento de éstos como base del análisis.
2. Análisis de calidad (QA): son las tareas necesarias para el aseguramiento de la calidad de los datos de transcriptómica y su corrección en caso de baja calidad de datos.
3. Análisis descriptivo: son las tareas para obtener información sobre cómo se agrupan las muestras a partir de diferentes métodos estadísticos.
4. Comparación entre grupos: es una de las tareas más importantes, ya que obtenemos como resultado la expresión diferencial de cada gen.
5. Análisis posterior (*Downstream Analysis*): nos permite analizar visualmente los resultados del paso anterior.
6. Caracterización del infiltrado inmune: tareas para identificar las variaciones del infiltrado inmune entre grupos por uno o más algoritmos.

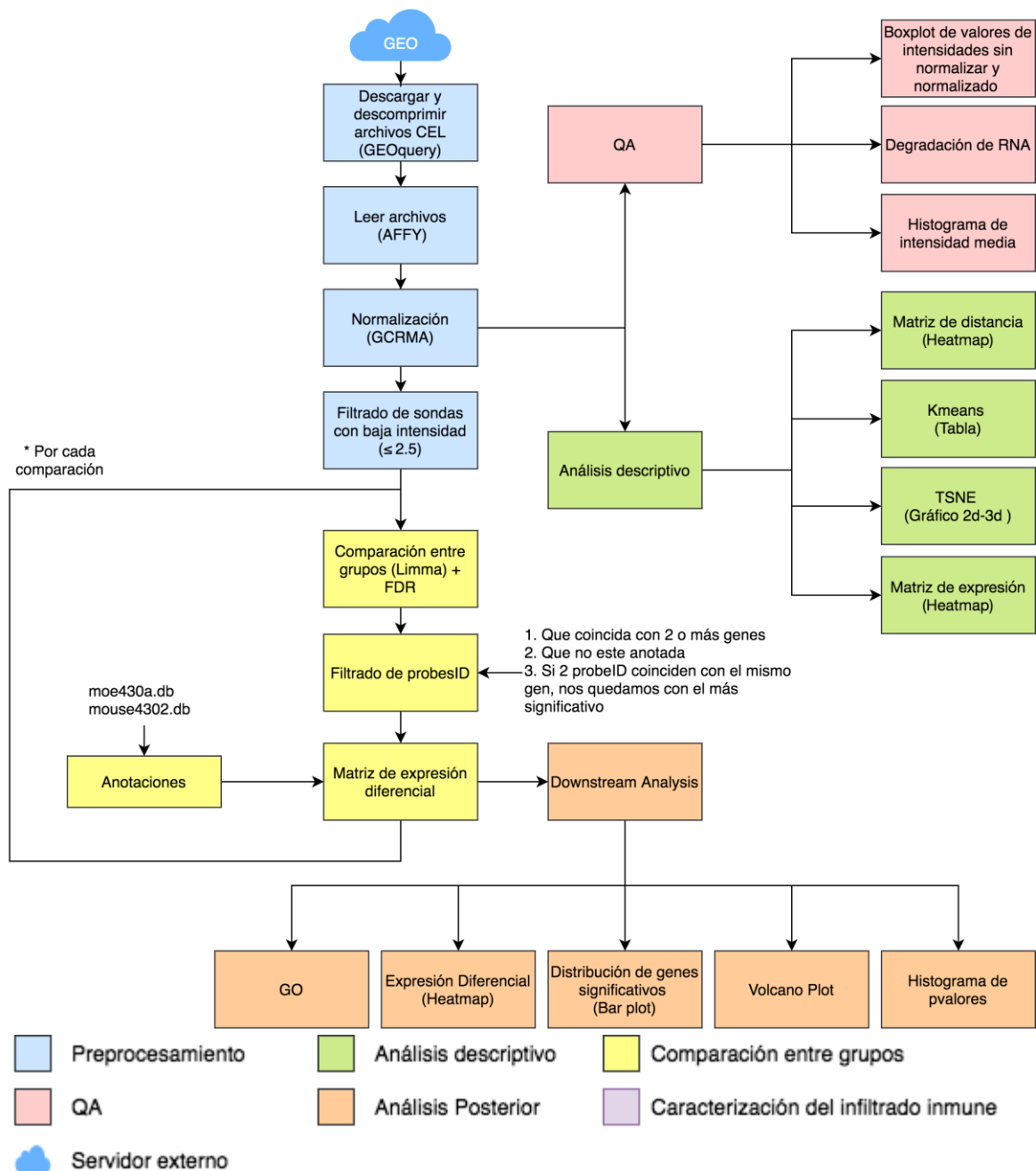


Figura 4: Flujo de trabajo utilizado para el procesamiento y análisis de la base de datos GSE43338, con estudios en modelos murinos experimentales de CCRAC (AOM-DSS) y CCRE (*Apc^{min/+}*).

Luego, se analizó la expresión diferencial de todos los genes y de los genes de interés glicobiológico seleccionados en la base de datos GSE43338 que contiene datos para modelos murinos experimentales de CCRAC (AOM-DSS) y CCRE (*Apc^{min/+}*). Los genes significativos ($p\text{-adj} \leq 0,05$) se filtraron según el valor de \log_2FC , tal como se indica en la Tabla V.

TABLA V: Resultados de la comparación entre grupos del modelo de CCRAC y CCRE (GSE43338).

Grupos	Filtro	Genes significativos (p-adj ≤ 0,05)	Glicogenes significativos (p-adj ≤ 0,05)
CCRAC vs. CCRAC Control	$ \log_2FC \geq 0$	4310	199
	$ \log_2FC \geq 1$	2485	131
	$ \log_2FC \geq 2$	861	60
CCRE vs. CCRE Control	$ \log_2FC \geq 0$	5390	253
	$ \log_2FC \geq 1$	2927	154
	$ \log_2FC \geq 2$	1157	70

4.1.1.1 Alteraciones en el transcriptoma de colon en un modelo experimental de CCRAC (AOM-DSS)

En el análisis de expresión diferencial de genes comparando muestras de tumor (animales tratados con AOM-DSS) vs. muestras control, obtuvimos un total de 4310 genes significativamente desregulados ($p\text{-adj} \leq 0,05$), de los cuales 199 codifican para glicoproteínas y lectinas, o están involucrados en biosíntesis y degradación de glicoconjugados (glicogenes). Aplicando un filtro con criterio $|\log_2FC| \geq 1$, obtuvimos 2485 genes, de los cuales 131 son glicogenes (Figura 5).

Notablemente, el análisis de los genes más desregulados ($|\log_2FC| \geq 2$) define una firma glicobiológica de 26 glicogenes aumentados y 34 disminuidos en su expresión (Figura 5). El análisis detallado de la funcionalidad de los mismos muestra sobreexpresión de genes de ciertas galectinas (*Lgals1* y *Lgals2*), clusters de diferenciación y receptores asociados a células inmunes (*Cd44*, *Mrc1*), biosíntesis de glicolípidos (*Degs1*), (glico)proteínas involucradas en el reconocimiento, degradación de proteoglicanos y glicosaminoglicanos (*Notum*, *Bgn*, *GusB*, *Ptn*, *Dcn*, *Sulf2*), proteoglicanos (*Gpc1*), metabolismo (*Pfkfb3*, *Plaur*), glicoproteínas de función inmune (*Srgn*, *Icam1*, *Vcam1*, *Pecam1*), lectinas tipo C (*Cd93*) con reconocimiento de glucanos (*Clec7a*) y mananos (*Clec4n*), glicosiltransferasas (*B4galt6* y *Galnt1*), receptores de células endoteliales involucrado en procesos de coagulación (*Thbd*) y glicoproteínas de degradación lisosomal (*Lamp2*). En tanto, se observa la disminución de la expresión génica de ciertas glicosiltransferasas (galactosiltransferasas *B3galt5* y *B4galt4*, GlcNAc transferasas

B3Gnt2, *A4GalT*, *Gcnt1* y 3, sialiltransferasas *St3Gal4* y *St6gal1*, GalNAc transferasas *Galnt6* y *Galnt10*, fucosiltransferasa *Fut4* y glucuroniltransferasas (*Ugt2b34*), enzimas involucradas en la biosíntesis de azúcares nucleótido y monosacáridos (*Gfpt1*, *Nans*, *Ugdh*), enzimas relacionadas con el metabolismo de azúcares (*Pcx*, *AldoB*, *Hk1*, *Chga*, *Klb*, *Thnsl2*), señalización por fgf (*Fgfr3*), sulfotransferasas (*Chst4*), proteoglicanos (*Sdc2*), glicosidasas (*Neu1*, *HexB*), lectinas tipo C (*Clec2h*), hidrólisis de glicolípidos (*Acer1*), lipasas (*Aadac*, *Ces2g*, *Ces1f*, *Ces1d*, *Ces2a*) y transportadores de sulfato y glucosa (*Slc26a2*, *Slc5a1*, respectivamente). En particular, *Lyz1* y *Lyz2*, dos glicosilhidrolasas de peptidoglicanos bacterianos, se encuentran marcadamente sobreexpresadas en este modelo experimental.

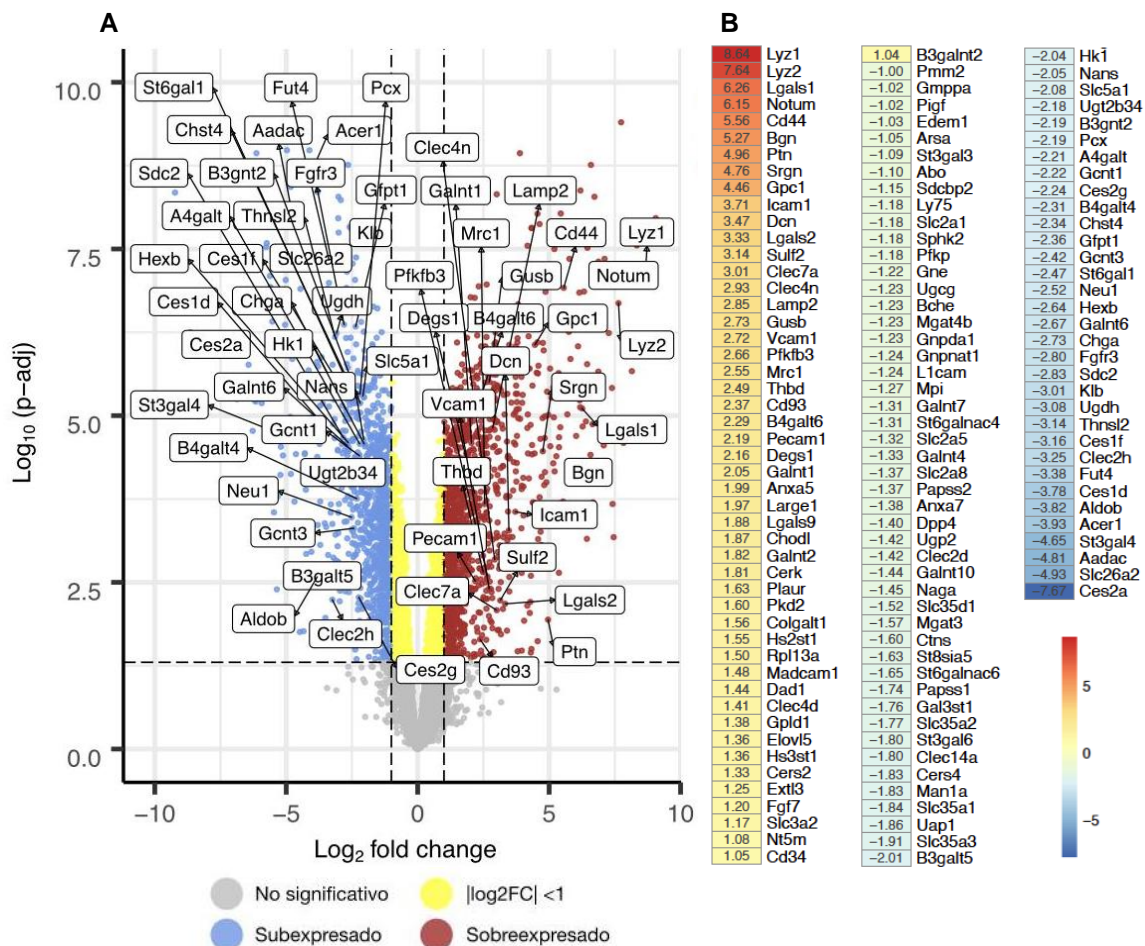


Figura 5: Genes desregulados en el modelo de cáncer colorrectal asociado a colitis por AOM-DSS, comparando con control. **A.** Volcano plot de glicogenes significativos detallando aquellos con $|\log_2FC| \geq 2$. **B.** Heatmap de glicogenes significativos con $|\log_2FC| \geq 1$.

4.1.1.2 Alteraciones en el transcriptoma de colon en un modelo experimental de CCRE (*APC^{min/+}*)

En el análisis de expresión diferencial de genes comparando muestras de tumor vs. muestras control en animales *Apc^{min/+}* obtuvimos un total de 5390 genes significativamente desregulados ($p\text{-adj} \leq 0,05$), de los cuales 253 clasifican como glicogenes. Aplicando un filtro con criterio $|\log_2FC| \geq 1$, la desregulación se reduce a 2927 genes de los cuales, 154 son glicogenes (Figura 6). Notablemente, el análisis de los genes más desregulados ($|\log_2FC| \geq 2$) define una firma glicobiológica de 29 genes con expresión aumentada y 41 con expresión disminuida (Figura 6A).

De particular interés, entre los genes cuya expresión se encuentra más aumentada (tanto en este modelo de CCRE como en el de CCRAC analizado en la sección anterior), el gen *Ptn* (*Pleiothropin*) codifica para una citoquina con actividad de lectina. Esta proteína se ha visto implicada en la angiogénesis tumoral, como agente proliferativo y anti-apoptótico. Por ejemplo, en células MCF7 (cáncer de mama) la expresión de *Ptn* remodela la matriz extracelular, aumentando la angiogénesis y estimulando las células estromales (Sorrelle *et al.*, 2017). Aunque los estudios en modelos animales de CCR centrados en el papel de *Ptn* son escasos, se ha identificado a *Ptn* como un factor clave en el desarrollo de CCR, donde nuevamente está involucrada en angiogénesis. En esta patología de hecho se ha propuesto a *Ptn* como factor de pronóstico, y como blanco terapéutico dada su baja expresión en tejido normal (Kong *et al.*, 2012).

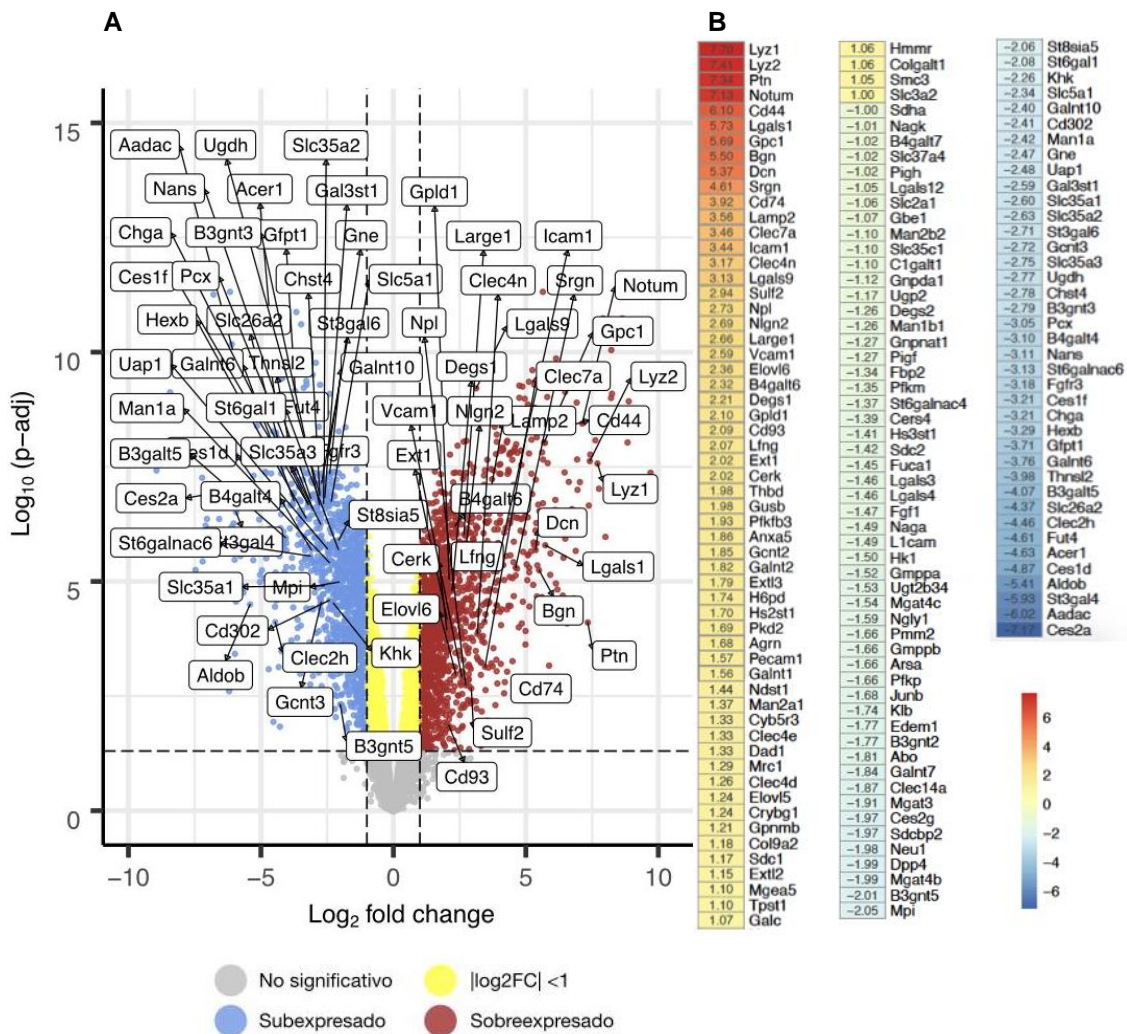


Figura 6: Genes desregulados en su expresión en el modelo experimental de cáncer colorrectal esporádico, comparando con control. **A.** Volcano plot de glicógenos significativos detallando aquellos con $|\log_2FC| \geq 2$. **B.** Heatmap de glicógenos significativos con $|\log_2FC| \geq 1$.

Al evaluar comparativamente los genes significativamente desregulados ($p \leq 0,05$) con una tasa de cambio $|\log_2FC| \geq 2$ entre el modelo de CCRAC (AOM-DSS) comparado con su control (Sección 4.1.1.1) y el modelo de cáncer colorrectal esporádico comparado con su control respectivo (esta sección), se observa que si bien cada modelo presenta singularidades (con tres glicógenos desregulados de manera particular para CCRAC y 12 glicógenos desregulados específicamente para CCRE), la mayor parte de los glicógenos (72) se mantienen sub- o sobreexpresados en ambos modelos (Figura 7), lo que indica la existencia de mecanismos de desregulación comunes a ambas patologías.

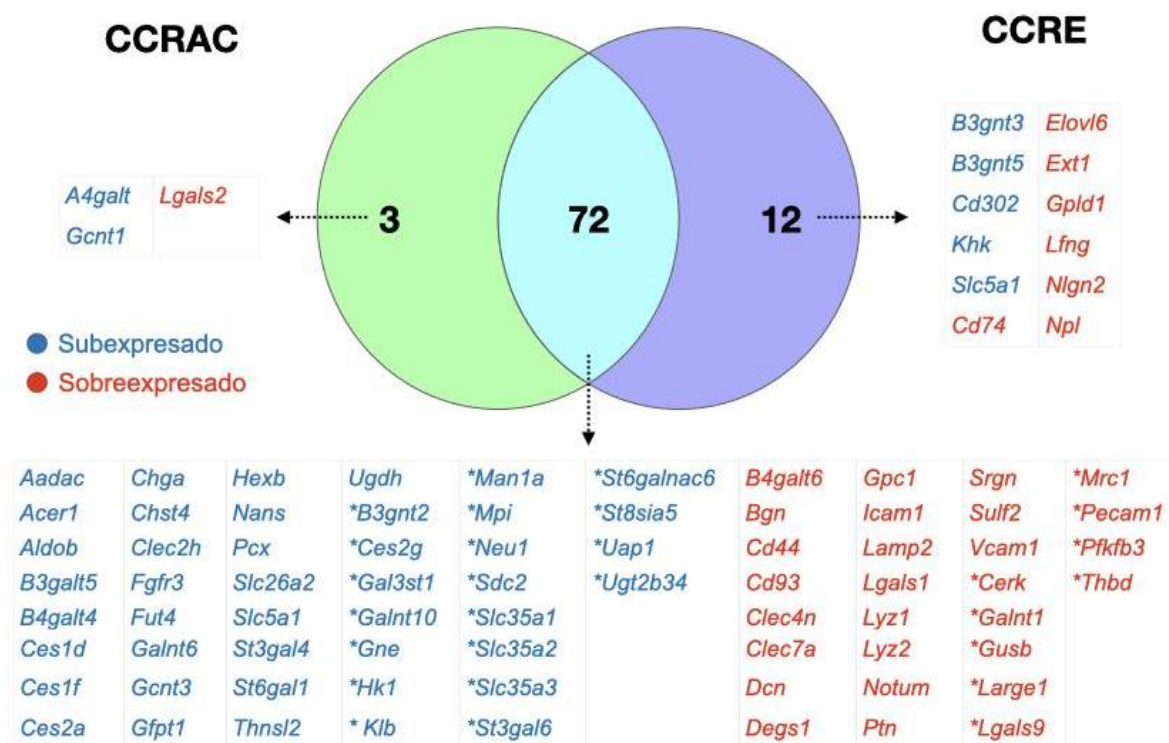


Figura 7: Diagrama de Venn de CCRAC vs. CCRAC Control U CCRE vs. CCRE Control, glicogenes significativos ($p\text{-adj} \leq 0,05$) y $|\log_2FC| \geq 2$. (*) Indica que el glicogen presenta valor de expresión con $|\log_2FC| \geq 2$ en uno de los modelos, pero que en el otro modelo su valor de expresión es $1 \leq |\log_2FC| \leq 2$.

En el análisis detallado del significado funcional de estos cambios, los 72 genes que se mantienen desregulados en ambos modelos incluyen la sobreexpresión de glicohidrolasas de peptidoglicanos bacterianos *Lyz1* y *Lyz2*, clusters de diferenciación (*CD44*, *CD93*), biosíntesis de glicolípidos (*Degs1*), (glico)proteínas involucradas en el reconocimiento, degradación de proteoglicanos y glicosaminoglicanos (*Notum*, *Bgn*, *Ptn*, *Dcn*, *Sulf2*), proteoglicanos (*Gpc1*), glicoproteínas de función inmune (*Srgn*, *Icam1*, *Vcam1*), Galectina-1, lectinas tipo C con reconocimiento de glucanos (*Clec7a*) y mananos (*Clec4n*), y la glicosiltransferasa *B4galt6* y glicoproteínas de degradación lisosomal (*Lamp2*). En tanto, la disminución en la expresión de genes común a ambas patologías se presenta en algunas glicosiltransferasas (*B3gnt2*, *B4galt4*, *Gcnt3*, *Galnt6*, *St3gal4*, *St6gal1* y *B3galt5*), en enzimas involucradas en la biosíntesis de azúcares nucleótido y monosacáridos (*Gfpt1*, *Nans*), enzimas relacionadas con procesos metabólicos (*Pcx*, *aldoB*, *Chga*, *Thns12*), sulfotransferasas (*Chst4*), glicosidasas (*Hexb*), lectinas tipo C (*Clec2h*), ceramidasa alcalina 1 (*Acer1*), lipasas (*Aadac*, *Ces2g*, *Ces1f*, *Ces1d*,

Ces2a, *Ces2s*), transportadores de sulfato y glucosa (*Slc26a2*, *Slc5a1*, respectivamente) y *LARGE1*, una glicosiltransferasa involucrada en la extensión de O-manosilación. Asimismo, ciertas enzimas que aparecen desreguladas en este modelo de cáncer colorrectal esporádico también aparecen desreguladas en el modelo de AOM-DSS pero con valores menores al corte de $|\log_2FC| \geq 2$, como la disminución de las enzimas involucradas en la biosíntesis de UDPGlcNAc (*Uap1*), Manosamina para ácido siálico (*Gne*), *Slc35a1*, *Slc35a2* y *Slc35a3* (Transportadores de CMP-siálico, UDP-gal y UDPGlcNAc, respectivamente), glicosiltransferasas *St3gal6*, *St6galnac6*, *Galnt10*, *St8sia5*, la galactosilceramida sulfotransferasa *Gal3st1*, y la galectina *Lgals9*.

Sin embargo, el desarrollo de las patologías de base genética presenta ciertas diferencias. En primer lugar, el aumento de *A4galt*, *Gcnt1* y *Lgals2* observado en el modelo de AOM-DSS no se reproduce en el modelo de CCRE. En segundo lugar, el modelo de CCRE se diferencia del de CCRAC (AOM-DSS) por un aumento de expresión de *CD74* (marcador de células B), la glicosiltransferasa *Lfn3* (lunatic fringe), *Npl* (una enzima involucrada en la degradación de ácido siálico), y la disminución de la lectina tipo C *CD302*, así como de ketokinasa (*Khk*, responsable de la transformación metabólica de fructosa por glucosa). En tanto, en el modelo de cáncer colorrectal esporádico, la disminución en la expresión de *B3gnt3* y *B3gnt5*, no están presentes en AOM-DSS, aunque ambos comparten la disminución de expresión en de *B3gnt2*.

4.1.1.3 Caracterización del infiltrado inmune

Dado que ImmuCC (software utilizado para el análisis del infiltrado inmune tumoral en modelos experimentales murinos, Chen et al., 2017; Chen et al., 2018) no soporta el procesamiento de microarray Affymetrix 430A, no se pudo caracterizar el infiltrado inmune de este estudio GSE43338. En consecuencia, la desregulación de la expresión de marcadores inmunes como los genes *CD302* (marcador de poblaciones mieloides) o *CD74* (marcador de linfocitos B) no pudo correlacionarse con el análisis de poblaciones inmunes.

4.1.2 *Desregulaciones en la expresión de glicogenes en un modelo de CCRAC (AOM-DSS) analizado según localización anatómica y cronicidad de la inflamación intestinal*

En segunda instancia, se analizó la expresión diferencial en modelos experimentales de CCRAC (AOM-DSS), pero en este caso considerando la influencia de la inflamación crónica (comparando dos y cuatro ciclos de inflamación intestinal con DSS, denominados temprano y tardío respectivamente) y de la zona del colon proximal o distal, utilizando la base de datos de los modelos experimentales murinos GSE64658 (Tabla I).

En el análisis realizado entre CCRAC Proximal Temprano vs. Control Proximal, se observaron 1922 genes desregulados de manera significativa, y al analizar la desregulación de genes en CCRAC Proximal Tardío vs. Control Proximal, se encontraron sólo 5 genes desregulados (Tabla VI). Cuando se comparó CCRAC Proximal Tardío vs CCRAC Proximal Temprano, sólo se vieron desregulados 505 genes.

Debido a que el modelo experimental de CCRAC basado en la administración de AOM-DSS genera tumores en el colon distal y raramente en la región proximal (De Robertis *et al.*, 2011; Gkouskou *et al.*, 2016), se decidió centrar el análisis en la región distal.

El flujo de trabajo es el mismo que el utilizado para el GSE anterior (Figura 4). Los genes significativos ($p\text{-adj} \leq 0,05$) se filtraron según el valor de \log_2FC como se indica en la Tabla VI.

TABLA VI: Resultados de la comparación entre grupos del modelo de CCRAC Proximal/Distal - Temprano/Tardío (GSE64658).

Grupos	Filtro	Genes significativos ($p\text{-adj} \leq 0,05$)	Glicogenes significativos ($p\text{-adj} \leq 0,05$)
CCRAC Proximal Temprano vs. CCRAC Control Proximal	$ \log_2FC \geq 0$	1922	59
	$ \log_2FC \geq 1$	964	38
	$ \log_2FC \geq 2$	355	18
CCRAC Proximal Tardío vs. CCRAC Control Proximal	$ \log_2FC \geq 0$	5	1
	$ \log_2FC \geq 1$	3	1
	$ \log_2FC \geq 2$	3	1

Grupos	Filtro	Genes significativos (p-adj ≤ 0,05)	Glicogenes significativos (p-adj ≤ 0,05)
CCRAC Distal Temprano vs. CCRAC Control Distal	$ \log_2FC \geq 0$	2150	95
	$ \log_2FC \geq 1$	1314	65
	$ \log_2FC \geq 2$	619	34
CCRAC Distal Tardío vs. CCRAC Control Distal	$ \log_2FC \geq 0$	7017	272
	$ \log_2FC \geq 1$	3247	129
	$ \log_2FC \geq 2$	1183	55

Para la caracterización del infiltrado inmune por ImmuCC se procedió con el flujo de trabajo representado en la Figura 8.

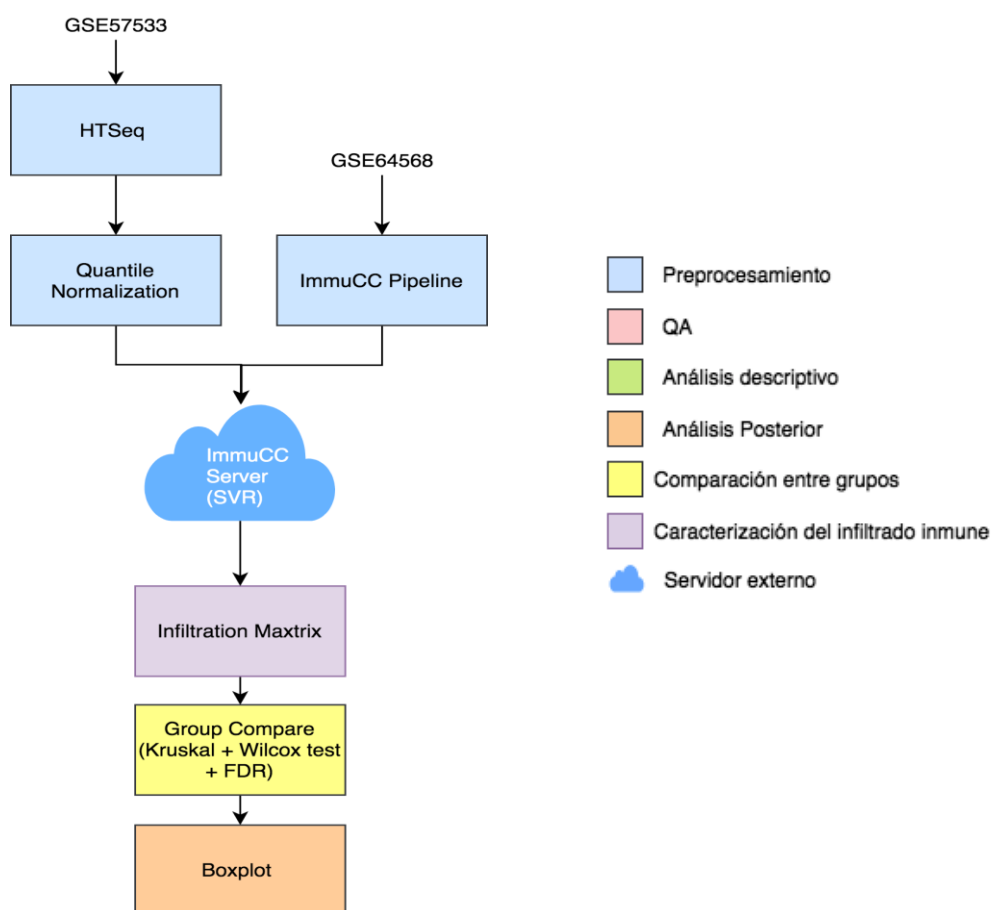


Figura 8: Flujo de trabajo utilizado para la caracterización del infiltrado inmune en la base de datos de modelos experimentales de CCRAC GSE64658.

4.1.2.1 Alteraciones en el transcriptoma de colon en modelo experimental de CCRAC (AOM-DSS) Distal Temprano

En el análisis de expresión diferencial de genes comparando tumor vs. control en animales tratados con AOM-DSS Temprano (dos ciclos de administración de DSS) y región distal del colon, obtuvimos un total de 2150 genes significativamente desregulados ($p\text{-adj} \leq 0,05$), de los cuales 95 son glicogenes. Notablemente, el análisis de los genes más desregulados ($|\log_2FC| \geq 2$) define una firma glicobiológica de 29 genes aumentados y 4 disminuidos en su expresión (Figura 9A). Aplicando un filtro con criterio $|\log_2FC| \geq 1$, obtuvimos 1314 genes, de los cuales 65 son glicogenes (Figura 9B).

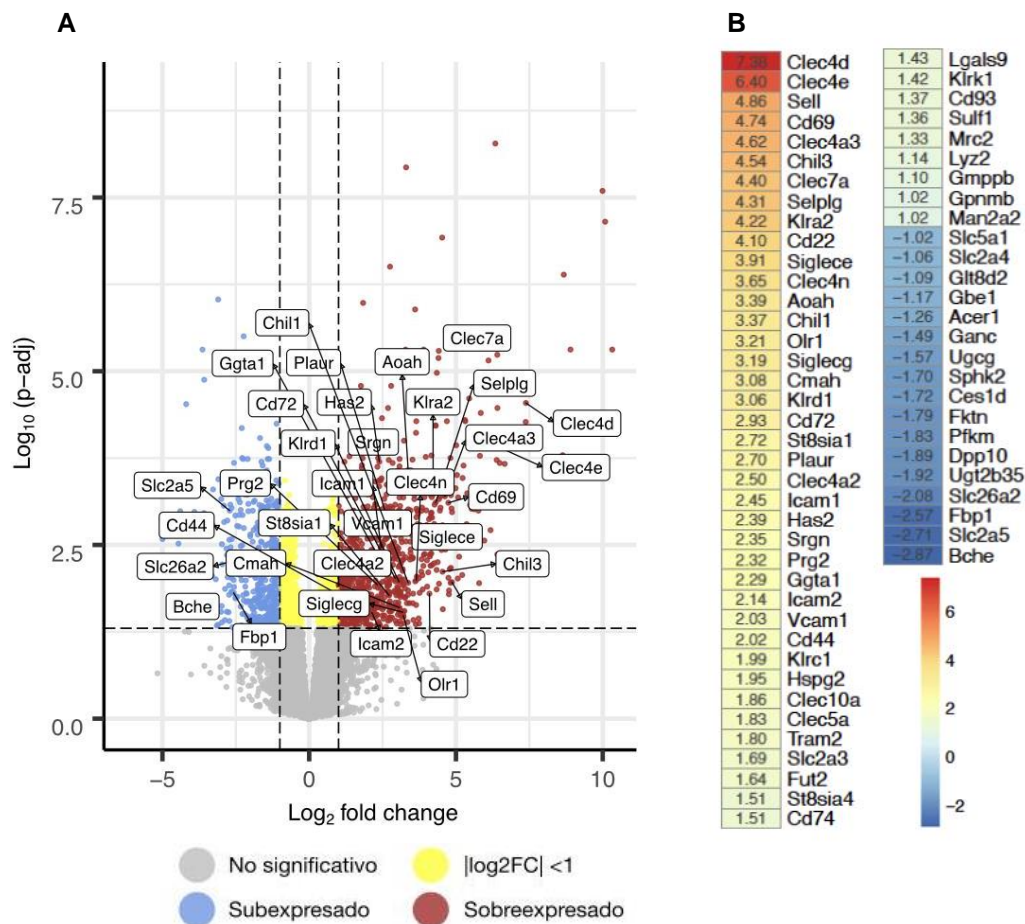


Figura 9: Genes desregulados en su expresión en cáncer colorrectal asociado a colitis distal temprano, comparado con control distal. **A.** Volcano plot de glicogenes significativos detallando aquellos con $|\log_2FC| \geq 2$. **B.** Heatmap de glicogenes significativos con $|\log_2FC| \geq 1$.

4.1.2.2 Alteraciones en el transcriptoma de colon en modelo experimental de CCRAC (AOM-DSS) Distal Tardío

En el análisis de expresión diferencial de genes comparando tumor vs. control en animales tratados con AOM-DSS ciclo Tardío y región Distal, obtuvimos un total de 7017 genes significativamente desregulados ($p\text{-adj} \leq 0,05$), de los cuales 272 codifican para glicogenes. Notablemente, el análisis de los genes más desregulados ($|\log_2FC| \geq 2$) define una firma glicobiológica de 40 genes aumentados y 15 disminuidos en su expresión (Figura 10A). Aplicando un filtro con criterio $|\log_2FC| \geq 1$, obtuvimos 3247 genes, de los cuales 129 son glicogenes (Figura 10B).

El análisis detallado de la funcionalidad de los mismos muestra sobreexpresión de glicoproteínas involucradas en el reconocimiento y degradación de proteoglicanos y glicosaminoglicanos (*Notum*, *Ptn*), lectina antibacteriana *Reg3b*, glucohidrolasa (*Chil3*), clusters de diferenciación y receptores asociados a células inmunes (*CD44*), proteoglicanos (*Gpc1*), metabolismo (*Pfkfb3*, *Plaur*), glicoproteínas de función inmune (*Srgn*, *Icam1*), lectinas tipo C (*Clec4d* y *Clec4e*), lectinas con reconocimiento de glucanos (*Clec7a*) y mananos (*Clec4n*), y subexpresión de hidrólisis de glicolípidos (*Acer1*) y lipasas (*Ces1d*), enzimas involucradas en el metabolismo de azúcares (*Klb*), metabolismo de ácido siálico (sialidasa *Neu1*), glicosiltransferasas (*Galnt6*, *St3gal4*) y carboxilesterasa (*Bche*).

En particular, *Notum* es una carboxilesterasa que se encuentra marcadamente sobreexpresada en este modelo experimental. Esta proteína se ha visto sobreexpresada en células metastásicas en pacientes de CCR y su inhibición suprime la tanto la proliferación como la migración tumoral y se ha propuesto como factor de pronóstico y blanco terapéutico (Yoon *et al.*, 2018).

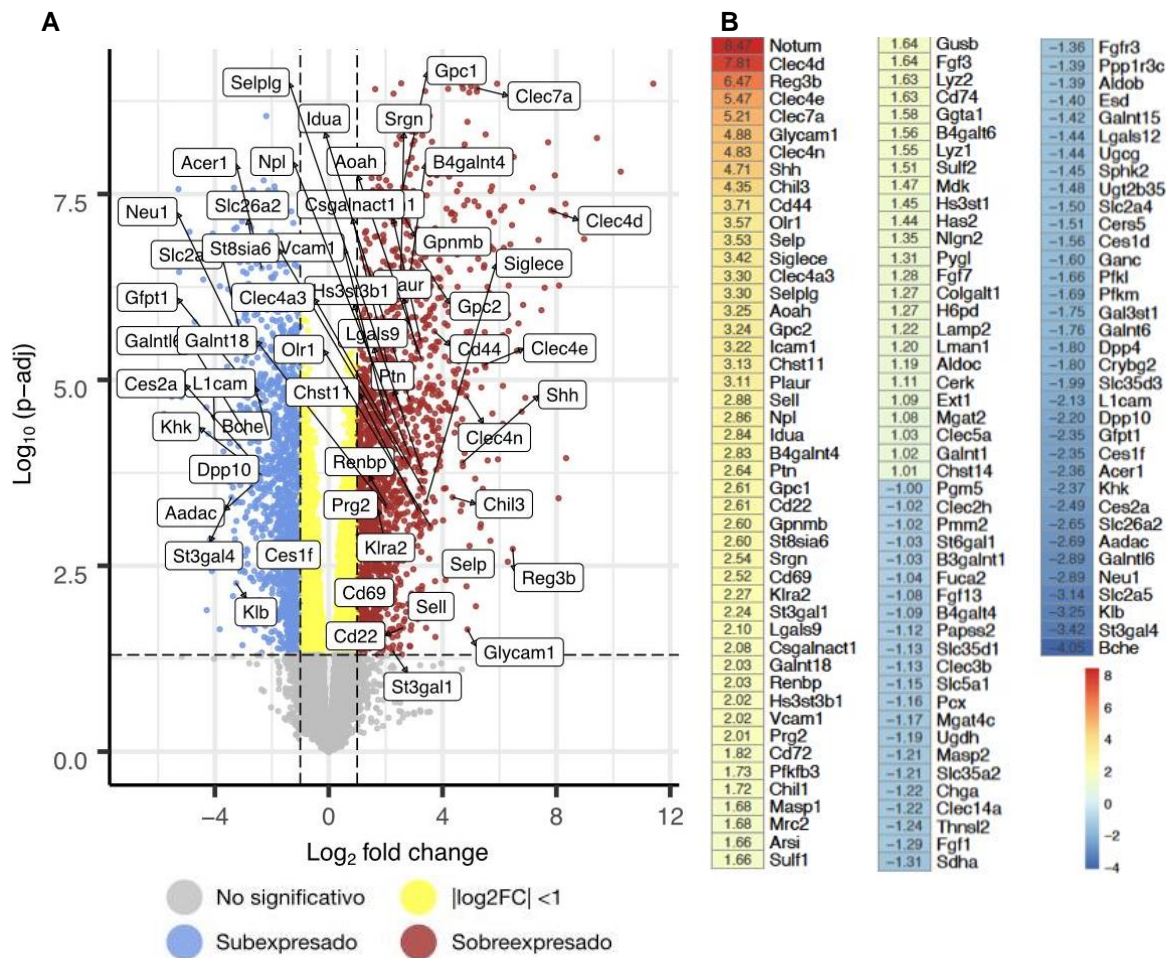


Figura 10: Genes desregulados en su expresión en el modelo de AOM-DSS distal tardío, comparado con control distal. **A.** Volcano plot de glicogenes significativos detallando aquellos con $|\log_2FC| \geq 2$. **B.** Heatmap de glicogenes significativos con $|\log_2FC| \geq 1$.

Al evaluar comparativamente los genes desregulados con una tasa de cambio $|\log_2FC| \geq 2$ entre los modelos de cáncer colorrectal asociado a colitis distal temprano y distal tardío (ambos relativizados a su control), notamos que cada modelo se desregula de manera singular. Con siete glicogenes desregulados de manera particular para CCRAC distal temprano, 30 genes desregulados para CCRAC distal tardío, y 29 genes desregulados para ambos modelos que se mantienen sub o sobreexpresados (Figura 11).

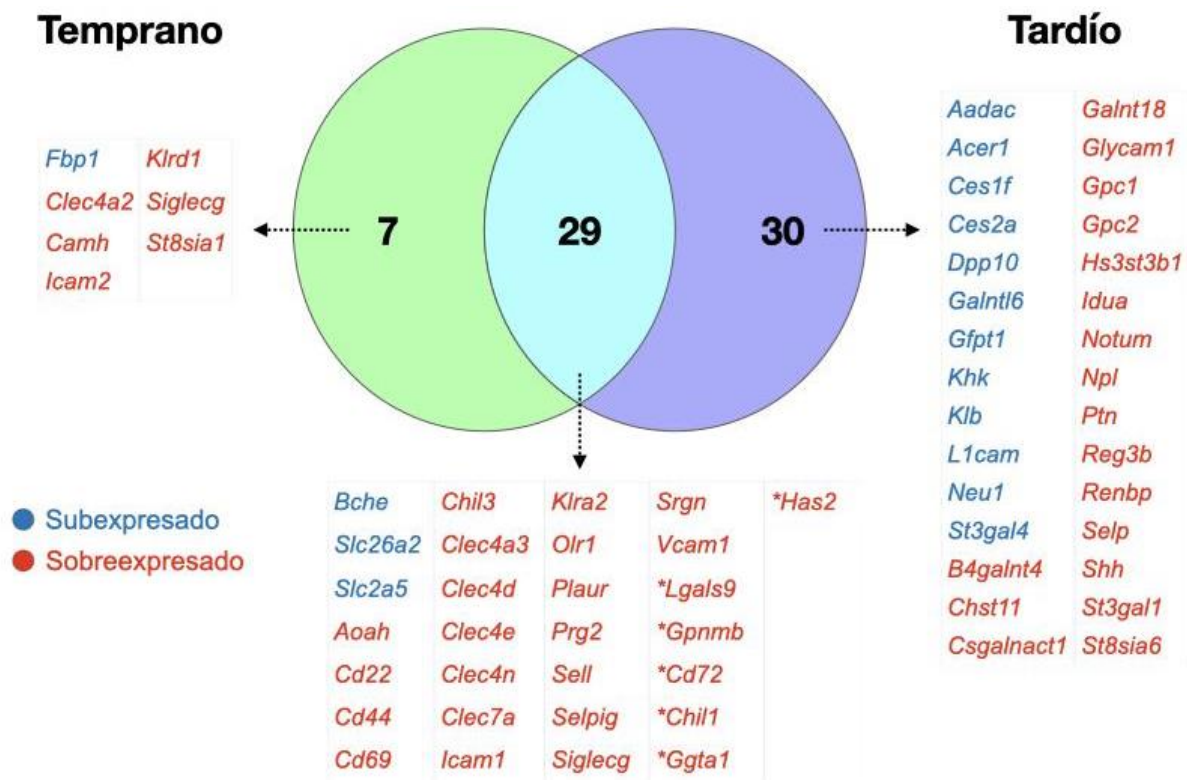


Figura 11: Diagrama de Venn de CCRAC Distal temprano vs. CCRAC Control U CCRAC Distal tardío vs. CCRAC Control, glicogenes significativos ($p\text{-adj} \leq 0,05$) y $|\log_2FC| \geq 2$. (*) Indica que para ese glicogen corresponde un valor de expresión $|\log_2FC| \geq 2$ en uno de los modelos, pero que en el otro modelo es $1 \leq |\log_2FC| \leq 2$.

Los resultados resumidos en la Figura 11 muestran que la inflamación crónica (cuatro ciclos de administración de DSS) provoca la desregulación de un mayor número de genes. En ese sentido, muchos de los mismos ya habían sido detectados en el primer análisis del modelo de AOM-DSS, ya que el mismo era basado en un diseño experimental para tres ciclos de DSS (como ejemplos, *Aadac*, *Acer1*, *Ces1f/2a*, *Klb*, *Notum*, *Neu1*, *Ptn*). En tanto, 29 genes aparecen desregulados ya desde el diseño experimental temprano (entre ellos *CD44*, *Lgals9*, *Icam1*, *Slc26a2*, *Srgn*, también detectados en el análisis del GSE anterior, Sección 4.1.1.1) lo cual valida su desregulación y también indica su papel en mecanismos de desarrollo neoplásico desde etapas iniciales de la patología. Aquellos genes desregulados a nivel distal temprano que no se encuentran en el modelo distal tardío (*Fbp1*, *Clec4a2*, *Camh*, *Icam2*, *Klr1*, *Siglecg*, *St8sia1*) y tampoco se observan desregulados en el primer análisis de AOM-DSS realizado (Sección 4.1.1.1), lo cual valida su desregulación temprana (antes de tres ciclos de DSS o más).

Estas desregulaciones tempranas ausentes tanto en este dataset en distal tardío como en el análisis del dataset inicial de AOM-DSS podrían explicarse desde un punto de vista biológico como eventos requeridos para el desarrollo inicial de la neoplasia. Las diferencias observadas con el análisis anterior pueden deberse a variaciones en la microbiota, las condiciones alimentarias y ambientales del bioterio, el sexo de los animales (machos en el primer GSE, hembras en éste) o la tecnología de transcriptómica utilizada en cada experimento (microarray o RNA-seq, respectivamente).

4.1.2.3 Caracterización del infiltrado inmune en modelo experimental de CCRAC (AOM-DSS) Distal

A diferencia del primer estudio realizado, GSE57533, el estudio transcriptómico GSE64658 se realizó con la técnica de RNA-seq, compatible con el método de deconvolución ImmuCC para el estudio del infiltrado tumoral en muestras murinas. Por lo tanto, se analizó el infiltrado inmune de los distintos modelos experimentales del estudio GSE64658 utilizando el método ImmuCC (Chen et al. 2017, 2018), focalizándose en la región distal del colon.

De las 25 poblaciones inmunes caracterizadas en ImmuCC, se encontraron diferencias significativas en células dendríticas activadas, neutrófilos, células asesinas naturales (*natural killers*, NK) en reposo, células NK activadas, mastocitos, células plasmáticas, linfocitos T $CD4^+$ foliculares (Thf) y linfocitos T $CD8^+$ vírgenes (*naive*). Por el contrario, no se encontraron diferencias significativas en otras poblaciones inmunes tales como macrófagos, linfocitos B, linfocitos T efectoras ni linfocitos T reguladores (Tregs) (Figura 12).

Tanto el aumento en la proporción de células dendríticas activadas y linfocitos T $CD4^+$ foliculares, como la disminución de las células NK en reposo, ya se evidencian en el modelo temprano de CCRAC, comparado con el control, y estas diferencias en el infiltrado inmune se mantienen y se hacen más pronunciadas en el modelo tardío. Por el contrario, las poblaciones restantes sólo se ven significativamente alteradas en el modelo tardío en comparación con el control. Al correlacionar las poblaciones inmunes alteradas según ImmuCC con el perfil de genes desregulados en el análisis de transcriptómica, se observa que *CD22*, *CD72*, *CD44*, *CD69*, *Clec4n* y *Clec7a* se encuentran sobreexpresados, tanto en el modelo de CCRAC temprano como tardío, en comparación con el control.

En primer lugar, *CD22* es un correceptor presente en linfocitos B cuya función es inhibir al receptor de linfocitos B (*B cell receptor*, BCR) y regular de manera negativa la activación de

estas células (Walker y Smith, 2008). Considerando que la expresión de *CD22* se encuentra disminuída en células plasmáticas, y que las otras poblaciones de linfocitos B no se encuentran alteradas, la sobreexpresión observada en CCRAC podría estar reflejando el aumento de células plasmáticas, dado que se sabe que en los modelos experimentales de AOM-DSS se encuentran aumentadas (Paradisi *et al.*, 2009).

Continuando con marcadores de linfocitos B, *CD72* es un correceptor para la activación de linfocitos B. Se expresa en todas las etapas del desarrollo de los linfocitos B, a excepción de las células plasmáticas. La unión con su ligando mejora el crecimiento y la diferenciación de linfocitos B (Wu y Bondada, 2002). En este caso, no se observa una correlación entre el resultado de transcriptómica y el descripto por el análisis de ImmuCC, ya que la población de linfocitos B en este algoritmo no se encuentra alterado. Por otro lado, *CD72* también se expresa en macrófagos, células dendríticas y algunas poblaciones de linfocitos T (Kumanagoh y Kikutani, 2002). Por lo tanto, la sobreexpresión de *CD72* observada en el análisis de expresión diferencial podría estar correlacionada con el aumento de células dendríticas activadas o de linfocitos Thf o $CD8^+$ *naive*, observado por ImmuCC.

En tanto, *CD44* es un receptor de membrana que está involucrado en numerosas funciones, entre las cuales está la activación de linfocitos T. Asimismo, la expresión de *CD44* es un marcador de linfocitos T efectores de memoria y se encuentra presente en linfocitos T auxiliares foliculares B (Thf). La sobreexpresión observada en este modelo de CCRAC podría correlacionar, de ese modo, con el aumento de Thf observado. Sin embargo, y considerando que este marcador también es expresado por células tumorales en cáncer colorrectal, la correlación no puede ser asociada solamente a esta población. Siguiendo con los marcadores de linfocitos T, *CD69* es un marcador de activación que se encuentra expresado en linfocitos T activados y su función es co-estimular la activación de otros linfocitos T. Asimismo, se encuentra expresado en linfocitos B, células NK, monocitos, neutrófilos y eosinófilos (Ziegler *et al.*, 1994). En ese sentido, la falta de alteración en la población de linfocitos T efectores y el aumento de células NK activadas podría correlacionar con el aumento de *CD69* descrito anteriormente.

Por último, los marcadores *Clec4n* y *Clec7a*, lectinas de tipo C, se correlacionan con la respuesta inmune innata, particularmente con células dendríticas y macrófagos. En ese sentido, se podría correlacionar la sobreexpresión de estos marcadores con el aumento de células dendríticas activadas, dado que no se observan alteraciones en la población de macrófagos.

Las diferencias encontradas en los genes analizados y los resultados obtenidos por ImmuCC, se pueden explicar porque el algoritmo utiliza una firma genética conformada por 511 genes y, a partir del nivel de expresión génica, aplica un algoritmo de deconvolución para la identificación y caracterización de 25 poblaciones inmunes (Chen *et al.*, 2017). En ese sentido, la variación estadísticamente significativa en la expresión diferencial de genes puede deberse a la variación en otras poblaciones celulares que expresen esos genes, pero para los cuales no forme parte de la firma en el algoritmo.

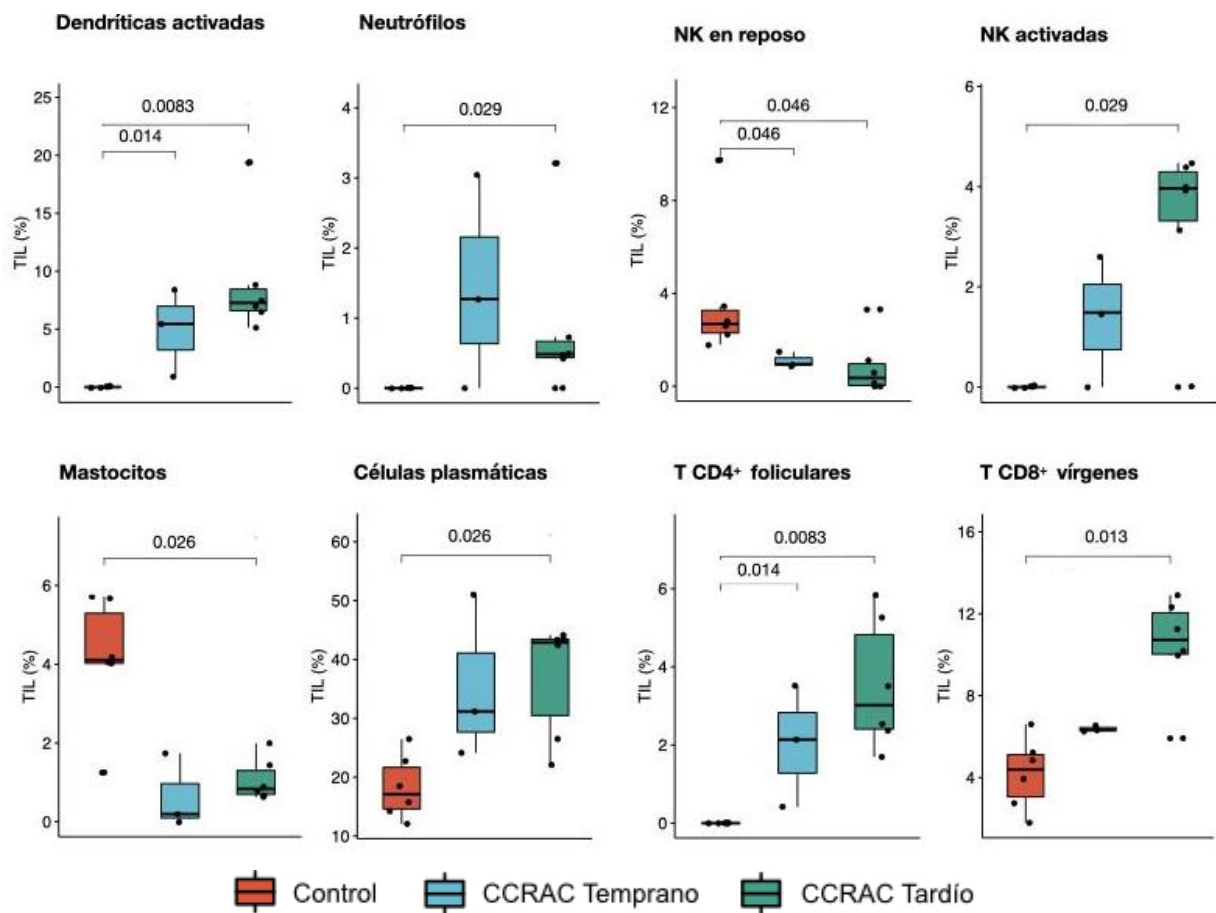


Figura 12: Caracterización del infiltrado inmune en el modelo de CCRAC Distal Temprano y Tardío por ImmuCC, graficando las distintas poblaciones inmunes caracterizadas que presentaron una variación significativa ($p_{adj} \leq 0,05$) en su proporción.

4.1.3 Comparación en la expresión de glicogenes en el modelo de inflamación intestinal por dextrano sulfato sódico (DSS) y en CCRAC (AOM-DSS)

En tercera instancia, y con el objetivo de encontrar mecanismos comunes a inflamación y neoplasia intestinal, se analizó la expresión diferencial en modelos experimentales de CCRAC (AOM-DSS), pero en este caso, comparando con un modelo de inflamación intestinal crónica por DSS, que representa a las EII, utilizando los datos de modelos experimentales murinos en el GSE57533 (Tabla I).

Para este experimento se diseñó un flujo de trabajo para procesar y analizar las muestras del experimento, el cual se presenta en la Figura 13.

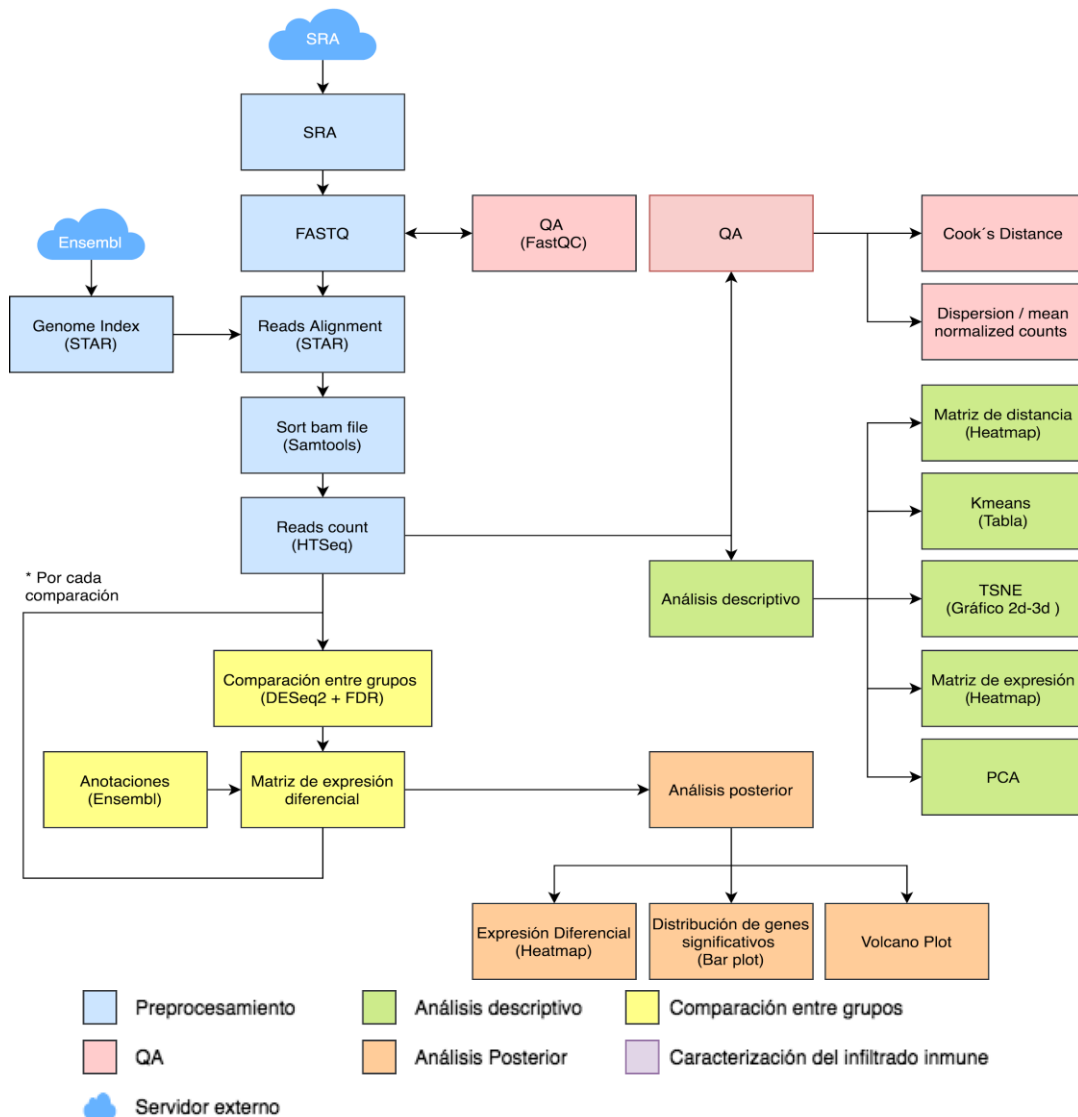


Figura 13: Flujo de trabajo utilizado para el procesamiento y análisis de GSE57533 (modelos experimentales de EII y CCRAC).

Luego, se analizó la expresión diferencial de todos los genes y de los genes de interés glicobiológico seleccionados en un modelo experimental de EII (DSS) y de CCRAC (AOM-DSS) (GSE57533). Los genes significativos ($p\text{-adj} \leq 0,05$) se filtraron según el valor de \log_2FC como se indica en la Tabla VII.

TABLA VII: Resultados de la comparación entre grupos de modelo de CCRAC (AOM-DSS) y EII (DSS)

Grupos	Filtro	Genes significativos ($p\text{-adj} \leq 0,05$)	Glicogenes significativos ($p\text{-adj} \leq 0,05$)
EII vs. Control	$ \log_2FC \geq 0$	2613	110
	$ \log_2FC \geq 1$	1401	57
	$ \log_2FC \geq 2$	679	22
CCRAC vs. Control	$ \log_2FC \geq 0$	3926	162
	$ \log_2FC \geq 1$	2113	91
	$ \log_2FC \geq 2$	1051	47
CCRAC vs. EII	$ \log_2FC \geq 0$	776	33
	$ \log_2FC \geq 1$	650	26
	$ \log_2FC \geq 2$	425	20

4.1.3.1 Alteraciones en el transcriptoma de colon en modelo experimental de CCRAC (AOM-DSS)

En el análisis de expresión diferencial de genes comparando tumor vs. control en animales tratados con AOM-DSS en la base de datos GSE57533, obtuvimos un total de 3926 genes significativamente desregulados ($p\text{-adj} \leq 0,05$), de los cuales 162 clasifican como glicogenes. Notablemente, el análisis de los genes más desregulados ($|\log_2FC| \geq 2$) define una firma glicobiológica de 41 genes aumentados y 6 disminuidos en su expresión (Figura 14A). Aplicando un filtro con criterio $|\log_2FC| \geq 1$, la desregulación se reduce a 2113 genes de los cuales, 91 son glicogenes (Figura 14B).

Al comparar estos resultados con los presentados en secciones anteriores, volvemos a observar, ratificando su relevancia en el desarrollo de neoplasia en el modelo de AOM-DSS, el

aumento de *Acer1*, *Clec4n*, *Clec4a2*, *Clec7a*, *Gpc1*, *Icam1*, *Lgals9*, *Notum*, *Ptn*, *Plaur* y *Srgn*, ya detectados en ambos GSE analizados previamente. En tanto, algunos genes como *Reg3b* (detectado en el modelo de AOM-DSS distal tardío, pero no en el primer GSE de AOM-DSS analizado) o *Lyz1* (detectado en el primer GSE de AOM-DSS, pero no en el modelo distal tardío) vuelven a aparecer desregulados en este tercer análisis.

El análisis detallado de la funcionalidad de los mismos muestra sobreexpresión de glicohidrolasas de peptidoglicanos bacterianos (*Lyz1*) y lectina antibacteriana *Reg3b*, clusters de diferenciación y receptores asociados a células inmunes (*CD44*), glicoproteínas involucradas en el reconocimiento, degradación de proteoglicanos y glicosaminoglicanos (*Notum*, *Ptn*), proteoglicanos (*Gpc1*), metabolismo (*Pfkfb3*, *Plaur*), glicoproteínas de función inmune (*Srgn*, *Icam1*), lectinas con reconocimiento de glucanos (*Clec7a*), mananos (*Clec4n*), hidrólisis de glicolípidos (*Acer1*). También se observa la subexpresión de enzimas involucradas en el metabolismo de ácido siálico (sialidasa *Neu3* y sialiltransferasa *St8Sia5*). En particular, la disminución en la expresión de *Klb* (*Klotho beta*) se observa en el análisis de los tres GSE. Esta es una potencial glicosiltransferasa que se encuentra marcadamente subexpresada.

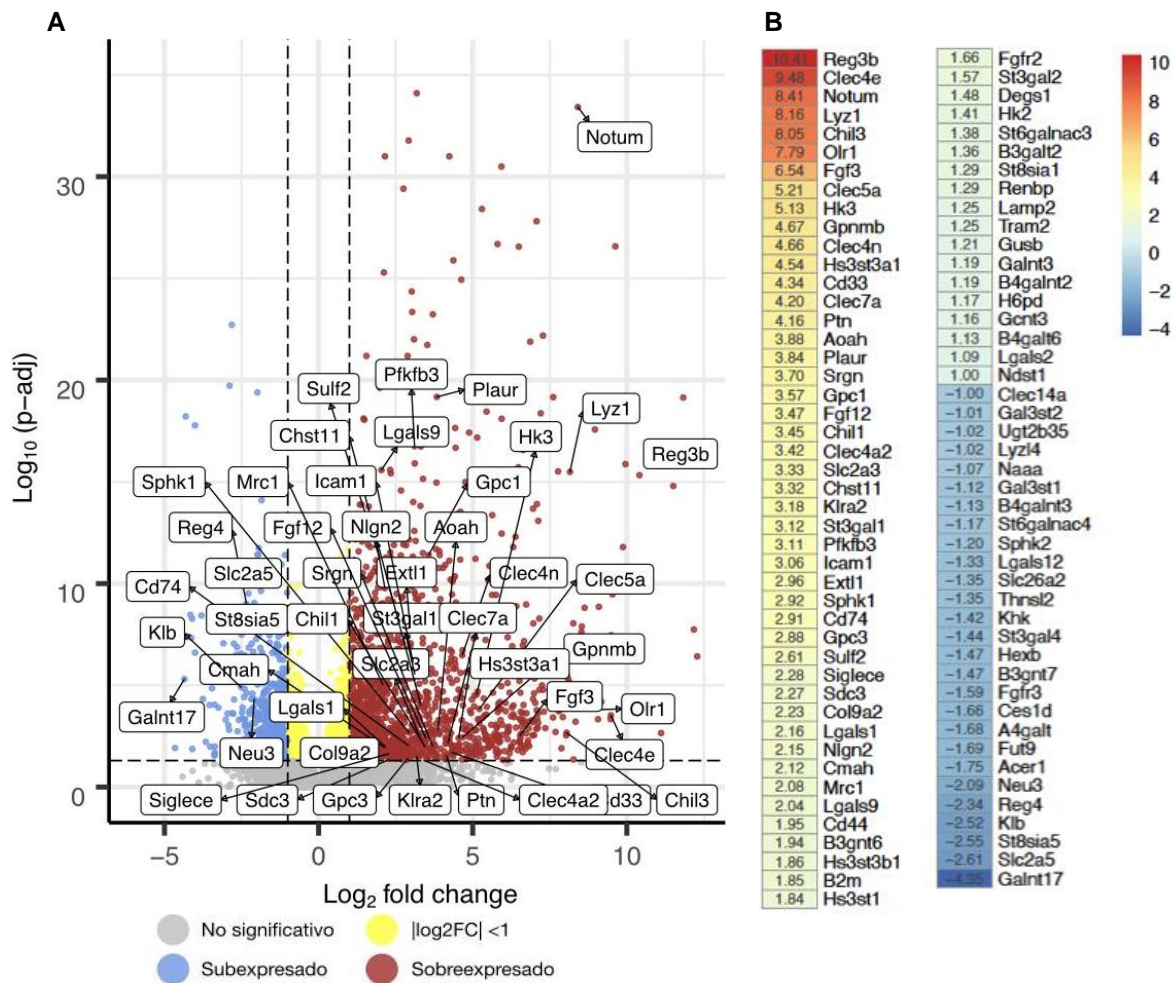


Figura 14: Genes desregulados en su expresión en cáncer colorrectal asociado a colitis, comparado con control. **A.** Volcano plot de glicogenes significativos detallando aquellos con $|\log_2FC| \geq 2$. **B.** Heatmap de glicogenes significativos con $|\log_2FC| \geq 1$.

4.1.3.2 Alteraciones en glicogenes en un modelo experimental de enfermedades inflamatorias intestinales por dextran sulfato sódico (DSS)

Con el objetivo de encontrar potenciales desregulaciones presentes en inflamación intestinal que sean comunes al desarrollo de neoplasia, analizamos datos de un modelo experimental de inflamación intestinal por dextrano sulfato sódico (DSS). En el análisis de expresión diferencial de glicogenes comparando tejido inflamado vs. control en animales tratados con DSS, obtuvimos un total de 2613 genes significativamente desregulados ($p\text{-adj} \leq 0,05$), de los cuales 110 clasifican como glicogenes. Notablemente, el análisis de los genes más desregulados ($|\log_2FC| \geq 2$) define una firma glicobiológica de 19 genes aumentados y 3

disminuidos en su expresión (Figura 15A). Aplicando un filtro con criterio $|\log_2FC| \geq 1$, la desregulación se reduce a 1401 genes de los cuales, 57 son glicogenes (Figura 15B).

El análisis detallado muestra glicogenes desregulados en común con los ya detallados en modelos de carcinogénesis asociada a colitis. En particular, se observó la sobreexpresión de lectinas tipo C (*Clec4e*), lectinas antibacterianas (*Reg3b*), sialiltransferasa *St3Gal1*, *Srgn*, *Icam1* y 2, *Plaur*, y subexpresión de *Acer1*, *Klb*, y de enzimas involucradas en el metabolismo de ácido siálico (sialidasa *Neu3* y sialiltransferasa *St8Sia5*). Aunque con menor desregulación que la observada en modelos de CCRAC, se encuentran también aumentadas en su expresión *Sulf2* y *Lgals9*.

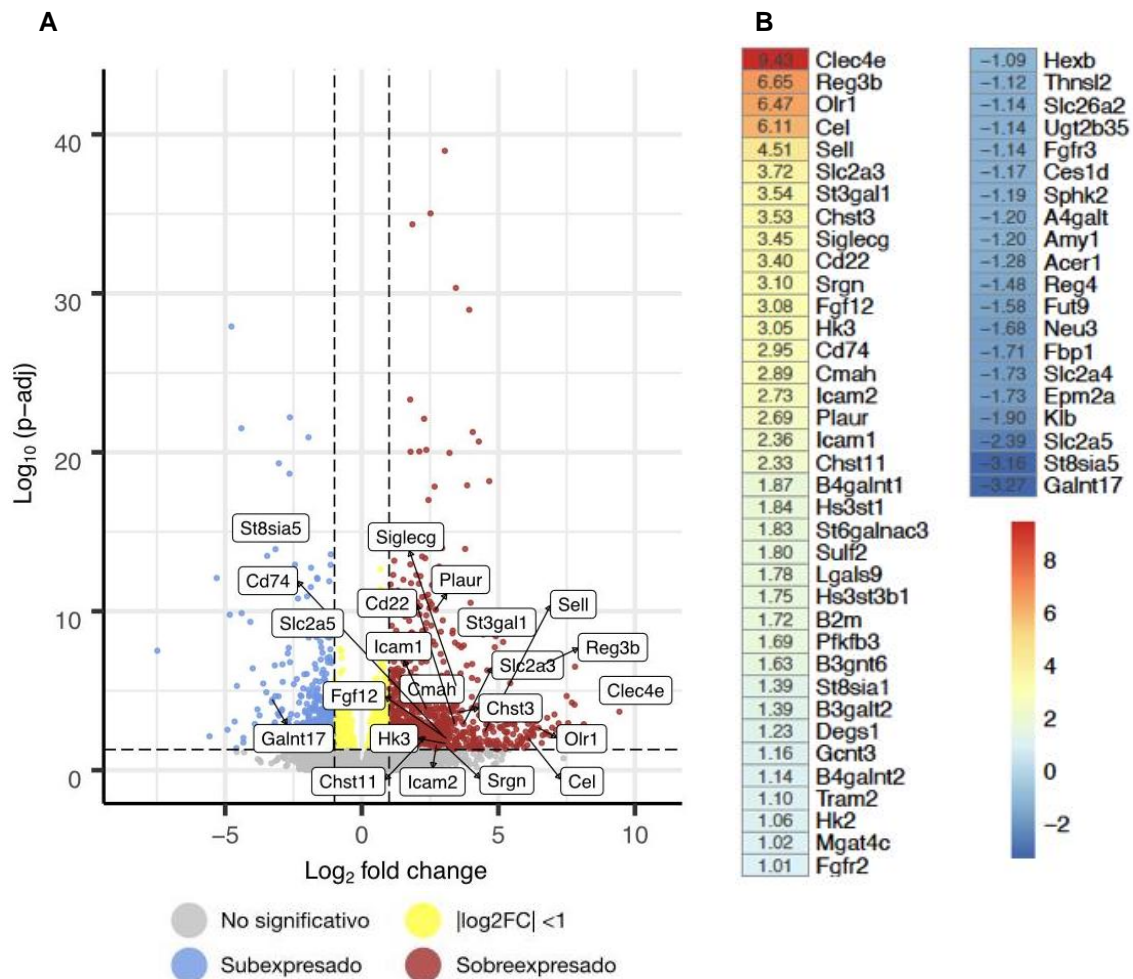


Figura 15: Genes desregulados en su expresión en modelo experimental de colitis por inducción con DSS, comparado con control. **A.** Volcano plot de glicogenes significativos detallando aquellos con $|\log_2FC| \geq 2$. **B.** Heatmap de glicogenes significativos con $|\log_2FC| \geq 1$.

Al evaluar los genes desregulados con una tasa de cambio $|\log_2FC| \geq 2$, entre el modelo de CCRAC comparado con su control y el modelo de EII comparado con su control dentro de este GSE en análisis (GSE 64658, Tabla I), notamos que 6 glicogenes aparecen desregulados solamente en colitis, 22 glicogenes se desregulan tanto en colitis como en carcinogénesis asociada a colitis, y 25 glicogenes se desregulan específicamente en CCRAC (Figura 16). Según este análisis parcial encontramos que *CD22*, *Cel*, *Chst3*, *Icam2*, *Sell* y *Siglecg*, parecerían estar sobreexpresados exclusivamente en colitis. Sin embargo, al comparar con los otros datasets de modelos de CCRAC analizados, cuatro de estos genes (*CD22*, *Icam2*, *Sell* y *Siglecg*) ya fueron identificados como desregulados en CCRAC (Sección 4.1.2).

En síntesis, hemos identificado 22 glicogenes que se desregulan en este modelo de enfermedad inflamatoria intestinal y que se demuestran desregulados al alcanzar el estado neoplásico. Los glicogenes más destacados son *Plaur*, glicoproteínas de función inmune (*Srgn*, *Icam1*), sulfatasas (*Sulf2*), galectinas (*Lgals9*) y lectinas antibacterianas (*Reg3b*). También se observa la subexpresión de la glicosiltransferasa *Galnt17*, Klotho-b (*Klb*), y la sialidasa *Neu3* y sialiltransferasa *St8Sia5*.

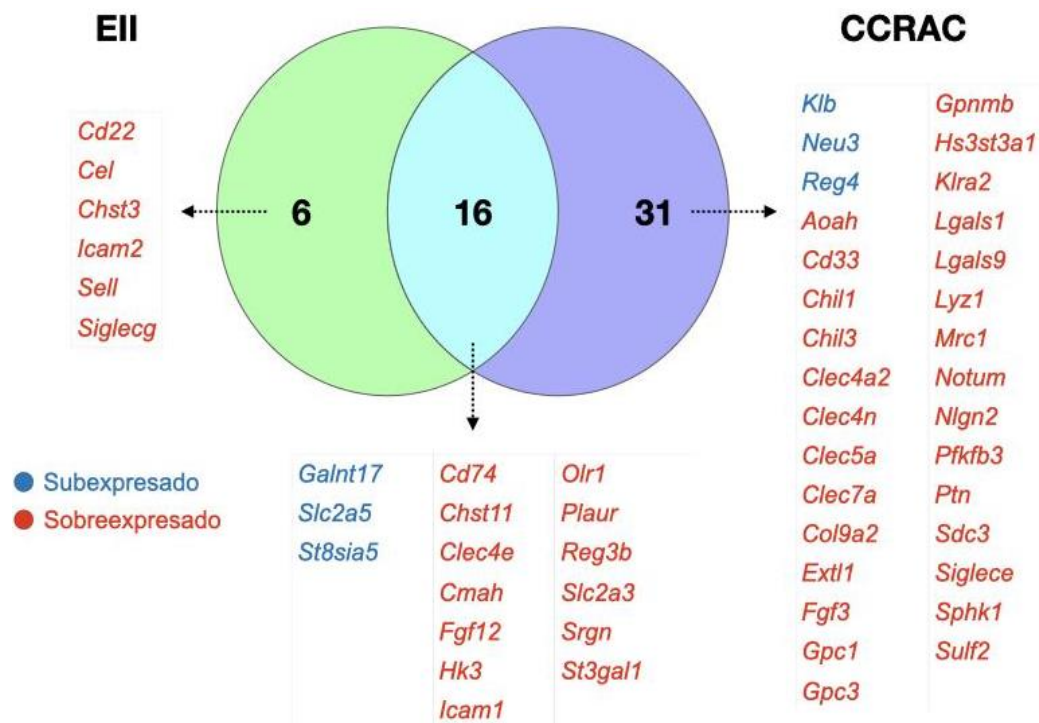


Figura 16: Diagrama de Venn de CCRAC vs. CCRAC Control \cup EII vs. EII Control, glicogenes significativos ($p\text{-adj} \leq 0,05$) y $|\log_2FC| \geq 2$. (*) Glicogenes significativos que en DSS o AOM-DSS están con $|\log_2FC| \geq 2$, mientras que en el otro $1 \leq |\log_2FC| \leq 2$.

4.1.3.3 Caracterización del infiltrado inmune

Se analizaron las poblaciones inmunes a partir de los datos de transcriptómica en los modelos de EII y CCRAC (GSE57533). Notablemente, no se encontraron variaciones significativas en ninguna de las poblaciones identificadas por InmuCC. Por lo tanto, no se pudo relacionar la expresión diferencial de los glicogenes asociados a estas poblaciones inmunes con la frecuencia de estas células. Estos resultados podrían deberse a la utilización de grupos con tamaño muestral muy chico (Control = 3, EII = 4, CCRAC = 3), falta de sensibilidad del algoritmo o la menor cantidad de poblaciones detectadas para RNA-seq en comparación con microarray (9 vs. 25).

4.1.4 Búsqueda de una firma robusta de glicogenes en modelos experimentales de CCRAC

En la Figura 17 puede observarse que, al comparar las tres bases de datos analizadas en esta tesis, las mismas se separan claramente según el estudio. En pos de lograr una firma robusta de glicogenes que pudiera mantenerse más allá de las condiciones del estudio (incluyendo las variaciones en la microbiota, las condiciones alimentarias y ambientales del bioterio), y a la vez sea independiente de la tecnología utilizada en cada experimento (microarray o RNA-seq), se evaluaron comparativamente los glicogenes desregulados para las tres transcriptómicas realizadas en modelos experimentales de cáncer colorrectal asociado a colitis (AOM-DSS, tomando solamente distal tardío en el caso del GSE64658 por criterio de similitud experimental). Los resultados obtenidos demuestran la existencia de una firma base conformada por 36 glicogenes desregulados significativamente ($p\text{-adj} \leq 0,05$ y $|\log_2FC| \geq 1$) (Figura 18).

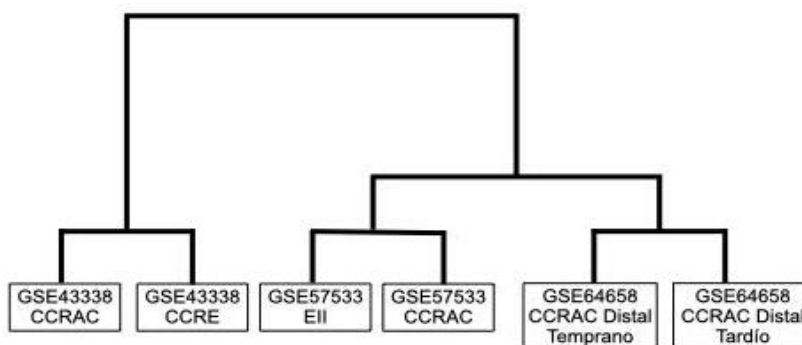


Figura 17: Dendrograma generado a partir de la matriz de expresión génica de las bases de datos analizadas, CCRAC GSE43338, GSE64658 y GSE57533, utilizando Hclust.

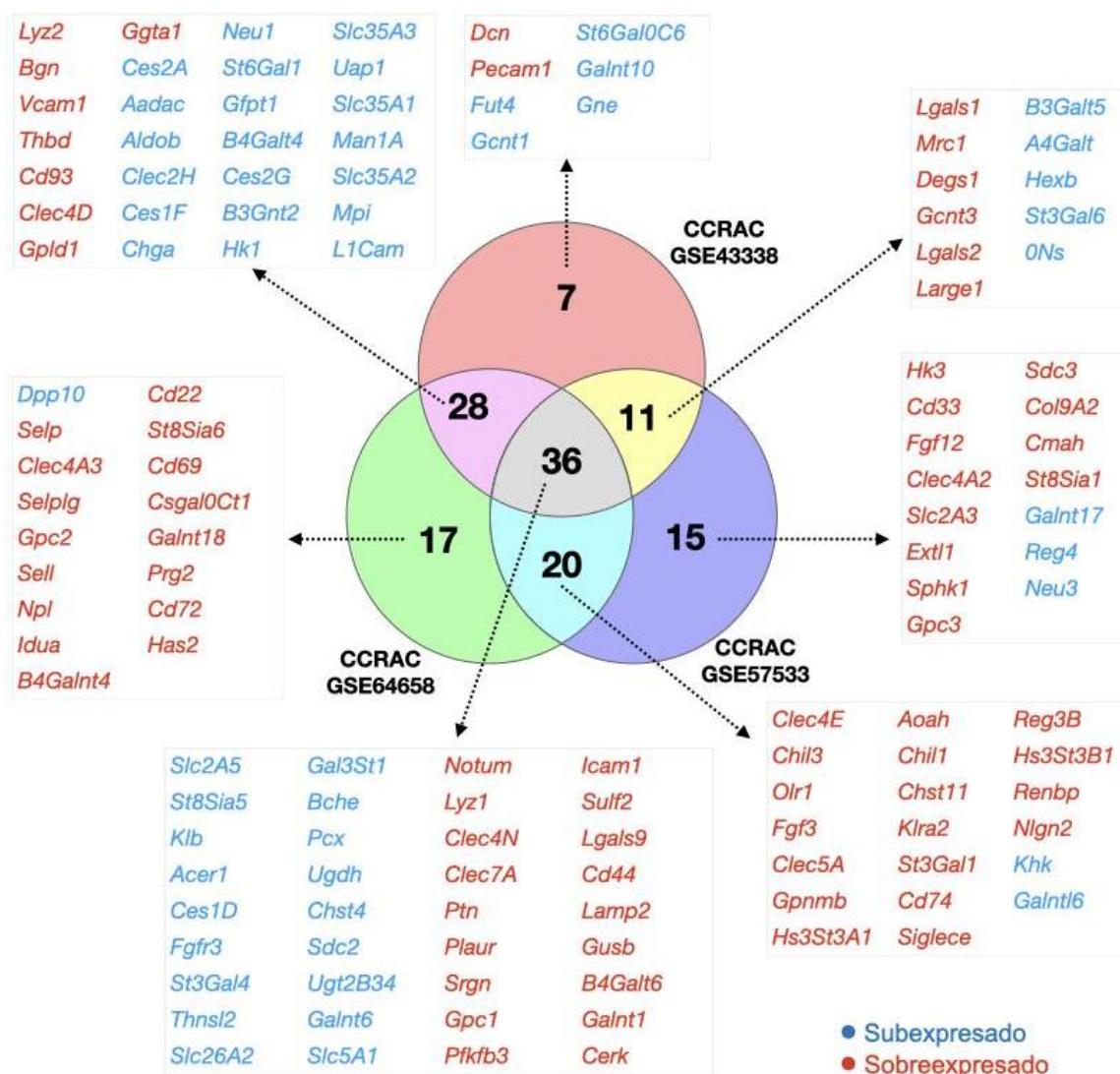


Figura 18: Diagrama de Venn con genes desregulados en forma comparativa para los tres modelos de CCRAC GSE43338, GSE64658 (distal tardío) y GSE57533. Para la comparación se tomaron solamente aquellos glicogenes significativos ($p\text{-adj} \leq 0,05$) y $|\log_2FC| \geq 1$.

Estableciendo un valor mínimo de $|\log_2FC| \geq 1.5$, obtenemos una firma donde la desregulación observada en los tres modelos de AOM-DSS está formada por 16 glicogenes: *Notum*, *Lyz1*, *Clec4n*, *Clec7a*, *Ptn*, *Plaur*, *Srgn*, *Pgc1*, *Pfkfb3*, *Icam1*, *Sulf2*, *Lgals9*, *Cd44*, *Ces1d*, *Acer1* y *Klb* (Figura 19). Cabe aclarar que *Plaur* tiene una desregulación $|\log_2FC| \geq 2$ en dos de los modelos experimentales de CCRAC (datasets GSE64658, GSE57533) pero con una desregulación menor en el GSE43338 ($1.5 \leq |\log_2FC| \leq 2$), y por eso no aparecía en el análisis de la sección 4.1.1. Lo mismo ocurre con *Lyz1*, que se encuentra sobreexpresado con

$|\log_2FC| \geq 2$ en los dataset GSE43338 y GSE57533, pero con una desregulación menor ($1.5 \leq |\log_2FC| \leq 2$) en el dataset GSE64658.

Para la determinación de esta firma, nos basamos en la consistencia de las desregulaciones y su significancia estadística, considerando que ciertas variaciones en el valor de *fold change* pueden ser producto de las condiciones experimentales.

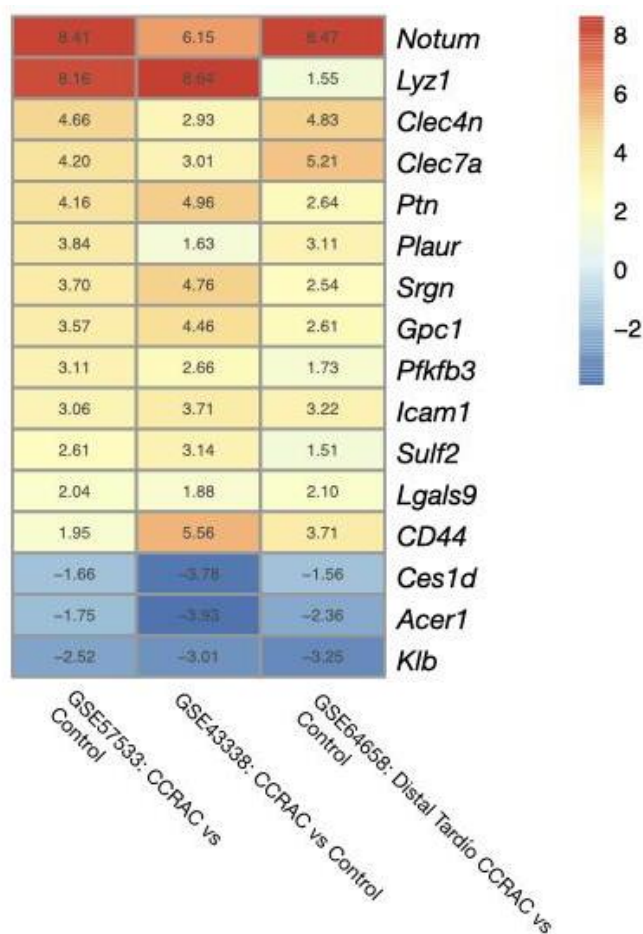


Figura 19: Firma de glicogenes significativamente desregulados en los tres GSE analizados (GSE57533, GSE43338 y GSE64658/distal tardío) para modelos experimentales de CCRAC, con $p\text{-adj} \leq 0,05$ y $|\log_2FC| \geq 1.5$.

Luego, se comparó esta firma con la desregulación de genes en el modelo esporádico de *Apc^{min/+}* ($|\log_2FC| \geq 1.5$). Notablemente, 15 de los 16 glicogenes de la firma obtenida para CCRAC también se encuentran desregulados en CCRE, indicando mecanismos comunes a ambos modelos (Figura 20A). En tanto, el gen *Plaur* pareciera desregularse específicamente en AOM-DSS, y podría estar asociado a mecanismos específicos de la neoplasia asociada a colitis.

Finalmente, comparando la firma obtenida para glicogenes en modelos experimentales de CCRAC, nuestro análisis en el modelo de inflamación intestinal por DSS muestra que siete de ellos se encuentran ya desregulados en el modelo de inflamación intestinal (*Srgn*, *Icam1*, *Sulf2*, *Pfkfb3*, *Lgals9*, *Plaur*, *Klb*) (Figura 20B). Además, 9 de los 15 genes en común entre los modelos de CCRE y CCRAC no están presentes en EII (*Acer1*, *Ces1D*, *Notum*, *Cd44*, *Gpc1*, *Clec4N*, *Clec7A*, *Ptn*, *Lyz1*), indicando mecanismos comunes a las neoplasias e independientes de la inflamación intestinal (Figura 20C).

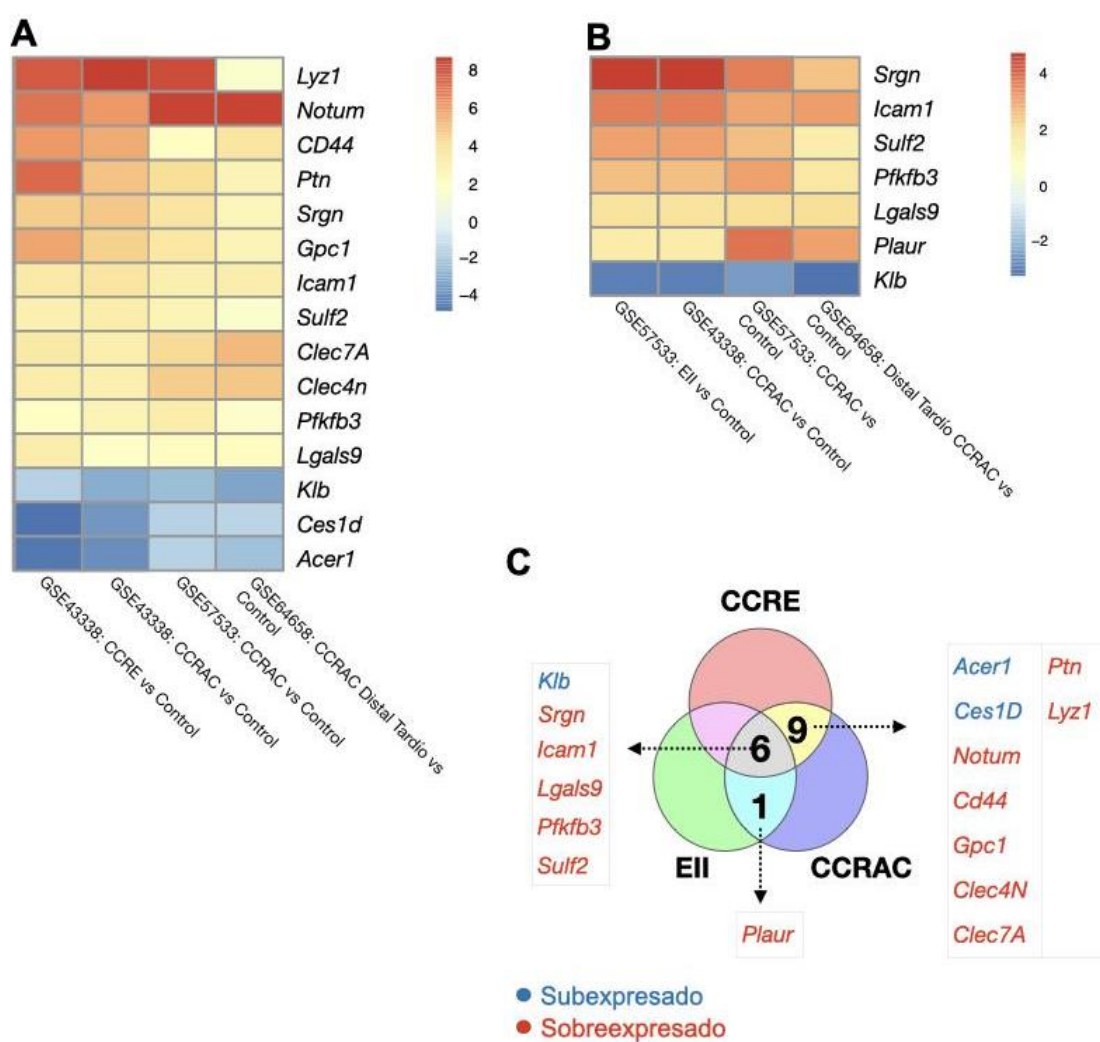


Figura 20: Firma de glicogenes significativos con $|\log_2FC| \geq 1.5$ entre los distintos modelos experimentales. **A.** Firma de glicogenes desregulados en común entre el modelo de CCRAC y el modelo de CCRE. **B.** Firma de glicogenes desregulados en común entre el modelo de EII y el modelo de CCRAC. **C.** Diagrama de Venn comparando firmas de glicogenes desregulados entre los modelos de EII, CCRAC y CCRE.

4.2 Análisis de la expresión de glicogenes en biopsias de pacientes con CCR

En esta última sección, procedimos a analizar muestras en pacientes obtenidas de la base de datos TCGA, siguiendo el siguiente flujo de trabajo (Figura 21).

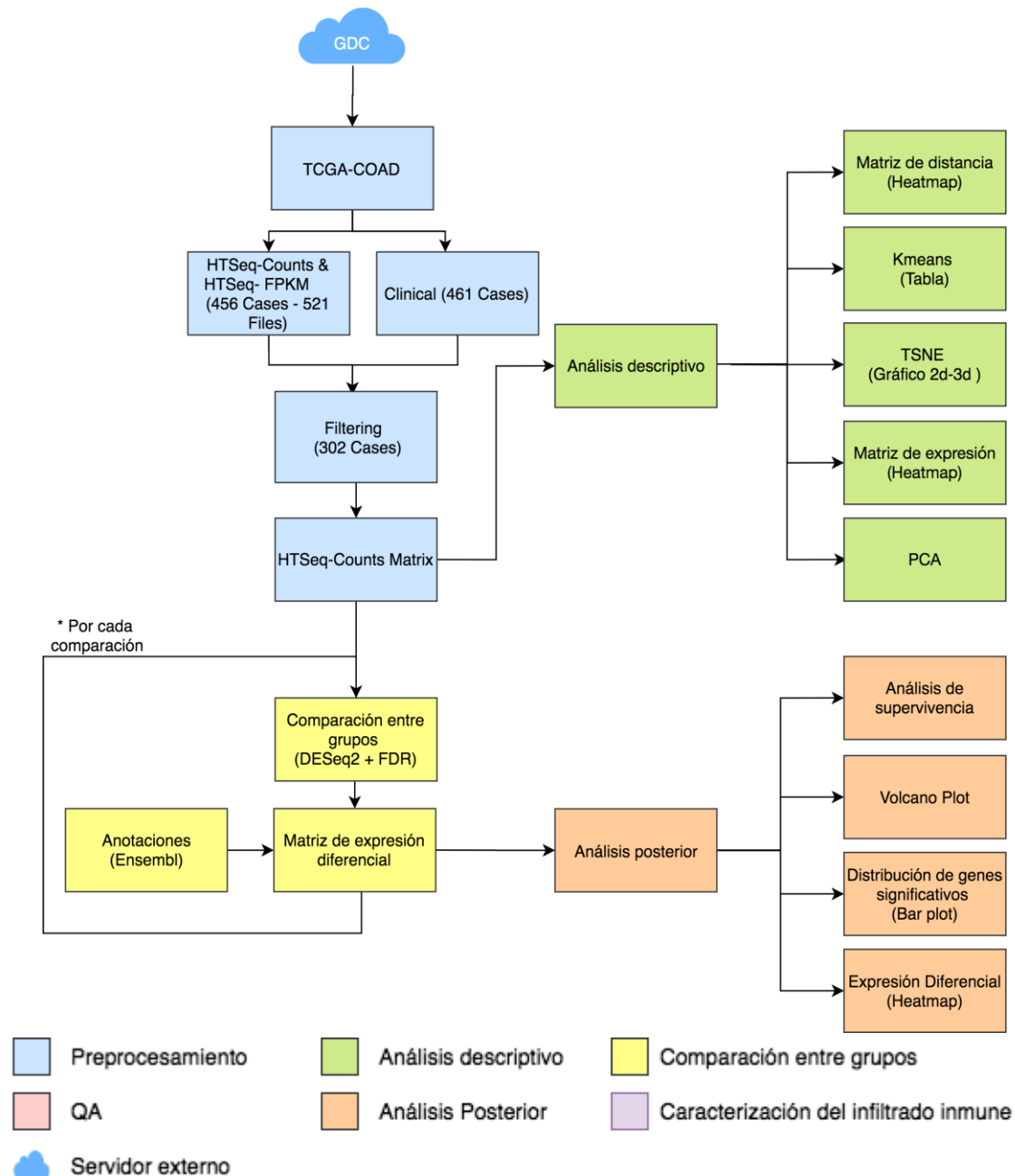


Figura 21: Flujo de trabajo utilizado para el análisis de expresión en muestras de pacientes (TCGA-COAD).

Asimismo, para la caracterización del infiltrado inmune se ejecutaron cuatro algoritmos distintos, cada uno capaz de distinguir distintos tipos de poblaciones inmunes, tal como se representa en la Figura 22.

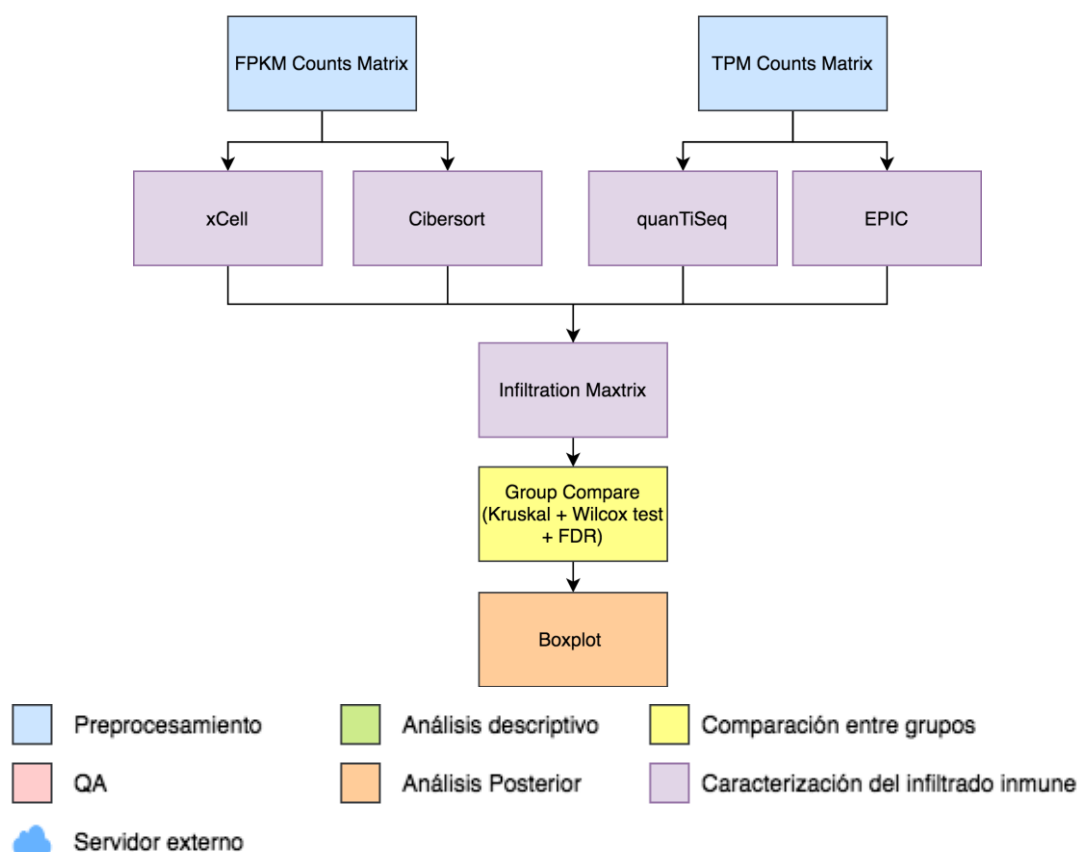


Figura 22: Flujo de trabajo utilizado para la caracterización del infiltrado inmune de TCGA-COAD.

4.2.1 Análisis descriptivo

En el análisis descriptivo de las variables clínicas, encontramos que la distribución de género (52,6% masculino, 47,4% femenino) y la edad de diagnóstico concuerdan con las estadísticas generales de pacientes de CCR (Markowitz y Bertagnolli, 2009), el cual presenta mayor incidencia en hombres que en mujeres (Bray *et al.*, 2018). Clasificando los diferentes grupos para el análisis de la expresión diferencial y la caracterización del infiltrado tumoral en las muestras de pacientes según tipo de muestra (control y tumor), estadio tumoral (I-IV), inestabilidad microsatelital (*Microsatellite Instability*, MSI) según bajo (L), alto (H) y estable (S), y deficiencia en el Sistema de Reparación de Emparejamiento de ADN (*deficient Mismatch*

Repair, dMMR) se obtiene una distribución como la que se observa en la Figura 23A-D. También se muestra la clasificación en base a las regiones de colon donde se tomaron las biopsias (Figura 23E), y por tratamiento (Figura 23 F). Cabe aclarar que las historias clínicas de pacientes no distinguen entre cáncer colorrectal esporádico o asociado a colitis.

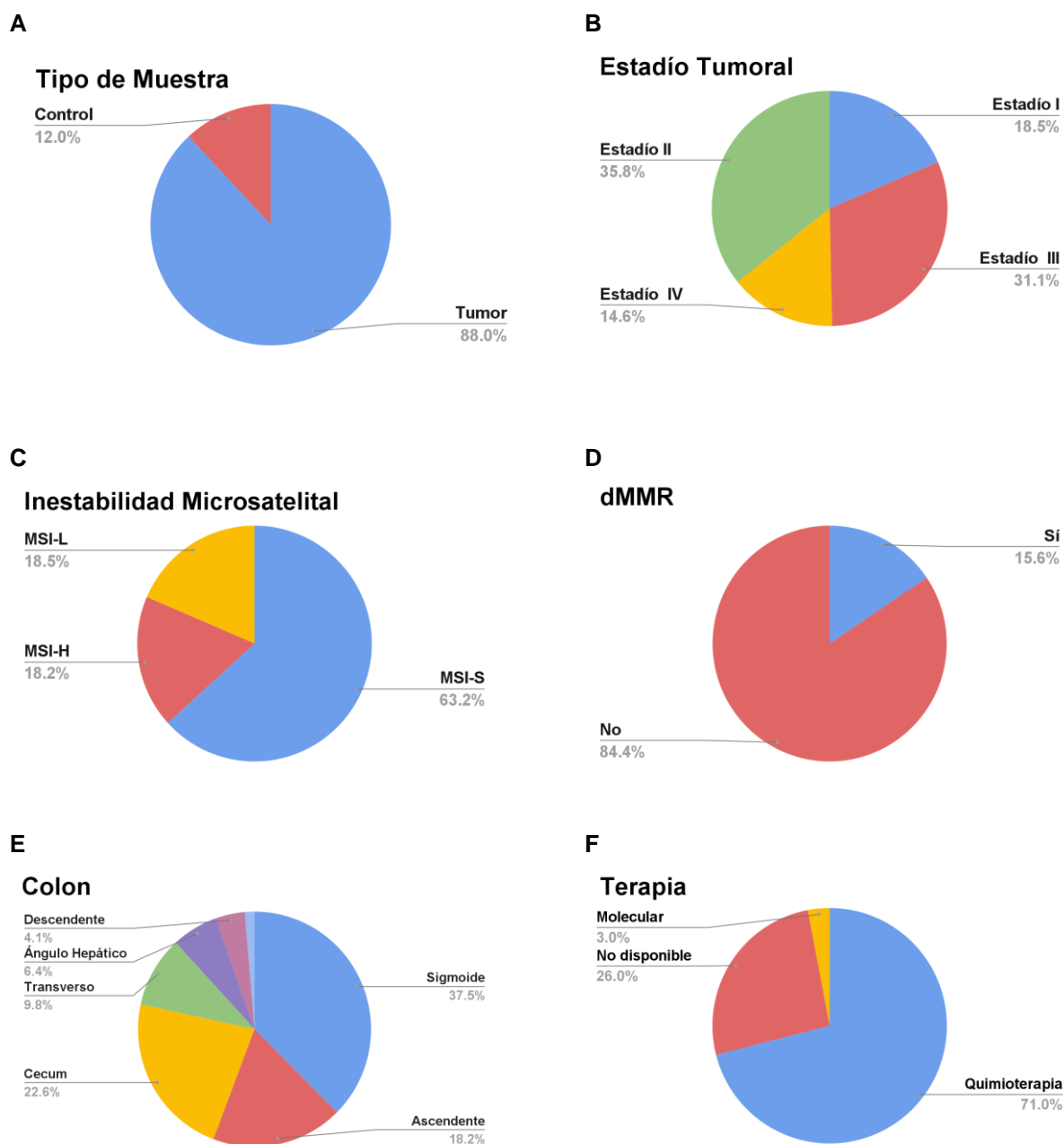


Figura 23: Análisis descriptivo de variables de interés en base de datos TCGA. Clasificación y distribución por: **A.** tipo de muestra; **B.** estadío tumoral; **C** inestabilidad microsatelital en muestras tumorales (MSI); **D.** deficiencia en el sistema de reparación de apareamiento de ADN (dMMR); **E.** Sección del colon donde se tomaron las biopsias; **F.** Terapias aplicadas.

Se analizó la expresión diferencial de todos los genes y de los glicogenes de interés seleccionados entre los distintos grupos de cada variable en los pacientes con cáncer colorrectal. Los genes significativos ($p\text{-adj} \leq 0,05$) se filtraron según el valor de \log_2FC (Tabla VIII).

TABLA VIII: Resultados de las comparaciones por cada variable.

Variable	Grupos	Filtro	Genes significativos ($p\text{-adj} \leq 0,05$)	Glicogenes significativos ($p\text{-adj} \leq 0,05$)
Tipo de Muestra	Tumor Primario vs. Control	$ \log_2FC \geq 1$	12887	182
		$ \log_2FC \geq 2$	4522	75
Estadio	I vs. Control	$ \log_2FC \geq 1$	12974	201
		$ \log_2FC \geq 2$	4906	82
	II vs. Control	$ \log_2FC \geq 1$	12061	185
		$ \log_2FC \geq 2$	4402	80
	III vs. Control	$ \log_2FC \geq 1$	12228	183
		$ \log_2FC \geq 2$	439	75
	IV vs. Control	$ \log_2FC \geq 1$	11839	208
		$ \log_2FC \geq 2$	4744	83
Inestabilidad microsatelital (MSI)	MSI Estable vs. Control	$ \log_2FC \geq 1$	12676	196
		$ \log_2FC \geq 2$	4496	78
	MSI Baja vs. Control	$ \log_2FC \geq 1$	11804	188
		$ \log_2FC \geq 2$	4485	83
	MSI Alta vs. Control	$ \log_2FC \geq 1$	1301	194
		$ \log_2FC \geq 2$	5352	85
Deficiencia en el sistema de reparación de emparejamiento de ADN (dMMR)	dMMR NO vs. Control	$ \log_2FC \geq 1$	13516	187
		$ \log_2FC \geq 2$	4851	75
	dMMR SI vs. Control	$ \log_2FC \geq 1$	9689	173
		$ \log_2FC \geq 2$	3499	75
	dMMR SI vs. dMMR NO	$ \log_2FC \geq 1$	1755	9
		$ \log_2FC \geq 2$	392	1

4.2.2 Caracterización del infiltrado inmune

Las células inmunes que componen el llamado infiltrado inmune pueden tener un impacto crucial en la progresión tumoral, la respuesta a la terapia y la probabilidad de recaída en el tiempo (Kamal *et al.*, 2021).

En la actualidad, los algoritmos para la caracterización del infiltrado tumoral se dividen básicamente en dos tipos: los basados en algoritmos de deconvolución y los basados en análisis de enriquecimiento. Los algoritmos de deconvolución (Cibersort, quanTIsec, EPIC) nos ofrecen como resultado el porcentaje en que se encuentra cada tipo de célula detectada por el algoritmo dentro del tumor sumando siempre un total de 100% por muestra, mientras que los basados en el análisis de enriquecimiento (xCell) resultan en un *score* que, basado en la firma génica, indica estrictamente puntuaciones de enriquecimiento y no pueden interpretarse como proporciones celulares.

Cibersort es un algoritmo de deconvolución basado en una firma de expresión formada por 547 genes (*Leukocyte signature Matrix*, LM22) que permite caracterizar 22 tipos de células inmunes humanas a partir de un transcriptoma de microarray o de RNA-Seq (Newman *et al.*, 2019). La caracterización se logra aplicando soporte de regresión vectorial (SVR), un enfoque de aprendizaje automático altamente robusto al ruido (Newman *et al.*, 2019). Adicionalmente, por cada muestra ofrece un P valor que nos permite descartar las muestras no significativas para el resto del análisis, un valor de correlación de Pearson para medir la concordancia entre las proporciones de tipos de células conocidas y las predichas, y un valor de error cuadrado medio (RMSE) para evaluar el sesgo de estimación que compara lo predicho con lo obtenido. Hay dos maneras para obtener los resultados: se pueden subir los datos de transcriptómica al sitio web (<https://cibersortx.stanford.edu>), o utilizar la librería en R.

En tanto, con el avance de la bioinformática y el interés generado por estos algoritmos, otras herramientas de análisis poblacional se fueron desarrollando. En este contexto, xCell (Aran *et al.*, 2017) es un algoritmo basado en el análisis de enriquecimiento y permite caracterizar 64 tipos de células inmunes y estromales (Aran *et al.*, 2017). Soporta tanto RNA-Seq como scRNA-Seq, y se basa en la utilización de distintas firmas genéticas aprendidas de múltiples fuentes de datos, lo que aumenta la capacidad de distinguir la señal del ruido y permite caracterizar la composición tumoral de manera más confiable (Aran *et al.*, 2017). Luego del análisis, a cada una de las 64 poblaciones, inmunes y estromales, se le asigna un score de

enriquecimiento en lugar de un porcentaje como en el caso de los métodos de deconvolución. En consecuencia, el uso principal de este método radica en la comparación del infiltrado entre muestras y no entre tipos de células. Además, éste método genera un valor de ImmunoScore por cada muestra, un sistema de puntuación estandarizado universal que se basa en la cuantificación de linfocitos T (CD3⁺) y linfocitos T CD8⁺ citotóxicos, tanto dentro del tumor como en los márgenes invasivos. El valor final se obtiene promediando los valores de densidad del infiltrado de cuatro percentiles y permite predecir el riesgo de recurrencia, definido como el tiempo transcurrido desde la cirugía hasta la recurrencia de la enfermedad (Pagès *et al.*, 2018).

En este contexto, comenzamos el análisis de las muestras de transcriptómica de la cohorte TCGA-COAD utilizando el método xCell basado en análisis de enriquecimiento. Elegimos este modelo porque es el que permite caracterizar el mayor número de poblaciones inmunes y estromales. Además, debido a la existencia previa de análisis de muestras de pacientes de CCR con el método de deconvolución Cibersort (Newman *et al.*, 2019), el estudio con xCell nos permitió analizar las consistencias y discrepancias entre ambos métodos.

En consecuencia, se filtraron las muestras con el criterio mencionado en la sección 3.2 y se caracterizó el infiltrado inmune tumoral mediante xCell para los siguientes grupos:

- tumor vs. control,
- estadio vs. control,
- inestabilidad microsatelital (MSI) vs. control,
- deficiencias en el sistema de reparación del ADN (dMMR) vs. control.

De estas comparaciones, nos enfocamos principalmente en el análisis de las diferencias entre tumor vs. control (Tabla IX) y la inestabilidad microsatelital (MSI-H) vs. MSS (Figura 24). Los análisis de los otros grupos se detallan en el Anexo D.

A nivel general, para las poblaciones inmunes analizadas que resultaron significativas, notamos que el valor medio de la mayoría de ellas se encuentra disminuidas en el tumor respecto al control, mientras que sólo cinco poblaciones se encuentran aumentadas, a saber: linfocitos pro B, linfocitos T reguladores (Tregs), linfocitos T cooperadores 1 y 2, y macrófagos tipo M2 (Tabla IX).

TABLA IX: Caracterización del infiltrado inmune tumoral en pacientes con cáncer colorrectal, comparando tumor vs. control, mostrando sólo aquellas poblaciones alteradas significativamente utilizando xCell.

Población		Tumor	
Linfocitos	B	B	↓
		B virgen	↓
		B de memoria	↓
		B de memoria conmutada de clase	↓
		Pro B	↑
	Plasmáticas	Plasmáticas	↓
	T CD4 ⁺	T CD4 ⁺ de memoria efectora	↓
	T CD8 ⁺	T CD8 ⁺	↓
		T CD8 ⁺ virgen	↓
		T CD8 ⁺ de memoria central	↓
	T reguladores	T reguladores	↑
	T cooperadores	T cooperadores 1	↑
		T cooperadores 2	↑
	T Gamma Delta	T Gamma Delta	↓
T Asesinos naturales	T Asesinos naturales	↓	
Leucocitos	Eosinófilos	Eosinófilos	↓
	Macrófagos	Macrófagos	↓
		Macrófagos M1	↓
		Macrófagos M2	↑
	Mastocitos	Mastocitos	↓
	Monocitos	Monocitos	↓
	Neutrófilos	Neutrófilos	↓
	Dendríticas	Dendríticas	↓
		Dendríticas activadas	↓
		Dendríticas convencionales	↓
Dendríticas inmaduras		↓	
Dendríticas plasmacitoides		↓	

Aumentadas en tumor vs. control ↑

Disminuidas en tumor vs. control ↓

Resulta esperable que las poblaciones inmunes que favorecen el escape y la progresión tumoral se encuentren aumentadas en el tumor, como son los Tregs y los macrófagos M2 antiinflamatorios (Lin *et al.*, 2019; Li *et al.*, 2020). Por otro lado, se observa que numerosas poblaciones inmunes que desarrollan respuestas antitumorales presentan un score disminuído en los tumores, como son los linfocitos T CD8⁺, los linfocitos T efectores, los macrófagos tipo M1 proinflamatorios y las células NK (Tabla IX).

A continuación, analizamos los resultados del infiltrado inmune mediante xCell en base al estatus de inestabilidad microsatelital (Control, MSS, MSI-L y MSI-H).

Al comparar el grupo MSI-H vs. MSS vemos un aumento significativo de enriquecimiento de genes de linfocitos T CD8⁺ y en particular, de linfocitos T CD8⁺ de memoria central y de memoria efectora, así como también linfocitos T CD4⁺ de memoria, linfocitos T cooperadores 1 y 2. Asimismo, se encuentran enriquecidos de manera significativa los genes correspondientes a linfocitos B, y en particular, los linfocitos B vírgenes, linfocitos B de memoria conmutada de clase y pro B. Además, las células dendríticas activadas, los basófilos, los macrófagos M1, mastocitos, monocitos y neutrófilos también se encuentran significativamente enriquecidos respecto a MSS. A su vez, vemos disminuidos los linfocitos NKT y linfocitos T gamma delta (Figura 24).

Estos resultados son congruentes con los hallazgos de otros estudios que han utilizado Cibersort para caracterizar el infiltrado inmune, demostrando que los pacientes con CCR y alta inestabilidad microsatelital (MSI-H) tienen asociada una importante carga mutacional (mutaciones por megabase) y, debido a su gran potencial de generar neoantígenos, con frecuencia son infiltrados por una alta proporción de linfocitos T citotóxicos (Narayanan *et al.*, 2019). Además, en otro estudio, analizando la expresión de genes e inmunohistoquímica, los tumores MSI-H exhiben densidades más altas de linfocitos T cooperadores tipo 1 (Th1), con alta producción de IFN- γ , linfocitos T de memoria efectora y linfocitos T proliferantes *in situ*, en comparación con los tumores con estabilidad microsatelital (MSS) (Catalano *et al.*, 2019).

En otro estudio de infiltrado inmune en pacientes con CCR utilizando el algoritmo de deconvolución Cibersort (Newman *et al.*, 2019), se observó que la proporción de linfocitos T fue significativamente mayor en el grupo MSI-H en comparación con MSS, lo cual resulta consistente con las proporciones de linfocitos T CD8⁺ y CD4⁺ de memoria, observadas en nuestro análisis por xCell. Sin embargo, en el estudio mencionado (Newman *et al.*, 2019; Angelova *et al.*, 2015) se observó también un aumento en la proporción de linfocitos T Gamma-

Delta, lo cual difiere de los resultados obtenidos por xCell en este trabajo. Asimismo, el estudio publicado por Angelova *et al.* (2015) describió una tendencia hacia una mayor proporción de linfocitos T CD8⁺ en pacientes MSI-H, aumento que en nuestro análisis por xCell fue significativo. Además, por Cibersort se observó un aumento de linfocitos T CD4⁺ de memoria activada (Newman *et al.*, 2019), y si bien xCell no detecta esta población inmune, en nuestro análisis con Cibersort (ver próxima sección) también se encontró significativamente aumentada. También se observó que en el grupo con MSI-H tenía una proporción más alta de macrófagos M1 y respecto del grupo MSS, lo que también coincide con nuestro análisis.

En resumen, en el análisis de las poblaciones inmunes de la base de datos de TCGA COAD mediante el algoritmo de enriquecimiento xCell se encontraron mayormente similitudes con el análisis previamente descrito con el algoritmo de deconvolución Cibersort.

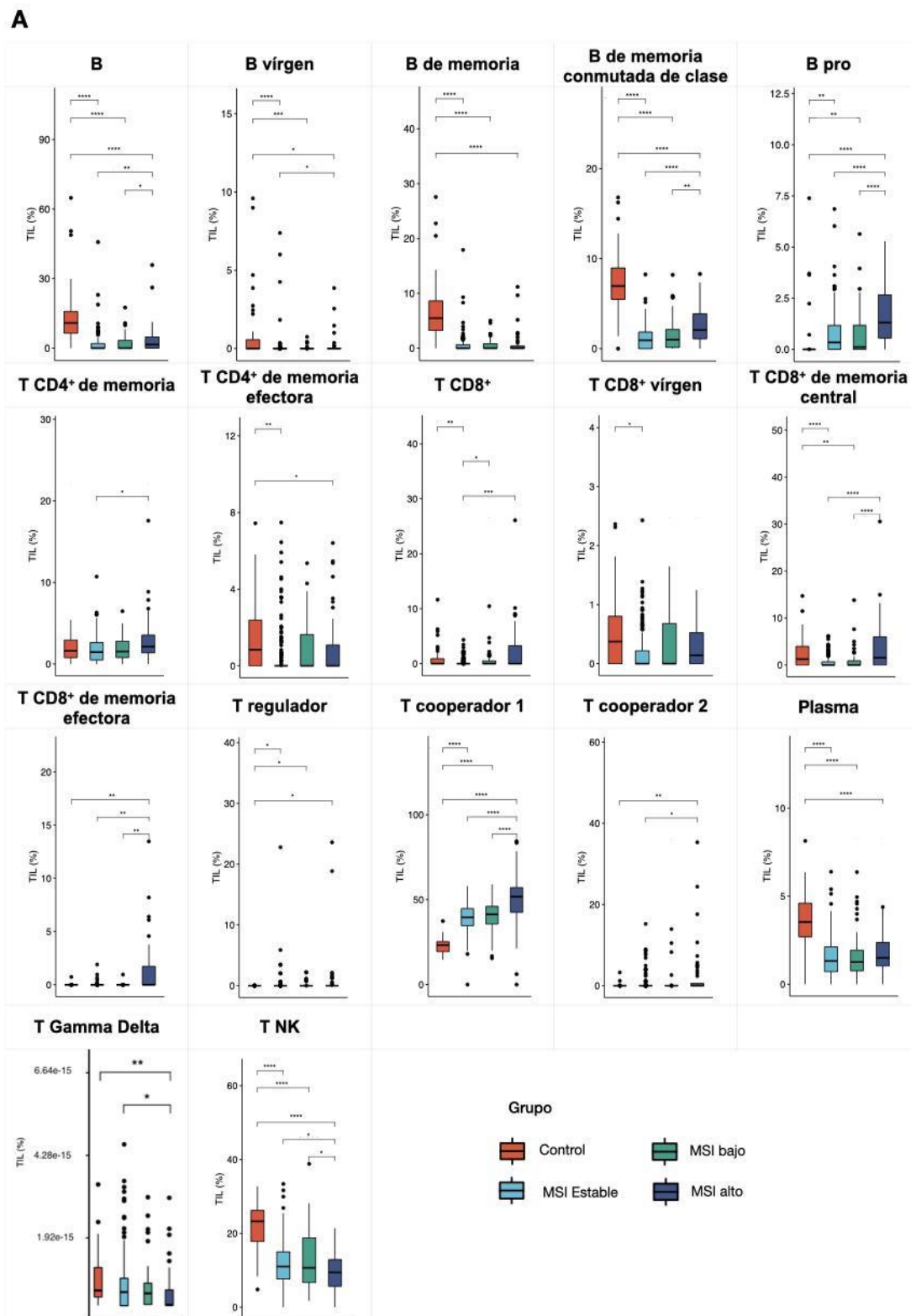


Figura 24A: Caracterización del infiltrado inmune en pacientes con cáncer colorrectal utilizando xCell en base al *status* de inestabilidad microsatelital con foco en poblaciones linfoides.

B

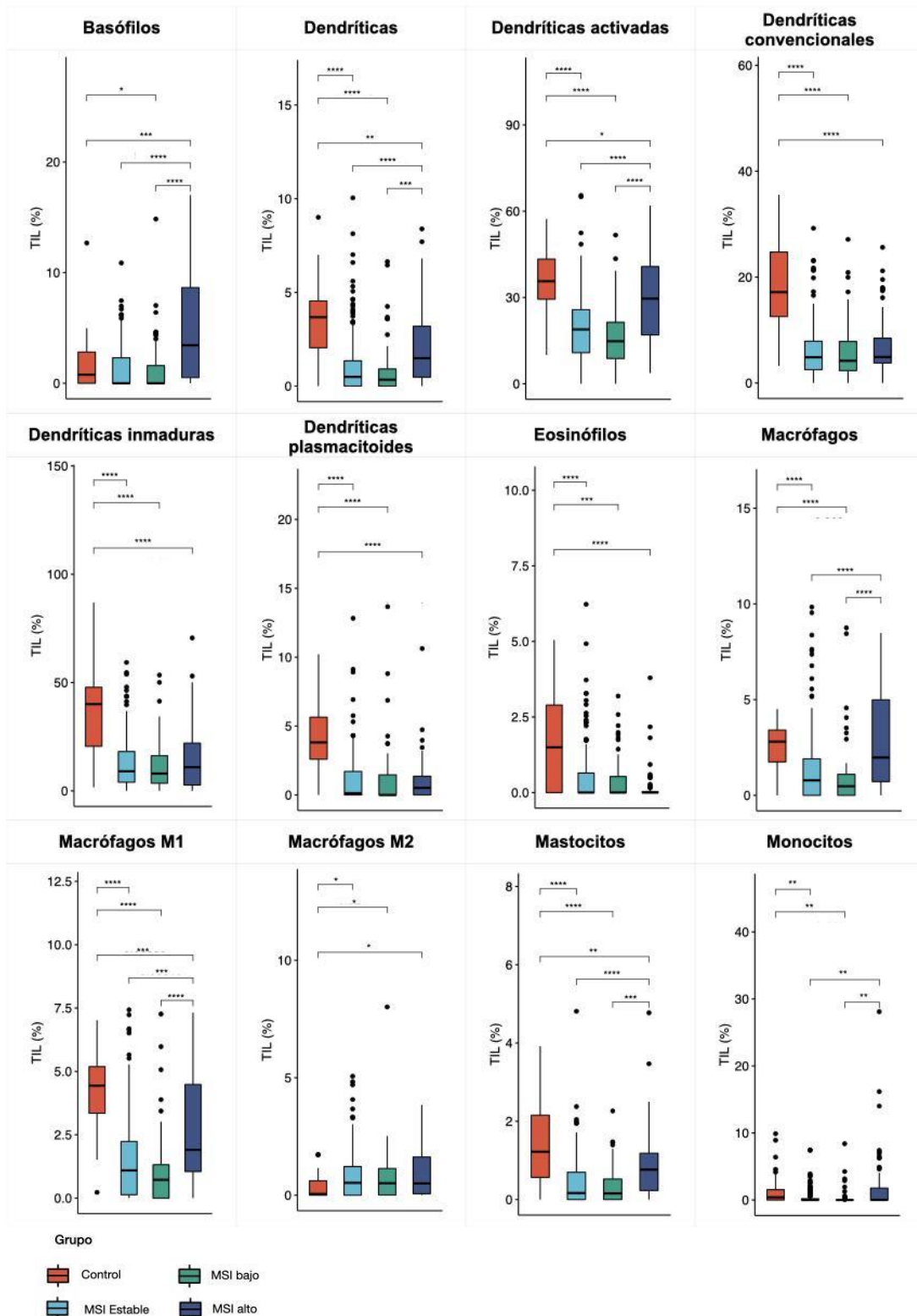


Figura 24B: Caracterización del infiltrado inmune en pacientes con cáncer colorrectal utilizando xCell en base al status de inestabilidad microsateles con foco en poblaciones mieloides.

4.2.3 Comparación del infiltrado inmune según distintos algoritmos

Con el objetivo de comparar los resultados de análisis poblacional del infiltrado inmune tumoral obtenidos a partir de la misma cohorte de pacientes, decidimos caracterizar este infiltrado inmune comparando muestras tumor vs. control con cuatro algoritmos de distinta naturaleza: por un lado, xCell (Aran *et al.*, 2017), un método de enriquecimiento de genes y por otro, los métodos de deconvolución como Cibersort (Newman *et al.*, 2019), descripto anteriormente, quanTISec (Finotello *et al.*, 2019) y EPIC (Racle *et al.*, 2017) (Tabla X).

Con respecto a estos otros dos métodos de deconvolución de análisis poblacional, quanTISec utiliza regresión de cuadrados mínimos restringidos (*constrained least squares regression*), y permite estimar las proporciones de 10 poblaciones del sistema inmune a partir de una firma de expresión genética formada por 170 genes (TIL10) para el análisis de microarray, y otra de 153 genes para el análisis de RNA-Seq (Wang *et al.*, 2009). Ésta difiere de la firma 547 genes *Leukocyte signature Matrix* (LM22) utilizada por Cibersort para caracterizar 22 tipos de células inmunes humanas. QuanTISec es el único método que ofrece un *pipeline* completo para obtener los resultados partiendo desde los datos crudos de secuenciación o partiendo desde una matriz de expresión.

Por último, EPIC es otro algoritmo de deconvolución para RNA-seq que permite estimar las proporciones de siete tipos de células del sistema inmune a partir de una firma de 98 genes. A cada una de estas siete poblaciones se le asigna un porcentaje y el restante se le asigna a una categoría “otros” que asume que pueden ser células tumorales y/o no tumorales. Asimismo, este método permite evaluar el infiltrado inmune en sangre circulante, donde puede detectar hasta seis tipos de células inmunes a partir de una firma de 65 genes. Para obtener los resultados, se pueden subir los datos de transcriptómica a <http://epic.gfellerlab.org> o bajar la librería y correrlo en R.

Se comparó entonces el infiltrado inmune de las muestras tumorales de CCR para nuestra cohorte de TCGA-COAD comparando tumor vs. control con los cuatro métodos de análisis antes mencionados: Cibersort, xCell, quanTISec y EPIC (Tabla X).

TABLA X: Comparación de la caracterización del infiltrado inmune tumoral en pacientes con CCR entre los cuatro distintos algoritmos utilizados. Las flechas indican la relación del tumor vs. control.

Poblaciones del sistema inmune			Algoritmo de enriquecimiento de genes	Algoritmos de deconvolución		
			xCell	Cibersort	quanTIsec	EPIC
Linfocitos	B	B	↓		↓	↓
		B virgen	↓			
		B de memoria	↓	↓		
		B de memoria conmutada de clase	↓			
		B pro	↑			
	Plasmática	Plasmática	↓	↓		
	T CD4 ⁺	T CD4 ⁺			↓	↓
		T CD4 ⁺ virgen				
		T CD4 ⁺ de memoria				
		T CD4 ⁺ de memoria descansando				
		T CD4 ⁺ de memoria activada		↑		
		T CD4 ⁺ de memoria central				
		T CD4 ⁺ de memoria efectora	↓			
	T CD8 ⁺	T CD8 ⁺	↓	↓		
		T CD8 ⁺ virgen	↓			
		T CD8 ⁺ de memoria central	↓			
		T CD8 ⁺ de memoria efectora				
	T regulador	T regulador	↑	↓	↑	
		T regulador ayudante folicular				
	T cooperador	T cooperador 1	↑			
		T cooperador 2	↑			
	T Gamma Delta	T Gamma Delta	↓	↓		
	Asesinos naturales (NK)	T Asesinos naturales	↓			
Asesinos naturales				↓		
Asesino natural descansando			↑			
Asesino natural activado			↓			

Poblaciones del sistema inmune			Algoritmo de enriquecimiento de genes	Algoritmos de deconvolución		
			xCell	Cibersort	quanTIsec	EPIC
Leucocitos	Basófilos	Basófilos				
	Dendríticas	Dendríticas	↓			
		Dendríticas descansando		↓		
		Dendríticas activadas	↓			
		Dendríticas convencionales	↓			
		Dendríticas inmaduras	↓			
		Dendríticas plasmacitoides	↓			
	Eosinófilos	Eosinófilos	↓	↓		
	Macrófagos	Macrófagos	↓			↓
		Macrófagos M0		↑		
		Macrófagos M1	↓	↑	↓	
		Macrófagos M2	↑	↓	↓	
	Mastocitos	Mastocitos	↓			
		Mastocitos descansando		↓		
		Mastocitos activados		↑		
	Monocitos	Monocitos	↓	↓		
Neutrófilos	Neutrófilos	↓	↑	↓		

	Detectable y significativo
	Detectable y no significativo
	Sin capacidad de detección

De manera general, podemos observar la mayor cantidad de poblaciones analizadas en XCell respecto de los otros métodos. Para una población inmune que puede ser detectada por dos o más algoritmos de análisis poblacional, los resultados reflejan consistencias, salvo en el caso de linfocitos B virgen (*naive*), donde xCell los detecta con diferencias estadísticamente significativas, pero Cibersort no. En tanto, los linfocitos T CD4⁺ son detectados por xCell, pero no significativamente modificados, mientras que quanTIsec y EPIC sí. Lo mismo ocurre con los linfocitos T Reguladores, que se encuentran significativamente aumentados para XCell y quanTIsec, mientras que para Cibersort no. Por último, los linfocitos T CD8⁺, son detectados por los 4 algoritmos, pero xCell y Cibersort los detectan con diferencias estadísticamente significativas, mientras que quanTIsec y EPIC no. También existen variaciones en las células NK, células dendríticas, células dendríticas activadas y monocitos.

4.2.4 Comparación del infiltrado inmune en un modelo experimental de CCRAC y pacientes de CCR

Uno de los objetivos de esta tesis era la caracterización del infiltrado inmune tumoral comparando los modelos experimentales murinos con los pacientes de CCR. Su correcta caracterización en modelos experimentales ayuda a comprender los mecanismos involucrados en la respuesta antitumoral, y la comparabilidad con la patología humana. En pacientes, es crucial para revelar los mecanismos de escape tumoral y determinar el pronóstico y su posible respuesta a inmunoterapias (Fernández *et al.*, 2020). En ese sentido, como se mencionó anteriormente, en los últimos años se han desarrollado numerosos algoritmos para el estudio de infiltrado tumoral en muestras de pacientes, a diferencia de los modelos experimentales donde sólo está disponible ImmuCC.

En consecuencia, se compararon las poblaciones estadísticamente significativas del infiltrado inmune tumoral de ratón en base al modelo experimental de CCRAC distal tardío (GSE64658) utilizando ImmuCC (sección 4.1.2.3, evaluando tumor vs. control) y las mismas poblaciones, pero en pacientes de CCR (TCGA-COAD) mediante xCell (sección 4.2.2). Se observaron numerosas diferencias entre las poblaciones significativamente alteradas, siendo los mastocitos los únicos significativamente disminuídos y detectados en ambos análisis (Tabla XI).

TABLA XI: Comparación de resultados en la caracterización del infiltrado inmune entre el modelo experimental de AOM-DSS distal tardío (GSE64658) y pacientes con CCR (TCGA-COAD), ambos evaluando Tumor vs. Control.

Población	Modelo Experimental	Pacientes
Células dendríticas activadas	↑	↓
Neutrófilos	↑	↓
NK activadas	↑	Sin capacidad de detección
NK en reposo	↓	Sin capacidad de detección
Mastocitos	↓	↓
Células plasmáticas	↑	↓
T CD4 ⁺ foliculares	↑	Sin capacidad de detección
T CD8 ⁺ virgen	↑	↓

4.2.5 Alteraciones en el glicotranscriptoma de CCR en pacientes

Al analizar los glicogenes con expresión diferencial y aplicando un filtro de $|\log_2FC| \geq 2$ entre los distintos grupos (tumor, clasificado por estadio tumoral, inestabilidad microsatelital o deficiencia en el sistema de reparación de errores de apareamiento de ADN) vs control en pacientes con cáncer colorrectal, obtenemos una firma de 45 glicogenes que se mantienen desde el comienzo de la neoplasia (Figura 25).

El análisis detallado de su funcionalidad muestra sobreexpresión de glicosiltransferasas (**B4GALNT4*, *FUT1*, *LARGE2*), lectinas tipo C (**CLEC5A*, **OLRI*, *ACAN*), sulfotransferasas (**CHST4*, *HS6ST2*), glicosilhidrolasas (**LYZL4*, *PGGHG*), factor de crecimiento (*FGF20*), carboxilesterasa (**CEL*) y degradación de proteoglicanos (**NOTUM*).

También se observa una desregulación negativa en glicosiltransferasas (*B3GALT1*, **B3GALT5*, **B3GNT7*, **B4GALNT2*, **GCNT2*, *PYGM*, **ST6GALNAC6*, *ST8SIA3*, *UGT1A10*, *UGT1A5*, *UGT2A3*), lectinas tipo I (**CD22*, *NCAMI*), lectinas tipo C (**CLEC10A*, *CLEC4G*, *ATRNL1*), glicosilhidrolasas (*NEU4*, *HPSE2*), hidrolasas (**NAAA*), transporte, síntesis y metabolismo de azúcares (**SLC26A2*, **SLC2A4*, **UGP2*, *G6PC*, **PGM5*, **CHGA*), carboxilesterasa (**BCHE*), sulfotransferasas (*CHST9*, *HS3ST6*), klotho-b (**KLB*) y degradación de glicosaminoglicanos (**PTN*). Cabe aclarar que, al comparar con los genes desregulados en modelos experimentales, los glicogenes indicados con asterisco (*) también aparecen diferencialmente expresados en alguno de los modelos experimentales murinos de CACRC o CCRE incluidos en esta tesis con valores de $|\log_2FC| \geq 2$. Teniendo en cuenta los resultados presentados anteriormente (Sección 4.1.4), donde el gen *Plaur* se encontraba en la firma común a los datasets de los modelos experimentales de CCRAC analizados, en pacientes también se observa significativamente aumentado, pero con $\log_2FC = 0,63$.

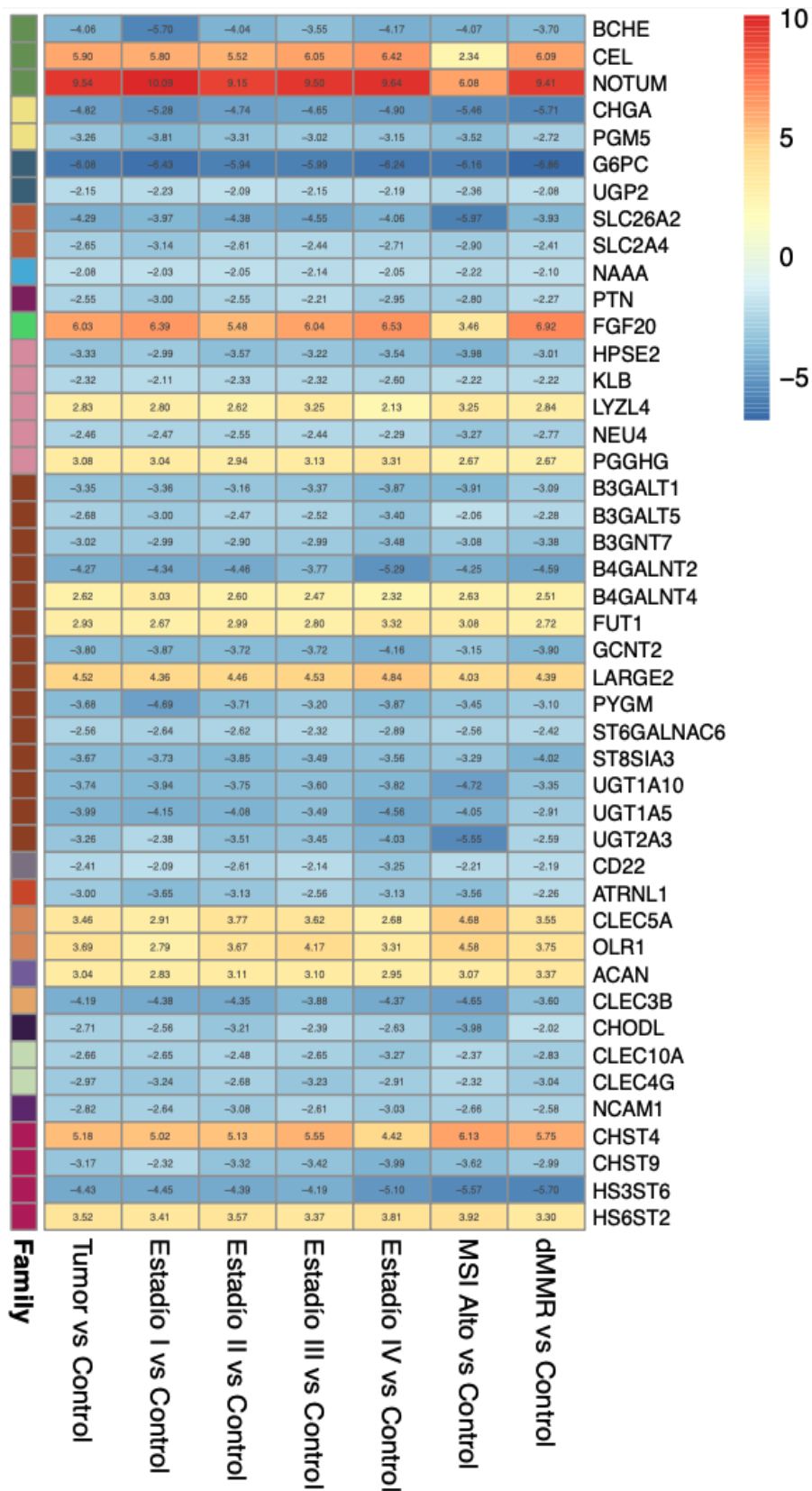


Figura 25: Heatmap de glicogenes significativos con $|\log_2FC| \geq 2$ en datos de pacientes con cáncer colorrectal, clasificados según estadio tumoral (I-IV), MSI alto y dMMR.

4.3 Comparación entre los modelos experimentales y pacientes con CCR

4.3.1 Firma de glicogenes desregulados compartidas entre modelos experimentales y pacientes con CCR

Considerando las limitaciones de los modelos experimentales en reflejar la patología humana (Bürtin *et al.*, 2020), nuestro objetivo final fue evaluar la validez de los modelos experimentales de AOM-DSS y *Apc^{min/+}* para estudiar la mecánica glicobiológica involucrada en el desarrollo neoplásico. En ese sentido, buscando desregulaciones génicas asociadas a procesos que se presenten independientemente del estadio, la inestabilidad microsatelital, o de la deficiencia en el sistema de reparación de emparejamiento de ADN de pacientes, se comparó la firma de 16 glicogenes obtenida para los modelos experimentales de CCRAC y la firma de 45 glicogenes obtenida de nuestra cohorte de pacientes con CCR (sección 4.2.5).

Los resultados de esta comparación nos muestran que hay una firma en común de 3 glicogenes: *Klb*, *Notum* y *Ptn* (Figura 26), donde *Klb* y *Notum* se desregulan negativamente tanto en los modelos experimentales de CCRAC y CCRE como en pacientes de CCR, mientras que el glicogen *Ptn* se encuentra sobreexpresado en los modelos experimentales, pero subexpresado en los pacientes.

Cabe destacar que *Clec4n* se encuentra incluido en la firma de los modelos experimentales, pero no tiene ortólogo en humanos, y que *Ces1d* también se encuentra incluido en la misma firma, pero su ortólogo en humanos es *Ces1*, el cual, sin embargo, no se encuentra en la lista de glicogenes diferencialmente expresados de pacientes con CCR.

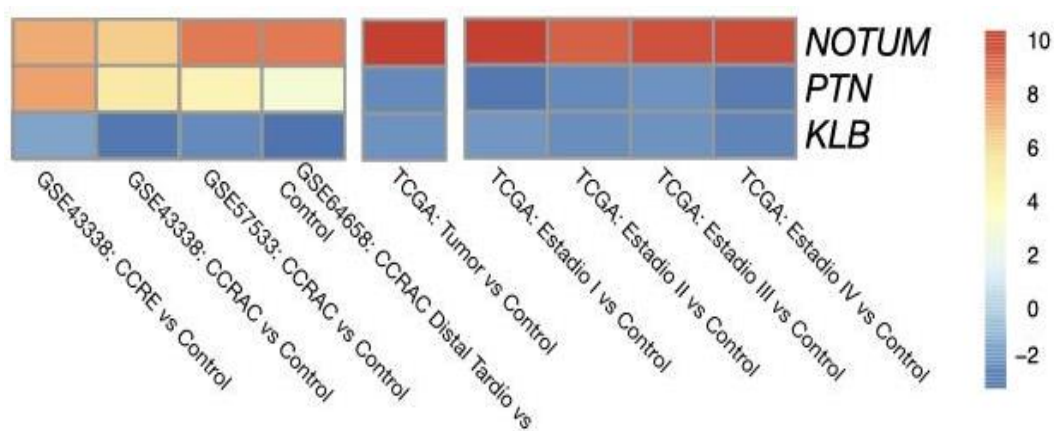


Figura 26: Firma de glicogenes desregulados tanto en modelos experimentales de CCRAC como en biopsias de pacientes con CCR.

4.3.2 *Curva de supervivencia para la firma de glicogenes en pacientes con CCR*

En base a la firma de glicogenes desregulados compartidos entre los modelos experimentales murinos y los pacientes con CCR de la cohorte de TCGA, que involucra a los genes de *NOTUM* (sobrexpresado), y *PTN* y *KLB* (subexpresados), realizamos un análisis de supervivencia. Se analizó cada gen de manera individual (Figura 27 A-C), de manera combinada filtrando por estadio I y II (Figura 27 D), comparando estadísticamente las curvas de referencia con las curvas con peor sobrevida que resultaron significativas. La transformación de los valores de expresión de estos tres glicogenes a su correspondiente variable categórica (baja y alta) se encuentran en el Anexo E, y el gráfico completo de todas las curvas de supervivencia, junto a los resultados estadísticos calculados por regresión de Cox, se encuentra en el Anexo F.

Si bien la utilización de un único gen como factor pronóstico de supervivencia global (OS) o supervivencia libre de enfermedad (DFS) no siempre resulta ideal, es interesante hacer el análisis de sobrevida para ver las diferencias en su expresión. Precisamente, en las curvas individuales de cada gen (Figura 27 A-C), en concordancia con la firma genética determinada, se observa peor pronóstico asociado a alta expresión de *NOTUM*, baja expresión de *KLB* y baja expresión de *PTN*. Asimismo, la curva combinada con alta expresión de *NOTUM* y baja expresión de *KLB* y *PTN* (Figura 27 D) presenta una sobrevida significativamente menor que la curva de referencia. Notablemente, en el análisis combinado de los resultados de sobrevida, también se observa peor pronóstico para pacientes con baja expresión de *KLB* y *NOTUM*, y alta expresión de *PTN*.

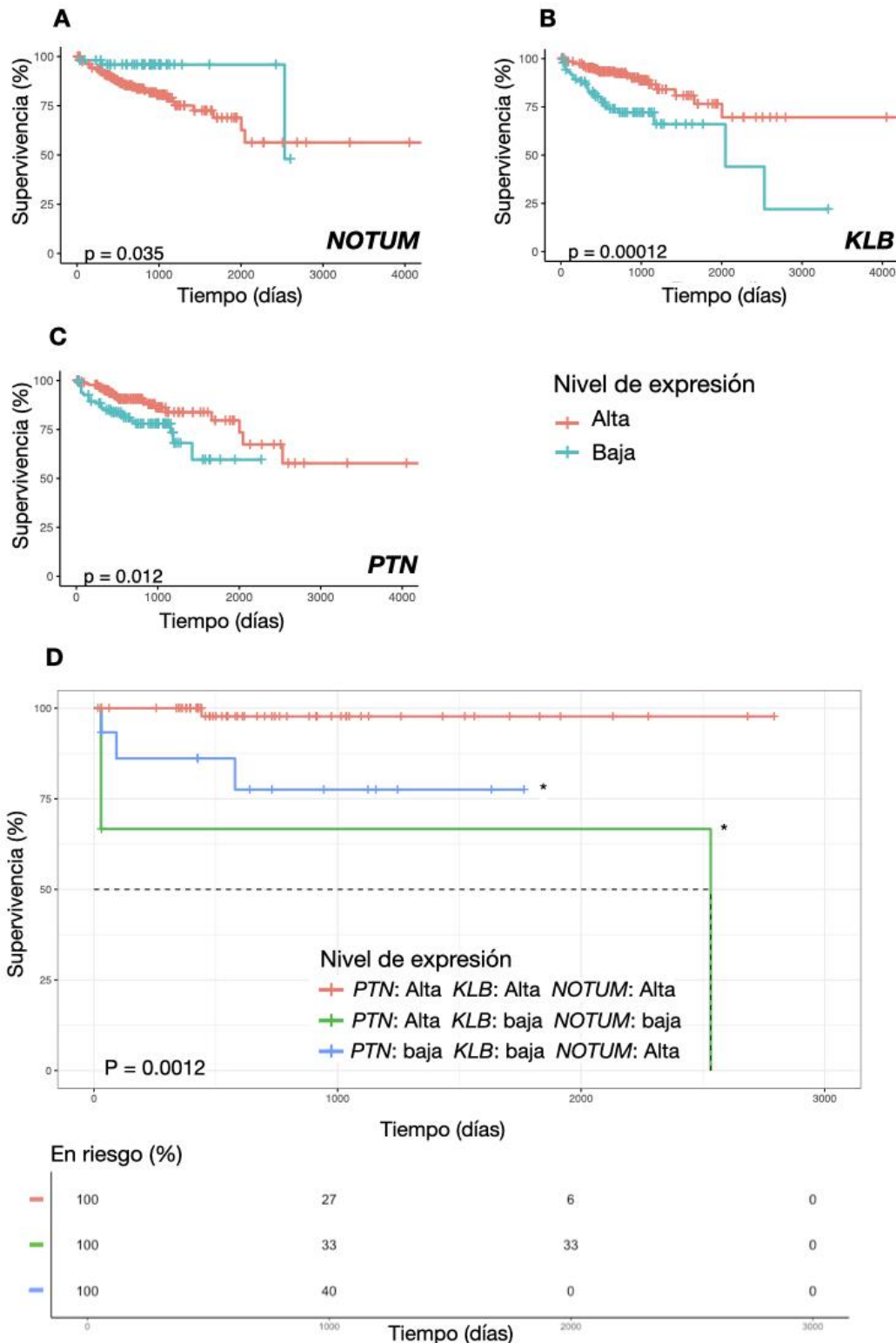


Figura 27: **A.** Curva de supervivencia en base a la expresión de *NOTUM*. **B.** Curva de supervivencia en base a la expresión de *KLB*. **C.** Curva de supervivencia en base a la expresión de *PTN*. **D.** Curva de supervivencia en base a la combinación de expresión de *NOTUM*, *KLB* y *PTN* filtrados por estadio I y II.

5. Discusión

5.1 Análisis de los algoritmos para caracterización del infiltrado inmune

Los métodos experimentales tradicionales para estudiar la heterogeneidad celular, tales como la inmunohistoquímica y la citometría de flujo, han permitido evaluar el infiltrado inmune tumoral en numerosas muestras de pacientes y modelos animales experimentales. Sin embargo, se basan en un repertorio limitado de marcadores fenotípicos. Asimismo, la desagregación del tejido como paso previo a la citometría de flujo puede provocar la pérdida o daño celular, alterando los resultados observados (Newman *et al.*, 2019).

A partir del advenimiento de los análisis transcriptómicos, el uso de métodos bioinformáticos de deconvolución o de análisis de enriquecimiento de grupo de genes para la caracterización del infiltrado inmune tumoral, junto con el resultado de la expresión diferencial génica, mejoró la estratificación de los pacientes, y en muchos casos, permite determinar *a priori* el potencial beneficio de una terapia, como por ejemplo la inmunoterapia para pacientes de CCR con alta inestabilidad microsatelital (MSI-H) (Guo *et al.*, 2020), e incluso predecir una posible recaída en el tiempo (Kamal *et al.*, 2021). En pacientes de CCR, el grupo de pacientes MSI-H (aproximadamente el 15% del total) presenta una carga mutacional tumoral entre 10 y 100 veces mayor que los tumores MSS, lo cual se asocia a una mayor presencia de neoantígenos y mayor infiltrado inmune (Ganesh *et al.*, 2019).

Estos métodos de análisis de poblaciones inmunes a partir de datos de transcriptómica también presentan ciertas limitaciones. En primer lugar, se basan fundamentalmente en la obtención de una firma de expresión génica para poder identificar las distintas poblaciones celulares. Por lo tanto, estos métodos dependen en gran medida de la fuente de origen, pudiendo así provocar un sobreajuste a la hora de caracterizar la composición del infiltrado inmune en distintas muestras (Aran *et al.*, 2017).

Otra de las importantes limitaciones de estos algoritmos radica en la detección de poblaciones celulares que están presentes en muy bajas proporciones (Shen-Orr y Gaujoux, 2013). Por ejemplo, las células asesinas naturales (NK) están presentes con baja frecuencia en tumores de melanoma, aunque desempeñan papeles importantes en el microambiente tumoral (Balch *et al.*, 1990; Sconocchia *et al.*, 2012), por lo que se requiere un conocimiento previo de la patología para poder evaluar qué algoritmo será el más útil, o en su defecto, estudios complementarios que permitan una visión más general y exploratoria. En el caso del método

EPIC, por ejemplo, las predicciones del algoritmo son confiables dentro de un error absoluto del 8%, estimado por el RMSE, lo que afecta mayormente las predicciones en poblaciones celulares que se encuentran en bajas proporciones. Asimismo, en el caso de pacientes de CCR, el centro de este trabajo, en los pacientes clasificados como MSS se espera en principio que el infiltrado inmune en el microambiente tumoral sea bajo.

Otra limitación importante de estos métodos de análisis resulta la detección de poblaciones con alta similitud transcripcional. En este sentido, por ejemplo, los linfocitos T cooperadores y Tregs solamente difieren en la expresión de pocos genes, haciendo difícil distinguir ambas poblaciones con precisión (Racle *et al.*, 2017), aunque xCell puede detectar ambas poblaciones.

En el caso particular de Cibersort, una de las desventajas de este algoritmo de deconvolución es que no brinda un P valor por cada tipo de célula inmune detectada, sino un P valor global. Además, este método tiende a sobrestimar o subestimar sistemáticamente algunos tipos de células (Fernández *et al.*, 2020) y aunque provee un RMSE no queda claro si la predicción es confiable dentro de un cierto error absoluto.

En el caso de quanTIsec, también sufre de limitaciones para distinguir poblaciones con similitud transcripcional, lo que puede dar como resultado una subestimación o la no detección de las células Treg cuando se encuentran baja abundancia (Finotello *et al.*, 2019).

Cabe notar que recientemente se describió un nuevo método de análisis poblacional mediante un algoritmo de deconvolución basado en regresión de soporte vectorial (SVR), denominado MIXTURE, el cual presenta alta robustez para la identificación de las poblaciones celulares y la estimación de sus proporciones (Fernández *et al.*, 2020).

Para métodos de enriquecimiento como xCell, es importante notar que las inferencias poblacionales son estrictamente puntuaciones de enriquecimiento y no pueden interpretarse como proporciones celulares. Esto se debe a la incapacidad de traducir los puntajes mínimos y máximos producidos por ssGSEA (Single Sample GSEA) a proporciones claras (Aran *et al.*, 2017).

De los cuatro algoritmos mencionado, xCell y Cibersort son los que mayor número de poblaciones detectan en la caracterización del infiltrado inmune. El uso simultáneo de ambos métodos puede resultar útil para su complementación, por ejemplo, para los Macrófagos M0, ya que Cibersort los puede detectar y xCell no. Sin embargo, debido a las características intrínsecas de cada método mencionadas anteriormente, también es posible obtener resultados

distintos para una misma población. Estas variaciones en los resultados pueden ser explicadas por el algoritmo empleado y la firma genética utilizada por cada uno.

Si quisiéramos comparar estrictamente las diferencias de cada método, se debería utilizar una misma firma para evaluar los tres algoritmos de deconvolución y constatarlo con datos obtenidos de bases públicas donde se haya caracterizado el infiltrado inmune por técnicas no bioinformáticas, como por ejemplo inmunohistoquímica.

Para los tres modelos experimentales murinos analizados, sólo en dos de ellos se pudo caracterizar el infiltrado inmune, ya que el tercero, GSE43338, no era compatible con ImmuCC. En el GSE64658, al comparar las muestras tumorales de los modelos temprano y tardío de CCRAC basados en la administración de AOM-DSS con el tejido control, se observaron variaciones significativas en células dendríticas activadas, neutrófilos, células Natural killers (NK) en reposo, células NK activadas, mastocitos, células plasmáticas, linfocitos T $CD4^+$ foliculares (Thf) y linfocitos T $CD8^+$ vírgenes (*naive*). Chen *et al.* (2018) describieron que los macrófagos, los linfocitos B y los linfocitos T $CD4^+$ representaban las poblaciones inmunes mayoritarias en colon normal, mientras que en el modelo murino de AOM-DSS se observaba un aumento en neutrófilos y linfocitos T $CD8^+$, lo cual concuerda con nuestros resultados.

En el caso particular del dataset GSE57533 (Sección 4.1.3.3), ninguna de las poblaciones analizadas se encontró significativamente alterada, lo cual, si bien resultó llamativo, podría ser consecuencia del pequeño tamaño muestral del estudio (Control = 3, EII = 4, CCRAC = 3), o debido a la falta de sensibilidad del algoritmo. Asimismo, cabe notar que la caracterización del infiltrado inmune en ratones a partir de RNA-Seq, a diferencia de microarray, ofrece resultados limitados en la detección de poblaciones (9 vs. 25, respectivamente). En ese sentido, los datasets basados en microarrays para la caracterización del infiltrado inmune en modelos experimentales murinos resulta más conveniente.

5.2 Glicofirmas en cáncer colorrectal

A partir de los resultados de esta Tesis, pudimos establecer una firma genética en común entre los modelos experimentales murinos de CCRAC y *Apc^{min/+}* y pacientes con cáncer colorrectal (*NOTUM*, *KLB* y *PTN*), de forma independientemente de las características del tumor en la patología humana, y proponerlos como posibles blancos moleculares. En el caso particular de *Ptn*, la desregulación se produce de manera inversa entre pacientes (subexpresado) y modelos experimentales (sobrexpresado).

Notum es un gen que codifica para una palmitoleoil-proteína carboxilesterasa. Su clasificación como glicogen obedece a su actividad como fosfolipasa, liberando glipicanos, una clase de proteoglicanos. Los mismos están implicados en la estabilización, movimiento y/o retención en superficie de Wnt, proteínas secretadas que presentan palmitoilación en residuos serina altamente conservados, y que han sido ampliamente implicadas en procesos de señalización celular (Kakugawa *et al.*, 2015). Si bien aún su función biológica no ha sido completamente establecida, esta proteína ha sido recientemente propuesta como potencial marcador de diagnóstico y de pronóstico en cáncer colorrectal (Yoon, *et al.*, 2018), ya que se ha demostrado que la expresión de la misma aumenta en células metastásicas, y la inhibición de su expresión reduce la proliferación del tumor. También se la ha implicado en otro tipo de enfermedades, como por ejemplo enfermedades neurodegenerativas, y recientemente se han descrito inhibidores de esta proteína (Zhao *et al.*, 2021), que ayudarían a comprender su potencial papel en patologías como CCR. Los resultados obtenidos en esta Tesis indican que el NOTUM se desregula de manera equivalente a la patología humana en modelos de cáncer colorrectal, por lo que la caracterización de su función en modelos es factible.

La familia de proteínas Klotho, incluyendo a β Klotho (KLB), son componentes esenciales de los complejos receptores de factor de crecimiento de fibroblasto (FGF, del inglés *fibroblast growth factor*), y se requieren para la afinidad de FGF19, FGF21 y FGF23 a sus receptores (FGFRs). Se considera que estas proteínas forman un sistema endócrino que gobierna el metabolismo en mamíferos. Por ejemplo, FGF19 es una hormona de saciedad que se secreta en el intestino ante la ingestión de alimentos y une el complejo β Klotho–FGFR4 en hepatocitos, despertando la respuesta a la misma. En cambio, en condiciones de ayuno, el hígado secreta FGF21, lo que induce la respuesta metabólica al ayuno que activa el eje hipotalamo-pituitaria-adenal vía unión del complejo β Klotho–FGFR1c en adipocitos. En los últimos años, se ha descrito a este sistema FGF–Klotho como crucial en las patologías asociadas al envejecimiento, como diabetes y cáncer (Kuro-o, 2019). En particular, la clasificación de Klotho-b como glicogen se basa en su homología de secuencia a la familia 1 de beta-glicosidasas, pero carece de los residuos de ácido glutámico esenciales para la actividad catalítica (Ito *et al.*, 2000). KLB se expresa mayoritariamente en páncreas, hígado y tejido adiposo, y es un correceptor clave para la unión de FGF19 y FGF21 a su receptor, lo que activa vías de señalización que regulan la síntesis de sales biliares y el metabolismo energético (Kurosu y Kuro-o, 2009). La interacción de KLB con FGFR4, mejora la afinidad de este

receptor por FGF19 (Liu *et al.*, 2020), y esta unión forma homodímeros o heterodímeros que activan vías de señalización que incluyen Ras-Raf- MAPK, PI3K-AKT y JAK/STAT. Se ha demostrado que la subexpresión de KLB en cáncer colorrectal (así como en cáncer de mama, pancreático y de estómago) se produce por la hipermetilación de su promotor. También se demostró que KLB interacciona con FGFR4 induciendo apoptosis e inhibiendo la proliferación de células de hepatoma, y que su baja expresión en carcinoma de mama invasivo se asoció negativamente con el grado patológico y la metástasis linfática (Dallol *et al.*, 2015), pero todavía su papel en cáncer colorrectal no ha sido estudiado en profundidad. Considerando su alta expresión en células endócrinas del intestino, y su presencia en células secretoras de mucinas, enterocitos y células paneth, todas de tejidos no tumorales, tal como se observa en <https://www.proteinatlas.org/ENSG00000134962-KLB/celltype> (Uhlén *et al.*, 2015), KLB se presenta como un blanco terapéutico interesante para estudiar en CCR, y según nuestros resultados, su función podría ser estudiada en modelos experimentales de esta patología.

Finalmente, PTN (Pleiothropin) es una citoquina considerada además una lectina, ya que reconoce proteoglicanos de superficie vía la porción glicosídica (Rauvala, 1989). En ese sentido, se han descrito muchas funciones para la misma en proliferación, crecimiento y migración celular. Su aumento de expresión se ha visto correlacionado con actividad proangiogénica en glioma y cáncer colorectal (Kong *et al.*, 2012; Zhang *et al.*, 2015), con progresión de tumores melanocíticos y potencial metastásico (Wu *et al.*, 2005) y con características histopatológicas, invasión perineural y prognosis en cáncer de páncreas (Yao *et al.*, 2009). Además, en células MCF7 (cáncer de mama) la expresión de PTN remodela la matriz extracelular, aumentando la angiogénesis y estimulando las células estromales (Sorrelle *et al.*, 2017), y recientemente se ha postulado a PTN como biomarcador en pacientes con cáncer de mama (Ma *et al.*, 2017).

El papel proangiogénico de esta proteína se asocia a la interacción con su receptor proteína tirosina fosfatasa beta/zeta (RPTPβ/ζ), y se la ha propuesto como un nuevo blanco de terapias antiangiogénicas (Papadimitriou *et al.*, 2016). Si bien fue propuesta como marcador pronóstico en cáncer colorrectal en 2012 (Kong *et al.*, 2012; Zhou, *et al.*, 2018), esta propuesta no ha llegado aún a la clínica. Dados los resultados obtenidos en esta Tesis, que demuestran que la desregulación en los modelos de AOM-DSS y *Apc^{min/+}* es inversa a la observada en pacientes, nos hace pensar que los estudios mecanísticos y/o preclínicos en modelos

experimentales se ven limitados debido a diferencias entre modelos experimentales y patología (Bürtin *et al.*, 2020).

5.3 *Plaur*: ¿La conexión glicobiológica entre EII y CCRAC?

Nuestros resultados en modelos experimentales mostraron que al comparar la firma glicobiológica establecida para CCRAC y CCRE, quince de los dieciséis genes (*Notum*, *Lyz1*, *Clec4n*, *Clec7a*, *Ptn*, *Srgn*, *Pgc1*, *Pfkfb3*, *Icam1*, *Sulf2*, *Lgals9*, *Cd44*, *Ces1d*, *Acer1* y *Klb*) se comparten en ambos modelos, indicando mecanismos comunes a ambas patologías, y que sientan las bases para la búsqueda de nuevos blancos terapéuticos en cáncer colorrectal. Por ello, en un futuro cercano en el laboratorio nos interesaría estudiar el papel de Galectina-9 (*Lgals9*), una lectina con afinidad por residuos β -galactósidos, en el contexto de EII y CCRAC. En el laboratorio ya contamos con un proyecto centrado en Galectina-1 como molécula clave en el escape tumoral en CCR y CCRAC, que se ha cristalizado recientemente en un artículo original (Cagnoni *et al.*, 2021), y nos encontramos llevando adelante un proyecto sobre la influencia de Galectina-4 (Tesis Doctoral CONICET Mora Massaro, datos no publicados) tanto en el contexto de las EII como en CCRAC.

En este sentido, al comparar el modelo de CCRAC (AOM-DSS) con el de EII (DSS, sección 4.1.3.2), el gen *Plaur* pareciera desregularse ya con la inflamación intestinal crónica, y podría ser un blanco terapéutico a estudiar específicamente en la neoplasia asociada a colitis, ya que no se encuentra desregulado en el modelo de carcinogénesis por *Apc^{min/+}*. Estos resultados son interesantes, ya que *Plaur* (también llamado uPAR, del inglés *urokinase plasminogen activator receptor*) ha sido descrito como un gen de relevancia en las EII y en modelos experimentales de las mismas (Genua *et al.*, 2015). La respuesta inmune exacerbada observada en pacientes de EII lleva a un proceso de remodelamiento tisular en el que la activación de la cascada de plasminógeno, incluyendo metaloproteasas y uPAR, juega un papel esencial. Los pacientes con EII activa presentan un aumento en la expresión de uPAR asociado a la presencia de macrófagos en las lesiones, y se demostró que al eliminar este gen en modelos experimentales de ratones se amplifica la liberación de citoquinas proinflamatorias (Baart *et al.*, 2020). En tanto, también se ha demostrado un papel importante de *Plaur* en cáncer: uPAR está ausente en tejidos normales, pero su papel en diferenciación y proliferación celular le ha garantizado atención en oncología, aunque los mecanismos involucrados no han sido completamente esclarecidos (Baart *et al.*, 2020). En CCR, la expresión de este gen presenta una

asociación negativa con la supervivencia de los pacientes (Boonstra *et al.*, 2014), y se han propuesto moléculas pequeñas que actúan como inhibidores de *Plaur* para el tratamiento de tumores de colon con mutaciones en RAS (Di Mauro *et al.*, 2017). Sin embargo, los estudios en CCR sobre *Plaur* no abordan la neoplasia asociada a la inflamación intestinal en particular. Los resultados obtenidos en los modelos experimentales relacionados con el gen *Plaur* (sección 4.1.3.2) no pudieron verificarse en el análisis de pacientes realizado en esta Tesis, ya que los datos clínicos de la cohorte de TCGA-COAD analizada no distingue entre cáncer colorectal asociado a colitis y esporádico. En el laboratorio nos proponemos seguir trabajando sobre el papel de este gen y discernir si el mismo podría ser un blanco terapéutico en CCRAC específicamente, ya que estos pacientes, debido a la inflamación intestinal que presentan, son una población compleja en su tratamiento. Los pacientes de CCRAC suelen tener pronósticos mas desfavorables que los de CCR esporádico, y la colectomía es mucho más frecuente en esta población. Además, los efectos de los inmunosupresores utilizados para tratar las EII disminuyen la capacidad de respuesta inmune antitumoral, por lo que presentan menor infiltrado y por lo tanto, las inmunoterapias no son eficaces en esta población. Estudios recientes, demuestran que altos niveles de expresión PD-L1⁺ se asocian a una mejor supervivencia en CCR esporádico, pero no en CCRAC, por lo tanto, el uso de inmunoterapias con anticuerpos anti-PD-L1 no sería efectivo en estos pacientes (Soh *et al.*, 2019). Un posible futuro estudio en el laboratorio, basado en este trabajo de Tesis, sería estudiar el potencial papel de inhibidores de *Plaur* en el modelo experimental de CCRAC.

Bibliografía

ABDUL-LATIFA, Mohammed, *et al.* Anti-tumour Treatment Immunotherapy in gastrointestinal cancer: The current scenario and future perspectives. *Cancer Treatment Reviews*. Ago. 2020, 88, 102030. <https://doi.org/10.1016/j.ctrv.2020.102030>

ABU-REMAILEH, Monther, *et al.* Chronic Inflammation Induces a Novel Epigenetic Program That Is Conserved in Intestinal Adenomas and in Colorectal Cancer. *Cancer Research*. May. 2015, 75 (10), 2120-2130. <https://doi.org/10.1158/0008-5472.can-14-3295>

ADAMCZYK, Barbara, THARMALINGAM, Tharmala y RUDD, Pauline M. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects*. Sep. 2012, 1820 (9), 1347–1353. <https://doi.org/10.1016/j.bbagen.2011.12.001>

ALTUNEL, E., *et al.* Development of a precision medicine pipeline to identify personalized treatments for colorectal cancer. *BMC Cancer*. Jun. 2020, 20, 592. <https://doi.org/10.1186/s12885-020-07090-y>

ANDERS, Simon, PYL, Paul Theodor y HUBER, Wolfgang. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. Ene. 2015, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>

ANDREWS, Simon. *FastQC: a quality control tool for high throughput sequence data*. Cambridge: Babraham Institute, 2010. Disponible en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

ANGELOVA, M. *et al.* Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol*. 16, 64 (2015). <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0620-6>

AXELRAD, Jordan E., LICHTIGER, Simon y YAJNIK, Vijay. Inflammatory bowel disease and cancer: The role of inflammation, immunosuppression, and cancer treatment. *World Journal of Gastroenterology*. May. 2016, 22(20), 4794-4801. <https://dx.doi.org/10.3748/wjg.v22.i20.4794>

ARAN, Dvir, HU, Zicheng y BUTTE, Atul J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*. Nov. 2017, 18, 220. <https://doi.org/10.1186/s13059-017-1349-1>

BAART, V. M., *et al.* Molecular imaging of the urokinase plasminogen activator receptor: opportunities beyond cancer. *EJNMMI Research*. Jul. 2020, 10, 87. <https://doi.org/10.1186/s13550-020-00673-7>

BARRETT, Tanya, *et al.* NCBI GEO: archive for functional genomics data sets - Update. *Nucleic Acids Research*. Ene. 2013, 41(D1), D991-D995. <https://doi.org/10.1093/nar/gks1193>

BEAUGERIE, Laurent e ITZKOWITZ, Steven H. Cancers Complicating Inflammatory Bowel Disease. *New England Journal of Medicine*. Abr. 2015, 372(15), 1441–1452. DOI: [10.1056/NEJMra1403718](https://doi.org/10.1056/NEJMra1403718)

BLIGHE Kevin, *et al.* *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling*. R package version 1.8.0. 2020. Disponible en: <https://github.com/kevinblighe/EnhancedVolcano>

BOONSTRA, Martin C., *et al.* Expression of uPAR in tumor-associated stromal cells is associated with colorectal cancer patient prognosis: a TMA study. *BMC Cancer*. Abr. 2017, 14, 269. <https://doi.org/10.1186/1471-2407-14-269>

BRAY, Freddie, *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. Sep. 2018, vol. 68(6), 394–424. <https://doi.org/10.3322/caac.21492>

BRISTER, J. Rodney, *et al.* NCBI viral genomes resource. *Nucleic Acids Research*. Ene. 2015, 43(Database issue), D571-D577. <https://doi.org/10.1093/nar/gku1207>

BÜRTIN, Florian, MULLINS, Christina S. y LINNEBACHER, Michael. Mouse models of colorectal cancer: Past, present and future perspectives. *World J Gastroenterol*. Abr. 2020, 26(13), 1394-1426. <https://doi.org/10.3748/wjg.v26.i13.1394>

CAGNONI, Alejandro J., *et al.* Turning-Off Signaling by Siglecs, Selectins, and Galectins: Chemical Inhibition of Glycan-Dependent Interactions in Cancer. *Frontiers in Oncology*. May. 2016, 6, 109. <https://doi.org/10.3389/fonc.2016.00109>

CAGNONI, Alejandro J., *et al.* Galectin-1 foster an immunosuppressive microenvironment in colorectal cancer by reprogramming CD8⁺ regulatory T cells. *Proceedings of the National Academy of Sciences of the United States of America*. May. 2021, 118(21), e2102950118. <https://doi.org/10.1073/pnas.2102950118>

CARLSON, Marc. *moe430a.db: Affymetrix Mouse Expression Set 430 annotation data (chip moe430a)*. R package version 3.2.3. 2016a. DOI: [10.18129/B9.bioc.moe430a.db](https://doi.org/10.18129/B9.bioc.moe430a.db)

CARLSON, Marc. *mouse4302.db: Affymetrix Mouse Genome 430 2.0 Array annotation data (chip mouse4302)*. R package version 3.2.3. 2016b. DOI: [10.18129/B9.bioc.mouse4302.db](https://doi.org/10.18129/B9.bioc.mouse4302.db)

CATALANO, Irene, GRASSI, Elena, BERTOTTI, Andrea, y TRUSOLINO, Livio. Immunogenomics of Colorectal Tumors: Facts and Hypotheses on an Evolving Saga. *Trends in Cancer*. Dic. 2019, vol. 5(12), 779-788. <https://doi.org/10.1016/j.trecan.2019.10.006>

CHANG, Wen-Wei, *et al.* Expression of Globo H and SSEA3 in breast cancer stem cells and the involvement of fucosyl transferases 1 and 2 in Globo H synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. Ago. 2008, vol. 105(33), 11667-11672. <https://doi.org/10.1073/pnas.0804979105>. Erratum in: *Proceedings of the National Academy of Sciences of the United States of America*. Nov. 2008, 105(44), 17206. <https://doi.org/10.1073/pnas.0808811105>

CHEN, Ziyi, *et al.* Inference of immune cell composition on the expression profiles of mouse tissue. *Scientific reports*. Ene. 2017, 7, 40508. <https://doi.org/10.1038/srep40508>

CHEN, Ziyi, *et al.* seq-ImmuCC: Cell-Centric View of Tissue Transcriptome Measuring Cellular Compositions of Immune Microenvironment From Mouse RNA-Seq Data. *Frontiers in Immunology*. Jun. 2018, 9, 1286. <https://doi.org/10.3389/fimmu.2018.01286>.

CHIK, Jenny H. L., *et al.* Comprehensive glycomics comparison between colon cancer cell cultures and tumours: implications for biomarker studies. *Journal of proteomics*. Ago. 2014, 108, 146-162. <https://doi.org/10.1016/j.jprot.2014.05.002>

CHRISTIANSEN Maja N., *et al.* Cell surface protein glycosylation in cancer. *Proteomics*. Mar. 2014, 14(4–5), 525–546. <https://doi.org/10.1002/pmic.201300387>

DALLOL, Ashraf, *et al.* Frequent methylation of the KLOTHO gene and overexpression of the FGFR4 receptor in invasive ductal carcinoma of the breast. *Tumor Biology*. Jul. 2015, 36, 9677–9683. <https://doi.org/10.1007/s13277-015-3733-3>

DAVIS, Sean y MELTZER, Paul S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. Jul. 2007, 23(14), 1846–1847. <https://doi.org/10.1093/bioinformatics/btm254>

DE FREITAS JUNIOR, Julio Cesar Madureira y MORGADO-DÍAZ, José Andrés. The role of *N-glycans* in colorectal cancer progression: potential biomarkers and therapeutic applications. *Oncotarget*. Ene. 2016, 7, 19395–19413. <https://doi.org/10.18632/oncotarget.6283>

DE ROBERTIS, Mariangela, *et al.* The AOM/DSS murine model for the study of colon carcinogenesis: From pathways to diagnosis and therapy studies. *Journal of Carcinogenesis*. Mar. 2011, 10(1), 9. <https://doi.org/10.4103/1477-3163.78279>

DI MAURO, Concetta, *et al.* Urokinase-type plasminogen activator receptor (uPAR) expression enhances invasion and metastasis in RAS mutated tumors. *Scientific Reports*. Ago. 2017, 7, 9388. <https://doi.org/10.1038/s41598-017-10062-1>

DIENSTMANN, Rodrigo, *et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer*. Ene. 2017, 17, 79–92. <https://doi.org/10.1038/nrc.2016.126>

DOBIN, Alexander, *et al.* STAR: ultrafast universal RNA-seq aligner, *Bioinformatics* [en línea]. Ene. 2013, 29 (1), 15–21, <https://doi.org/10.1093/bioinformatics/bts635>

DURINCK, Steffen, *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. Ago. 2005, 21(16), 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>

DURINCK, Steffen, *et al.* Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*. Jul. 2009, 4(8), 1184–1191. <https://doi.org/10.1038/nprot.2009.97>

ELIOPOULOS, Aristides G. y PAVLOPOULOS, Georgios A. *Gene expression profiling of proximal versus distal mouse colon during AOM/DSS-induced colitis-associated carcinogenesis* [en línea]. Heraklion: University of Crete. Ene. 2015. Disponible en: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64658>

FAKIH, Marwan, *et al.* Immune overdrive signature in colorectal tumor subset predicts poor clinical outcome. *The Journal of Clinical Investigation*. Oct. 2019, 129(10), 4464–4476. <https://doi.org/10.1172/JCI127046>.

FERNÁNDEZ, Elmer A., *et al.* Unveiling the immune infiltrate modulation in cancer and response to immunotherapy by MIXTURE—an enhanced deconvolution method. *Briefings in Bioinformatics*. Dic. 2020, bbaa317. <https://doi.org/10.1093/bib/bbaa317>

FINOTELLO, Francesca, *et al.* Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Medicine*. May. 2019, 11, 34. <https://doi.org/10.1186/s13073-019-0638-6>

FRITZ, Jill M., y LENARDO, Michael J. Development of immune checkpoint therapy for cancer. *The Journal of Experimental Medicine*. Jun. 2019, 216(6), 1244–1254. <https://doi.org/10.1084/jem.20182395>

GANESH, Karuna, *et al.* A rectal cancer organoid platform to study individual responses to chemoradiation. *Nature Medicine*. Oct. 2019, 25, 1607–1614 <https://doi.org/10.1038/s41591-019-0584-2>

GAUTIER, Laurent, *et al.* affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. Feb. 2004, 20(3), 307–315. <https://doi.org/10.1093/bioinformatics/btg405>

GENUA, *et al.* The urokinase plasminogen activator receptor (uPAR) controls macrophage phagocytosis in intestinal inflammation. *Gut*. 2015, 64(4), 589–600. <http://dx.doi.org/10.1136/gutjnl-2013-305933>

GKOUSKOU, Kalliopi, *et al.* Apolipoprotein A-I inhibits experimental colitis and colitis-propelled carcinogenesis. *Oncogene*, May. 2016, 35(19), 2496–2505. <https://doi.org/10.1038/onc.2015.307>

GHOSN, Marwan, TABCHI, Samer, y KOURIE, Hampig Raphael. Immunotherapy in gastrointestinal cancers. En: TYAGI, Amit Kumar y PRASAD Sahdeo (Editores). *Gastrointestinal Cancers: Prevention, Detection and Treatment, Volumen 1*. Houston: Department of Experimental Therapeutics, The University of Texas MD Anderson Center, 2016, 343–356. <https://doi.org/10.1891/9781617052736.0043>.

GILEWSKI, Teresa, *et al.* Immunization of metastatic breast cancer patients with a fully synthetic globo H conjugate: A phase I trial. *Proceedings of the National Academy of Sciences of the United States of America*. Mar. 2001, 98(6), 3270–3275. <https://doi.org/10.1073/pnas.051626298>

GUO, Liang, WANG, Chuanlei, QIU, Xiang, PU, Xiaoyu, y CHANG, P. Colorectal Cancer Immune Infiltrates: Significance in Patient Prognosis and Immunotherapeutic Efficacy. *Frontiers in Immunology*. May. 2020, 11(1050), 1–13. <https://doi.org/10.3389/fimmu.2020.01052>

GUINNEY, Jutin, *et al.* The consensus molecular subtypes of colorectal cancer. *Nature Medicine*. Oct. 2015, 21, 1350–1356. <https://doi.org/10.1038/nm.3967>

HAKOMORI, Senitiroh, WANG, S. M., y YOUNG W. W. Jr. Isoantigenic expression of Forssman glycolipid in human gastric and colonic mucosa: its possible identity with "A-like antigen" in human cancer. *Proceedings of the National Academy of Sciences of the United States of America*. Jul. 1977, 74(7), 3023–3027. <https://doi.org/10.1073/pnas.74.7.3023>

HAKOMORI, Senitiroh. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proceedings of the National Academy of Sciences of the United States of America*. Ago. 2002, 99(16), 10231–10233. <https://doi.org/10.1073/pnas.172380699>

HARVEY, David J., *et al.* Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics*. Ago. 2009, 9(15), 3796–3801. <https://doi.org/10.1002/pmic.200900096>

HAZAMA, Shoichi, *et al.* Current status of immunotherapy against gastrointestinal cancers and its biomarkers: Perspective for precision immunotherapy. *Annals of Gastroenterological Surgery*. Jun. 2018, 2(4), 289–303. <https://doi.org/10.1002/ags3.12180>.

HUBBARD, Tim, *et al.* The Ensembl genome database project. *Nucleic Acids Research*. Ene. 2002, 30(1), 38–41. <https://doi.org/10.1093/nar/30.1.38>

HUBER, Wolfgang, *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*. Ene. 2015, 12, 115–121. <https://doi.org/10.1038/nmeth.3252>

HULL, M.A., *et al.* A risk-stratified approach to colorectal cancer prevention and diagnosis. *Nature Reviews Gastroenterology & Hepatology*. Oct. 2020, 17, 773–780. <https://doi.org/10.1038/s41575-020-00368-3>

HUMAN PROTEIN ATLAS [en línea]. [consulta: 2 agosto 2021]. Disponible en: <http://www.proteinatlas.org>

ITO, Shinji, *et al.* Molecular cloning and expression analyses of mouse *βklotho*, which encodes a novel Klotho family protein. *Mechanisms of Development*. Nov. 2000, 98(1-2), 115-119. [https://doi.org/10.1016/S0925-4773\(00\)00439-1](https://doi.org/10.1016/S0925-4773(00)00439-1)

KAKUGAWA, Satoshi, *et al.* Notum deacylates Wnt proteins to suppress signalling activity. *Nature*. Mar. 2015, 519(7542), 187-192. <https://doi.org/10.1038/nature14259>

KAMAL, Yasmin, *et al.* Tumor immune infiltration estimated from gene expression profiles predicts colorectal cancer relapse. *OncoImmunology*. Mar. 2021, 10(1), e1862529. <https://doi.org/10.1080/2162402X.2020.1862529>

KASSAMBARA, Alboukadel. *ggpubr: 'ggplot2' Based Publication Ready Plots*. R Package version 0.4.0. 2020a. Disponible en: <https://rpkgs.datanovia.com/ggpubr/>

KASSAMBARA, Alboukadel. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R Package version 0.6.0. 2020b. Disponible en: <https://rpkgs.datanovia.com/rstatix/>

KIM, Chang Gong, *et al.* VEGF-A drives TOX-dependent T cell exhaustion in anti-PD-1-resistant microsatellite stable colorectal cancers. *Science Immunology*. Nov. 2019, 4(41), 1–18. <https://doi.org/10.1126/sciimmunol.aay0555>

KIRWAN, Alan, *et al.* Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *BioMed Research International*. Oct. 2015, 2015, 490531. <https://doi.org/10.1155/2015/490531>

KOLDE Raivo. *pheatmap: Pretty Heatmaps*. R package version 1.0.12. 2019. Disponible en: <https://CRAN.R-project.org/package=pheatmap>

KONG, Ying, *et al.* Pleiotrophin is a potential colorectal cancer prognostic factor that promotes VEGF expression and induces angiogenesis in colorectal cancer. *International Journal of Colorectal Disease*. Mar. 2012, 27(3), 287-298. <https://doi.org/10.1007/s00384-011-1344-z>

KRUSKAL, William H., y WALLIS, W. Allens. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*. Dic. 1952, 47, 583– 621.

KURO-O Makoto. The Klotho proteins in health and disease. *Nature Reviews Nephrology*. Ene. 2019, 15, 27-44. <https://doi.org/10.1038/s41581-018-0078-3>

KUROSU, Hiroshi, y KURO-O Makoto. The Klotho gene family as a regulator of endocrine fibroblast growth factors. *Molecular and cellular endocrinology*. Feb. 2009, 299(1), 72-78. <https://doi.org/10.1016/j.mce.2008.10.052>

LEINONEN, Rasko, *et al.* The sequence read archive. *Nucleic Acids Research*. Ene. 2011, 39(1), D19-D21. <https://doi.org/10.1093/nar/gkq1019>

LI, Heng, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. Ago. 2009, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

LI, Xin-Xiang, *et al.* Klotho suppresses growth and invasion of colon cancer cells through inhibition of IGF1R-mediated PI3K/AKT pathway. *International Journal of Oncology*. May. 2014, 45(2), 611-618. <https://doi.org/10.3892/ijo.2014.2430>

LI C, *et al.* Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. *Mol Cancer*. 2020; 19(1):116. <https://doi.org/10.1186/s12943-020-01234-1>

LIAO, Yu-Mai, *et al.* High B3GALT5 expression confers poor clinical outcome and contributes to tumor progression and metastasis in breast cancer. *Breast Cancer Research*. Ene. 2021, 23(1), 5. <https://doi.org/10.1186/s13058-020-01381-9>

LIN, Otto S. Acquired Risk Factors for Colorectal Cancer. *Methods in Molecular Biology*. Feb. 2009, 472, 361-372. https://doi.org/10.1007/978-1-60327-492-0_16

LIN, Y, *et al.* Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *J Hematol Oncol*. 2019, 12(1):76. <https://doi.org/10.1186/s13045-019-0760-3>

LIU, Ming y GUO, Fukun. Recent updates on cancer immunotherapy. *Precision Clinical Medicine*. Sep. 2018, 1(2), 65-74. <https://doi.org/10.1093/pcmedi/pby011>

LIU, Yanan, *et al.* Dissecting the Role of the FGF19-FGFR4 Signaling Pathway in Cancer Development and Progression. *Frontiers in cell and developmental biology*. Feb. 2020, 8, 95. <https://doi.org/10.3389/fcell.2020.00095>

LOVE, Michael I., HUBER, Wolfgang y ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. Dic. 2014, 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>

LYNCH, David y MURPHY, Adrian. The emerging role of immunotherapy in colorectal cancer. *Annals of Translational Medicine*. Ago. 2016, 4(16), 305. <https://doi.org/10.21037/atm.2016.08.29>

LONG, Junyu, *et al.* PD-1/PD-L blockade in gastrointestinal cancers: Lessons learned and the road toward precision immunotherapy. *Journal of Hematology and Oncology*. Ago. 2017, 10(1), 1–21. <https://doi.org/10.1186/s13045-017-0511-2>

MA, Jiequn, *et al.* Pleiotrophin as a potential biomarker in breast cancer patients. *Clinica chimica acta; international journal of clinical chemistry*. Mar. 2017, 466, 6–12. <https://doi.org/10.1016/j.cca.2016.12.030>.

MARIÑO, Karina, *et al.* A systematic approach to protein glycosylation analysis: a path through the maze. *Nature Chemical Biology*. Sep. 2010, 6(10), 713–723. <https://doi.org/10.1038/nchembio.437>

MARKOWITZ, Sanford D., *et al.* Focus on colon cancer. *Cancer Cell*. Abr. 2002, 1(3), 233-236. [https://doi.org/10.1016/S1535-6108\(02\)00053-3](https://doi.org/10.1016/S1535-6108(02)00053-3)

MARKOWITZ, Sanford D. y BERTAGNOLLI, Monica M. Molecular Basis of Colorectal Cancer. *New England Journal of Medicine*. Dic. 2009, 361(25), 2449–2460. <https://doi.org/10.1056/nejmra0804588>

MEREITER, Stefan, *et al.* Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading?. *Cancer Cell*. Jul. 2019, 36(1), 6-16. <https://doi.org/10.1016/j.ccell.2019.06.006>

NAIRN, Alison V., *et al.* Regulation of Glycan Structures in Animal Tissues. *Journal of Biological Chemistry*. Jun. 2008, 283(25), 17298–17313. <https://doi.org/10.1074/jbc.m801964200>

NARAYANAN, Sumana, *et al.* Tumor Infiltrating Lymphocytes and Macrophages Improve Survival in Microsatellite Unstable Colorectal Cancer. *Scientific Report*. Sep. 2019, 9, 13455. <https://doi.org/10.1038/s41598-019-49878-4>

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. *Genome Assembly Model* [internet]. Actualizado 26 octubre 2017 [consulta: 11 agosto 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/assembly/model/>

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. *Gene database. Dcn decorin [Mus musculus]* [en línea]. Bethesda (MD): National Library of Medicine (US). Actualizada 23 febrero 2021 [consulta: 17 marzo 2021]. Disponible en: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=13179#general-gene-info> Gene ID: 13179

NEELAMEGHAM, Sriram, *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology*. Ago. 2019, 29(9), 620–624. <https://doi.org/10.1093/glycob/cwz045>

NEUFERT, Clemens, *et al.* Tumor fibroblast–derived epiregulin promotes growth of colitis-associated neoplasms through ERK. *The Journal of Clinical Investigation*. Mar. 2013, 123(4), 1428-1443. <https://doi.org/10.1172/JCI63748>

NEWMAN, Aaron M., *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*. Mar. 2015, 12, 453–457. <https://doi.org/10.1038/nmeth.3337>

NEWMAN, Aaron M., *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*. May. 2019, 37, 773–782. <https://doi.org/10.1038/s41587-019-0114-2>

OHHARA, Yoshihito, *et al.* Role of targeted therapy in metastatic colorectal cancer. *World Journal Gastrointestinal Oncology*. Sep. 2016, 8(9), 642-655. <https://doi.org/10.4251/wjgo.v8.i9.642>

O'LEARY, Nuala A., *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*. Ene 2016, 44(D1), D733-D745. <https://doi.org/10.1093/nar/gkv1189>

PAGÈS, Franck, *et al.* International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet*. May. 2018, 391(10135), P2128-2139. [https://doi.org/10.1016/S0140-6736\(18\)30789-X](https://doi.org/10.1016/S0140-6736(18)30789-X)

PAPADIMITRIOU, Evangelia, *et al.* Pleiotrophin and its receptor protein tyrosine phosphatase beta/zeta as regulators of angiogenesis and cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. Dic. 2016, 1866(2), 252-265. <https://doi.org/10.1016/j.bbcan.2016.09.007>

PARADISI, Andrea, *et al.* Netrin-1 up-regulation in inflammatory bowel diseases is required for colorectal cancer progression. *Proceedings of the National Academy of Sciences*. Oct. 2009, 106(40), 17146-17151. <https://doi.org/10.1073/pnas.0901767106>

PERACLA, Rosa, *et al.* Altered glycosylation in tumours focused to cancer diagnosis. *Disease Markers*. Dic. 2008, 25(4–5), 207–218. <https://doi.org/10.1155/2008/797629>

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. 2019. Disponible en: <https://www.R-project.org/>

RACLE, Julien, *et al.* Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife*. Nov. 2017, 6, e26476. <https://doi.org/10.7554/eLife.26476>

RABINOVICH, Gabriel A. y CROCI, Diego O. Regulatory Circuits Mediated by Lectin-Glycan Interactions in Autoimmunity and Cancer. *Immunity*. Mar. 2012, 36(3), 322–335. <https://doi.org/10.1016/j.immuni.2012.03.004>

RAUVALA, H. An 18-kd heparin-binding protein of developing brain that is distinct from fibroblast growth factors. *The EMBO journal*. OCT. 1989, 8(10), 2933-2941.

REEDIJK, Michael, *et al.* Activation of Notch signaling in human colon adenocarcinoma. *International Journal of Oncology*. Dic. 2008. 33(6), 1223–1229. https://doi.org/10.3892/ijo_00000112

RITCHIE, Matthew E., *et al.* *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. Abr. 2015, 43(7), e47. <https://doi.org/10.1093/nar/gkv007>

ROBERTSON, David S., WILDENHAIN, Jan, JAVANMARD, Adel, KARP, Natasha A. onlineFDR: an R package to control the false discovery rate for growing data repositories, *Bioinformatics*. Oct. 2019, 35(20), 4196–4199. <https://doi.org/10.1093/bioinformatics/btz191>

SHARMA, Padmanee y ALLISON, James P. The future of immune checkpoint therapy. *Science*. Abr. 2015, 348(6230), 56–61. <https://doi.org/10.1126/science.aaa8172>

SHIMIZU, K., *et al.* (2001). Manic fringe and lunatic fringe modify different sites of the Notch2 extracellular region, resulting in different signaling modulation. *The Journal of biological chemistry*. Jul. 2001, 276(28), 25753–25758. <https://doi.org/10.1074/jbc.M103473200>

SOH, *et al.* Immunoprofiling of Colitis-associated and Sporadic Colorectal Cancer and its Clinical Significance. *Scientific Reports*. May. 2019, 9, 6833. <https://doi.org/10.1038/s41598-019-42986-1>

SOLIS HERNÁNDEZ, María del Pilar, *et al.* Small leucine-rich proteoglycans expression in colorectal cancer: Differences by primary tumor location. *Journal of Clinical Oncology*. Feb. 2018, 36(4_suppl), 721-721. DOI: 10.1200/JCO.2018.36.4_suppl.721

SORRELLE, Noah, DOMINGUEZ, Adrián T. A., y BREKKEN, Rolf A. From top to bottom: midkine and pleiotrophin as emerging players in immune regulation. *Journal of Leukocyte Biology*. Mar. 2017, 102(2), 277-286. <https://doi.org/10.1189/jlb.3MR1116-475R>

STIDHAM, Ryan W. y HIGGINS, Peter D. R. Colorectal Cancer in Inflammatory Bowel Disease. *Clinics in Colon and Rectal Surgery*. Ene. 2018, 31(03), 168–178. <https://doi.org/10.1055/s-0037-1602237>

TANIGUCHI, Naoyuki, *et al.* Expression of Forssman Antigen Synthesis and Degradation in Human Lung Cancer, *JNCI: Journal of the National Cancer Institute*. Sep. 1981, vol. 67(3), 577–583. <https://doi.org/10.1093/jnci/67.3.577>

TANNAPFEL, Andrea, y REINACHER-SCHICK, Anke. Immunotherapy in Gastrointestinal Cancer: Where Do We Stand? *Visceral Medicine*. Febrero 2019, 35(1), 1–2. <https://doi.org/10.1159/000497294>.

TATUSOVA, Tatiana, *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*. 19 agosto 2016, 44(14), 6614–6624. <https://doi.org/10.1093/nar/gkw569>

THE CANCER GENOME ATLAS RESEARCH NETWORK, *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*. Sep. 2013, 45, 1113–1120. <https://doi.org/10.1038/ng.2764>

THE UNIVERSAL PROTEIN RESOURCE. *Protein Knowledgebase (UniProtKB) - O09010 (LFNG_MOUSE)* [en línea]. Actualizada 2 diciembre 2020 [consulta: 20 marzo 2021]. Disponible en: <https://www.uniprot.org/uniprot/O09010> UniProtKB entry: O09010

THORSSON, Vésteinn, *et al.* The Immune Landscape of Cancer. *Immunity*. Abr. 2018, 48(4), 812–830.e14. <https://doi.org/10.1016/j.immuni.2018.03.023>

TOSCANO, Marta A., *et al.* Untangling Galectin-Driven Regulatory Circuits in Autoimmune Inflammation. *Trends Molecular Medicine*. Abr. 2018, 24(4), 348–363. <https://doi.org/10.1016/j.molmed.2018.02.008>

UHLÉN, Mathias, *et al.* Tissue-based map of the human proteome. *Science*. Ene. 2015, 347(6220), 1260419–1260419. <http://dx.doi.org/10.1126/science.1260419>

VAN DER MAATEN, Laurens y HINTON, Geoffrey. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. Nov. 2008, 9(86), 2579–2605. <https://www.jmlr.org/papers/v9/vandermaaten08a.html>

VARKI, Ajit, *et al.* (Ed.), *Essentials of Glycobiology* [internet] 3rd Ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2015–2017. Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK310274/>

VASTA, Gerardo R., *et al.* Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. *Frontiers in Immunology*. Jul. 2012, 3,199. <https://doi.org/10.3389/fimmu.2012.00199>

VENKITACHALAM, Srividya, *et al.* Biochemical and functional characterization of glycosylation-associated mutational landscapes in colon cancer. *Scientific Reports*. Mar. 2016, 6, 23642. <https://doi.org/10.1038/srep23642>

WALKER, Jennifer y SMITH, Kenneth G. C. CD22: an inhibitory enigma. *Ene.* 2008, 123(3), 314-325. <https://doi.org/10.1111/j.1365-2567.2007.02752.x>

WANG, Cheng-Chi, *et al.* Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*. Ago. 2008, 105(33), 11661-11666; <https://doi.org/10.1073/pnas.0804923105>

WANG, Zhong, GERSTEIN, Mark y SNYDER, Michael. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*. Ene. 2009, 10, 57-63. <https://doi.org/10.1038/nrg2484>

WENK Jonathan, *et al.* Glycolipids of germ cell tumors: extended globo-series glycolipids are a hallmark of human embryonal carcinoma cells. *International Journal of Cancer*. Jul. 1994, 58(1), 108-115. <https://doi.org/10.1002/ijc.2910580118>

WILCOXON, Frank. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*. Dic. 1945, 1(6), 80-83. <https://doi.org/10.2307/3001968>

WU, H, *et al.* Pleiotrophin expression correlates with melanocytic tumor progression and metastatic potential. *Journal of cutaneous pathology*. Feb. 2005, 32(2), 125-30. <https://doi.org/10.1111/j.0303-6987.2005.00282.x>

WU, Hsin-Jung y BONDADA, Subbarao. Positive and negative roles of CD72 in B cell function. *Immunologic Research*. Abr. 2002, 25(2), 155-166. <https://doi.org/10.1385/IR:25:2:155>

WU, Jean, *et al.* *gcrma: Background Adjustment Using Sequence Information*. R package version 2.56.0. 2020. DOI: [10.18129/B9.bioc.gcrma](https://doi.org/10.18129/B9.bioc.gcrma)

YAO, Jun, *et al.* Pleiotrophin expression in human pancreatic cancer and its correlation with clinicopathological features, perineural invasion, and prognosis. *Digestive diseases and sciences*. Abr. 2009, 54(4), 895–901. <https://doi.org/10.1007/s10620-008-0433-5>

YOON, Jong Hyuk, *et al.* NOTUM Is Involved in the Progression of Colorectal Cancer. *Cancer genomics & proteomics*. Nov. - Dic. 2018, 15(6), 485–497. <https://doi.org/10.21873/cgp.20107>

ZANINI, Fabio, *et al.* HTSeq 2.0 – Efficient manipulation of high-throughput sequencing data with long genomes [En preparación]. Sidney: University of New South Wales, 2021. Disponible en: <https://htseq.readthedocs.io/en/master/>

ZHANG, Lei, *et al.* Pleiotrophin promotes vascular abnormalization in gliomas and correlates with poor survival in patients with astrocytomas. *Science Signaling*. Dic. 2015, 8(406), ra125. <https://doi.org/10.1126/scisignal.aaa1690>

ZHANG, Xiangli, *et al.* ELBOW: ELBOW - Evaluating foLd change By the lOgit Way. R package version 1.26.0. 2020. DOI: [10.18129/B9.bioc.ELBOW](https://doi.org/10.18129/B9.bioc.ELBOW)

ZHAO, Yuguang, *et al.* Small-molecule inhibitors of carboxylesterase Notum. *Future medicinal chemistry*. Jun. 2021, 13(11), 1001–1015. <https://doi.org/10.4155/fmc-2021-0036>

ZHOU, Jiupeng, *et al.* A meta-analysis on the role of pleiotrophin (PTN) as a prognostic factor in cancer. *PLOS ONE*. Nov. 2018, 13(11), e0207473. <https://doi.org/10.1371/journal.pone.0207473>

ZHOU, Julie Y., *et al.* The Glycoscience of Immunity. *Trends in Immunology*. May. 2018, 39(7), 523–535. <https://doi.org/10.1016/j.it.2018.04.004>

ZIEGLER, Steven F, RAMSDELL, Fred y ALDERSON Mark R. The activation antigen CD69. *Stem Cells*. Sep. 1994, 12(5), 456–465. <https://doi.org/10.1002/stem.5530120502>

**Anexo A: Listado de pacientes con cáncer colorrectal de la base de datos
TCGA (COAD)**

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-3L-AA1B-01A	Tumor	Estadio i	mss	Si	--	475
TCGA-4N-A93T-01A	Tumor	Estadio iii	mss	Si	--	146
TCGA-5M-AAT4-01A	Tumor	Estadio iv	mss	No	49	--
TCGA-5M-AAT6-01A	Tumor	Estadio iv	msi-h	No	290	--
TCGA-5M-AATE-01A	Tumor	Estadio ii	mss	No	--	1200
TCGA-A6-3807-01A	Tumor	Estadio iii	mss	No	--	1054
TCGA-A6-3808-01A	Tumor	Estadio ii	msi-l	No	--	1014
TCGA-A6-3809-01B	Tumor	Estadio ii	msi-h	No	--	996
TCGA-A6-4105-01A	Tumor	Estadio ii	mss	No	442	--
TCGA-A6-4107-01A	Tumor	Estadio iii	mss	Si	--	987
TCGA-A6-5656-01B	Tumor	Estadio i	mss	No	--	1001
TCGA-A6-5657-01A	Tumor	Estadio iii	mss	No	--	962
TCGA-A6-5660-01A	Tumor	Estadio iii	mss	No	--	888
TCGA-A6-5661-01A	Tumor	Estadio ii	msi-h	Si	--	1020
TCGA-A6-5664-01A	Tumor	Estadio iii	mss	Si	--	672
TCGA-A6-5666-01A	Tumor	Estadio ii	msi-l	No	--	995
TCGA-A6-6141-01A	Tumor	Estadio ii	mss	Si	--	255
TCGA-A6-6142-01A	Tumor	Estadio iv	mss	Si	--	763
TCGA-A6-6648-01A	Tumor	Estadio iv	mss	Si	--	766
TCGA-A6-6651-01A	Tumor	Estadio iii	mss	No	--	662
TCGA-A6-6652-01A	Tumor	Estadio iv	mss	Si	--	751
TCGA-A6-6653-01A	Tumor	Estadio i	msi-h	Si	--	742
TCGA-A6-6654-01A	Tumor	Estadio iii	mss	Si	--	726
TCGA-A6-6781-01B	Tumor	Estadio iii	msi-h	No	--	598
TCGA-A6-6782-01A	Tumor	Estadio ii	mss	No	--	617
TCGA-A6-A5ZU-01A	Tumor	Estadio iii	mss	Si	--	293
TCGA-AA-3488-01A	Tumor	Estadio iv	mss	No	153	--
TCGA-AA-3492-01A	Tumor	Estadio ii	msi-h	No	92	0
TCGA-AA-3494-01A	Tumor	Estadio iv	mss	No	--	31
TCGA-AA-3495-01A	Tumor	Estadio i	mss	No	--	1127
TCGA-AA-3502-01A	Tumor	Estadio i	msi-l	No	--	1065
TCGA-AA-3506-01A	Tumor	Estadio i	mss	No	--	1765
TCGA-AA-3509-01A	Tumor	Estadio ii	mss	No	--	1915
TCGA-AA-3510-01A	Tumor	Estadio ii	mss	No	--	1946
TCGA-AA-3519-01A	Tumor	Estadio iii	mss	No	--	276

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-AA-3521-01A	Tumor	Estadio ii	mss	No	--	--
TCGA-AA-3524-01A	Tumor	Estadio ii	mss	No	--	1096
TCGA-AA-3526-01A	Tumor	Estadio i	msi-l	No	--	580
TCGA-AA-3529-01A	Tumor	Estadio iii	msi-l	No	0	--
TCGA-AA-3530-01A	Tumor	Estadio i	mss	No	--	580
TCGA-AA-3532-01A	Tumor	Estadio ii	mss	No	--	882
TCGA-AA-3538-01A	Tumor	Estadio i	mss	No	--	791
TCGA-AA-3542-01A	Tumor	Estadio iii	mss	No	--	395
TCGA-AA-3543-01A	Tumor	Estadio i	msi-h	No	--	30
TCGA-AA-3544-01A	Tumor	Estadio i	mss	No	--	426
TCGA-AA-3548-01A	Tumor	Estadio iii	mss	No	--	1034
TCGA-AA-3549-01A	Tumor	Estadio i	mss	No	--	639
TCGA-AA-3552-01A	Tumor	Estadio iii	mss	No	396	0
TCGA-AA-3553-01A	Tumor	Estadio i	msi-l	No	--	730
TCGA-AA-3554-01A	Tumor	Estadio ii	msi-h	No	--	546
TCGA-AA-3555-01A	Tumor	Estadio ii	mss	No	--	911
TCGA-AA-3556-01A	Tumor	Estadio i	mss	No	--	700
TCGA-AA-3560-01A	Tumor	Estadio iii	mss	No	--	608
TCGA-AA-3561-01A	Tumor	Estadio ii	mss	No	--	424
TCGA-AA-3562-01A	Tumor	Estadio iii	mss	No	--	608
TCGA-AA-3664-01A	Tumor	Estadio ii	mss	No	--	1643
TCGA-AA-3666-01A	Tumor	Estadio iii	mss	No	61	--
TCGA-AA-3667-01A	Tumor	Estadio i	msi-l	No	--	426
TCGA-AA-3672-01A	Tumor	Estadio iii	msi-h	No	--	28
TCGA-AA-3673-01A	Tumor	Estadio ii	mss	No	--	1522
TCGA-AA-3675-01A	Tumor	Estadio ii	mss	No	--	1431
TCGA-AA-3678-01A	Tumor	Estadio iii	mss	No	--	1430
TCGA-AA-3679-01A	Tumor	Estadio iv	mss	No	--	457
TCGA-AA-3680-01A	Tumor	Estadio iv	msi-l	No	335	0
TCGA-AA-3681-01A	Tumor	Estadio iii	mss	No	--	182
TCGA-AA-3684-01A	Tumor	Estadio iv	mss	No	--	0
TCGA-AA-3685-01A	Tumor	Estadio ii	mss	No	--	1127
TCGA-AA-3688-01A	Tumor	Estadio iv	msi-l	No	--	578
TCGA-AA-3692-01A	Tumor	Estadio iv	msi-l	No	1095	0
TCGA-AA-3693-01A	Tumor	Estadio iv	mss	No	--	0
TCGA-AA-3696-01A	Tumor	Estadio iv	mss	No	153	0
TCGA-AA-3710-01A	Tumor	Estadio ii	msi-h	No	--	821
TCGA-AA-3715-01A	Tumor	Estadio ii	msi-h	No	579	30

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-AA-3811-01A	Tumor	Estadio iii	msi-h	No	306	30
TCGA-AA-3812-01A	Tumor	Estadio ii	mss	No	--	1066
TCGA-AA-3814-01A	Tumor	Estadio ii	mss	No	--	0
TCGA-AA-3815-01A	Tumor	Estadio ii	msi-h	No	--	1005
TCGA-AA-3818-01A	Tumor	Estadio ii	mss	No	30	--
TCGA-AA-3819-01A	Tumor	Estadio ii	msi-l	No	--	761
TCGA-AA-3821-01A	Tumor	Estadio i	msi-h	No	--	31
TCGA-AA-3831-01A	Tumor	Estadio ii	mss	No	--	547
TCGA-AA-3833-01A	Tumor	Estadio ii	msi-h	No	--	485
TCGA-AA-3837-01A	Tumor	Estadio ii	mss	No	--	1186
TCGA-AA-3841-01A	Tumor	Estadio ii	mss	No	--	1124
TCGA-AA-3842-01A	Tumor	Estadio iii	mss	No	--	1126
TCGA-AA-3844-01A	Tumor	Estadio iii	mss	No	--	454
TCGA-AA-3845-01A	Tumor	Estadio ii	msi-h	No	0	--
TCGA-AA-3846-01A	Tumor	Estadio ii	mss	No	--	518
TCGA-AA-3848-01A	Tumor	Estadio iii	mss	No	306	30
TCGA-AA-3850-01A	Tumor	Estadio i	mss	No	0	--
TCGA-AA-3851-01A	Tumor	Estadio ii	mss	No	--	1006
TCGA-AA-3852-01A	Tumor	Estadio ii	msi-l	No	0	--
TCGA-AA-3854-01A	Tumor	Estadio i	msi-l	No	--	1096
TCGA-AA-3855-01A	Tumor	Estadio i	msi-l	No	--	975
TCGA-AA-3856-01A	Tumor	Estadio ii	mss	No	--	30
TCGA-AA-3858-01A	Tumor	Estadio i	mss	No	--	945
TCGA-AA-3860-01A	Tumor	Estadio iii	mss	No	--	945
TCGA-AA-3861-01A	Tumor	Estadio ii	msi-l	No	--	914
TCGA-AA-3862-01A	Tumor	Estadio ii	mss	No	--	914
TCGA-AA-3864-01A	Tumor	Estadio ii	msi-h	No	--	1612
TCGA-AA-3866-01A	Tumor	Estadio i	msi-l	No	--	518
TCGA-AA-3867-01A	Tumor	Estadio iv	mss	No	--	731
TCGA-AA-3869-01A	Tumor	Estadio iv	mss	No	822	0
TCGA-AA-3870-01A	Tumor	Estadio iv	mss	No	--	912
TCGA-AA-3872-01A	Tumor	Estadio iv	mss	No	--	0
TCGA-AA-3875-01A	Tumor	Estadio i	mss	No	--	549
TCGA-AA-3877-01A	Tumor	Estadio i	msi-h	No	--	943
TCGA-AA-3930-01A	Tumor	Estadio iv	msi-l	No	61	0
TCGA-AA-3939-01A	Tumor	Estadio ii	mss	No	--	395
TCGA-AA-3941-01A	Tumor	Estadio iv	msi-l	No	--	730
TCGA-AA-3947-01A	Tumor	Estadio ii	msi-h	No	--	1004

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-AA-3949-01A	Tumor	Estadio iii	msi-h	No	--	791
TCGA-AA-3950-01A	Tumor	Estadio ii	msi-h	No	--	730
TCGA-AA-3952-01A	Tumor	Estadio iii	mss	No	61	--
TCGA-AA-3955-01A	Tumor	Estadio iii	mss	No	--	638
TCGA-AA-3956-01A	Tumor	Estadio ii	mss	No	--	1035
TCGA-AA-3966-01A	Tumor	Estadio ii	msi-h	No	--	61
TCGA-AA-3968-01A	Tumor	Estadio i	mss	No	--	669
TCGA-AA-3970-01A	Tumor	Estadio ii	mss	No	--	1096
TCGA-AA-3971-01A	Tumor	Estadio iii	mss	No	--	489
TCGA-AA-3972-01A	Tumor	Estadio iv	msi-l	No	--	1551
TCGA-AA-3973-01A	Tumor	Estadio iv	msi-l	No	--	397
TCGA-AA-3975-01A	Tumor	Estadio i	mss	No	--	1036
TCGA-AA-3976-01A	Tumor	Estadio iii	mss	No	--	791
TCGA-AA-3979-01A	Tumor	Estadio ii	mss	No	--	730
TCGA-AA-3980-01A	Tumor	Estadio i	mss	No	--	822
TCGA-AA-3982-01A	Tumor	Estadio iii	msi-l	No	--	822
TCGA-AA-3984-01A	Tumor	Estadio ii	mss	No	--	0
TCGA-AA-3986-01A	Tumor	Estadio i	mss	No	--	580
TCGA-AA-3989-01A	Tumor	Estadio iv	mss	No	242	0
TCGA-AA-3994-01A	Tumor	Estadio iii	mss	No	--	822
TCGA-AA-A004-01A	Tumor	Estadio ii	msi-l	No	--	424
TCGA-AA-A00A-01A	Tumor	Estadio ii	msi-h	No	--	1157
TCGA-AA-A00D-01A	Tumor	Estadio i	mss	No	--	578
TCGA-AA-A00E-01A	Tumor	Estadio ii	msi-h	No	--	913
TCGA-AA-A00F-01A	Tumor	Estadio iii	mss	No	--	1035
TCGA-AA-A00J-01A	Tumor	Estadio iii	msi-h	No	--	549
TCGA-AA-A00K-01A	Tumor	Estadio ii	msi-l	No	--	549
TCGA-AA-A00L-01A	Tumor	Estadio ii	mss	No	--	1157
TCGA-AA-A00N-01A	Tumor	Estadio ii	msi-l	No	122	--
TCGA-AA-A00O-01A	Tumor	Estadio iii	msi-l	No	--	822
TCGA-AA-A00Q-01A	Tumor	Estadio iii	mss	No	--	1278
TCGA-AA-A00R-01A	Tumor	Estadio i	msi-h	No	--	30
TCGA-AA-A00U-01A	Tumor	Estadio iii	mss	No	--	518
TCGA-AA-A00W-01A	Tumor	Estadio i	mss	No	--	456
TCGA-AA-A00Z-01A	Tumor	Estadio ii	mss	No	--	669
TCGA-AA-A010-01A	Tumor	Estadio ii	msi-l	No	--	1064
TCGA-AA-A017-01A	Tumor	Estadio ii	mss	No	--	457
TCGA-AA-A01C-01A	Tumor	Estadio iii	mss	No	--	457

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-AA-A01D-01A	Tumor	Estadio iii	mss	No	334	30
TCGA-AA-A01F-01A	Tumor	Estadio iii	mss	No	--	974
TCGA-AA-A01G-01A	Tumor	Estadio ii	msi-l	No	--	365
TCGA-AA-A01I-01A	Tumor	Estadio i	mss	No	--	943
TCGA-AA-A01K-01A	Tumor	Estadio iii	mss	No	--	943
TCGA-AA-A01P-01A	Tumor	Estadio iii	msi-h	No	1158	0
TCGA-AA-A01Q-01A	Tumor	Estadio ii	msi-h	No	--	31
TCGA-AA-A01R-01A	Tumor	Estadio iii	msi-h	No	--	1065
TCGA-AA-A01S-01A	Tumor	Estadio iii	msi-l	No	--	31
TCGA-AA-A01T-01A	Tumor	Estadio iii	mss	No	--	1005
TCGA-AA-A01V-01A	Tumor	Estadio i	mss	No	--	31
TCGA-AA-A01X-01A	Tumor	Estadio iii	mss	No	--	791
TCGA-AA-A01Z-01A	Tumor	Estadio ii	mss	No	--	1126
TCGA-AA-A022-01A	Tumor	Estadio ii	msi-h	No	--	0
TCGA-AA-A024-01A	Tumor	Estadio ii	msi-l	No	1188	61
TCGA-AA-A029-01A	Tumor	Estadio ii	msi-l	No	--	1581
TCGA-AA-A02E-01A	Tumor	Estadio iv	msi-l	No	90	--
TCGA-AA-A02F-01A	Tumor	Estadio iv	mss	No	--	1216
TCGA-AA-A02H-01A	Tumor	Estadio iv	mss	No	61	--
TCGA-AA-A02J-01A	Tumor	Estadio iv	mss	No	153	--
TCGA-AA-A02K-01A	Tumor	Estadio iv	mss	No	426	0
TCGA-AA-A02O-01A	Tumor	Estadio ii	mss	No	--	28
TCGA-AA-A02R-01A	Tumor	Estadio ii	msi-h	No	670	--
TCGA-AA-A02W-01A	Tumor	Estadio i	msi-l	No	--	1247
TCGA-AA-A02Y-01A	Tumor	Estadio i	mss	No	--	1216
TCGA-AA-A03F-01A	Tumor	Estadio iii	mss	No	549	0
TCGA-AA-A03J-01A	Tumor	Estadio i	mss	No	--	1246
TCGA-AD-6888-01A	Tumor	Estadio iii	mss	No	472	155
TCGA-AD-6889-01A	Tumor	Estadio ii	msi-h	No	2532	2056
TCGA-AD-6895-01A	Tumor	Estadio iii	msi-h	No	--	763
TCGA-AD-6965-01A	Tumor	Estadio iii	mss	No	--	805
TCGA-AD-A5EK-01A	Tumor	Estadio i	mss	Si	--	500
TCGA-AM-5820-01A	Tumor	Estadio iv	mss	No	--	14
TCGA-AM-5821-01A	Tumor	Estadio ii	msi-h	No	--	28
TCGA-AU-3779-01A	Tumor	Estadio ii	mss	No	--	441
TCGA-AY-4070-01A	Tumor	Estadio iii	mss	No	496	--
TCGA-AY-4071-01A	Tumor	Estadio i	mss	No	29	--
TCGA-AY-6196-01A	Tumor	Estadio iii	msi-l	Si	--	6

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-AY-6197-01A	Tumor	Estadio ii	msi-h	Si	--	652
TCGA-AY-6386-01A	Tumor	Estadio iii	mss	No	--	542
TCGA-AY-A54L-01A	Tumor	Estadio i	msi-l	Si	--	525
TCGA-AY-A69D-01A	Tumor	Estadio ii	msi-l	Si	--	543
TCGA-AY-A71X-01A	Tumor	Estadio i	mss	Si	--	588
TCGA-AY-A8YK-01A	Tumor	Estadio iv	mss	Si	--	573
TCGA-AZ-4308-01A	Tumor	Estadio iii	mss	Si	--	3324
TCGA-AZ-5407-01A	Tumor	Estadio i	mss	Si	--	2683
TCGA-CA-5256-01A	Tumor	Estadio ii	mss	No	--	379
TCGA-CK-4948-01B	Tumor	Estadio iii	msi-l	No	--	4502
TCGA-CK-4952-01A	Tumor	Estadio iii	mss	No	--	475
TCGA-CK-5913-01A	Tumor	Estadio ii	msi-h	Si	--	1561
TCGA-CK-6746-01A	Tumor	Estadio ii	msi-h	Si	--	0
TCGA-CK-6748-01A	Tumor	Estadio iv	mss	No	--	61
TCGA-CM-4743-01A	Tumor	Estadio ii	msi-h	No	--	701
TCGA-CM-4744-01A	Tumor	Estadio i	mss	No	--	609
TCGA-CM-4746-01A	Tumor	Estadio i	msi-h	No	--	1126
TCGA-CM-4747-01A	Tumor	Estadio iv	msi-l	Si	--	761
TCGA-CM-4748-01A	Tumor	Estadio iii	mss	Si	--	792
TCGA-CM-4750-01A	Tumor	Estadio iii	msi-l	Si	--	244
TCGA-CM-4751-01A	Tumor	Estadio iii	mss	No	--	822
TCGA-CM-4752-01A	Tumor	Estadio ii	mss	No	--	396
TCGA-CM-5341-01A	Tumor	Estadio iii	mss	No	--	884
TCGA-CM-5344-01A	Tumor	Estadio iii	mss	Si	--	670
TCGA-CM-5348-01A	Tumor	Estadio iii	mss	No	--	699
TCGA-CM-5349-01A	Tumor	Estadio ii	mss	Si	--	915
TCGA-CM-5860-01A	Tumor	Estadio ii	mss	No	--	974
TCGA-CM-5861-01A	Tumor	Estadio ii	msi-h	No	--	457
TCGA-CM-5862-01A	Tumor	Estadio iv	mss	No	153	--
TCGA-CM-5863-01A	Tumor	Estadio iii	mss	No	--	457
TCGA-CM-5864-01A	Tumor	Estadio i	mss	Si	--	457
TCGA-CM-5868-01A	Tumor	Estadio iv	mss	No	--	518
TCGA-CM-6161-01A	Tumor	Estadio i	msi-l	Si	--	457
TCGA-CM-6162-01A	Tumor	Estadio iii	msi-h	Si	--	365
TCGA-CM-6163-01A	Tumor	Estadio i	mss	No	--	90
TCGA-CM-6164-01A	Tumor	Estadio ii	mss	Si	--	883
TCGA-CM-6165-01A	Tumor	Estadio ii	mss	No	--	488
TCGA-CM-6166-01A	Tumor	Estadio i	mss	Si	--	669

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-CM-6167-01A	Tumor	Estadio iii	mss	Si	--	456
TCGA-CM-6168-01A	Tumor	Estadio ii	msi-l	No	--	395
TCGA-CM-6169-01A	Tumor	Estadio ii	mss	No	--	396
TCGA-CM-6170-01A	Tumor	Estadio i	mss	No	--	457
TCGA-CM-6171-01A	Tumor	Estadio i	msi-h	No	--	427
TCGA-CM-6172-01A	Tumor	Estadio iii	mss	No	--	335
TCGA-CM-6674-01A	Tumor	Estadio ii	msi-h	Si	--	394
TCGA-CM-6675-01A	Tumor	Estadio iv	mss	Si	--	397
TCGA-CM-6676-01A	Tumor	Estadio i	msi-l	No	--	337
TCGA-CM-6677-01A	Tumor	Estadio ii	mss	No	--	337
TCGA-CM-6678-01A	Tumor	Estadio iv	mss	No	--	335
TCGA-CM-6679-01A	Tumor	Estadio ii	mss	Si	--	306
TCGA-CM-6680-01A	Tumor	Estadio iii	mss	No	--	366
TCGA-D5-5538-01A	Tumor	Estadio iii	mss	No	1661	373
TCGA-D5-5540-01A	Tumor	Estadio ii	msi-l	No	--	1706
TCGA-D5-5541-01A	Tumor	Estadio iii	mss	No	--	1701
TCGA-D5-6531-01A	Tumor	Estadio ii	mss	No	--	540
TCGA-D5-6538-01A	Tumor	Estadio iii	msi-l	No	--	521
TCGA-D5-6541-01A	Tumor	Estadio ii	mss	No	--	474
TCGA-D5-6898-01A	Tumor	Estadio i	mss	No	--	229
TCGA-D5-6920-01A	Tumor	Estadio ii	mss	No	--	377
TCGA-D5-6922-01A	Tumor	Estadio iii	mss	No	--	308
TCGA-D5-6923-01A	Tumor	Estadio i	mss	No	--	378
TCGA-D5-6924-01A	Tumor	Estadio ii	mss	No	--	435
TCGA-D5-6926-01A	Tumor	Estadio iii	mss	No	--	275
TCGA-D5-6927-01A	Tumor	Estadio ii	msi-h	No	--	287
TCGA-D5-6928-01A	Tumor	Estadio ii	msi-h	No	--	354
TCGA-D5-6929-01A	Tumor	Estadio iv	msi-l	No	--	408
TCGA-D5-6930-01A	Tumor	Estadio ii	msi-h	No	--	406
TCGA-D5-6931-01A	Tumor	Estadio iii	mss	No	--	365
TCGA-D5-6932-01A	Tumor	Estadio ii	mss	No	--	346
TCGA-F4-6459-01A	Tumor	Estadio iii	mss	No	262	0
TCGA-F4-6460-01A	Tumor	Estadio iii	mss	No	972	407
TCGA-F4-6461-01A	Tumor	Estadio iii	mss	No	338	0
TCGA-F4-6805-01A	Tumor	Estadio ii	mss	No	--	1047
TCGA-F4-6806-01A	Tumor	Estadio i	mss	No	--	1260
TCGA-F4-6807-01A	Tumor	Estadio iii	mss	No	--	1309
TCGA-F4-6808-01A	Tumor	Estadio i	mss	No	--	1024

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-F4-6809-01A	Tumor	Estadio iv	mss	No	403	39
TCGA-F4-6854-01A	Tumor	Estadio ii	mss	No	--	16
TCGA-G4-6293-01A	Tumor	Estadio iii	msi-l	Si	--	4051
TCGA-G4-6294-01A	Tumor	Estadio iv	mss	No	858	--
TCGA-G4-6295-01A	Tumor	Estadio ii	msi-l	No	--	254
TCGA-G4-6297-01A	Tumor	Estadio iv	mss	No	--	2506
TCGA-G4-6298-01A	Tumor	Estadio iii	msi-l	No	715	--
TCGA-G4-6299-01A	Tumor	Estadio iii	mss	No	--	2268
TCGA-G4-6302-01A	Tumor	Estadio ii	msi-h	No	2047	1310
TCGA-G4-6303-01A	Tumor	Estadio iv	msi-l	No	2003	1882
TCGA-G4-6304-01A	Tumor	Estadio ii	msi-h	No	--	1631
TCGA-G4-6309-01A	Tumor	Estadio iii	msi-h	Si	--	2600
TCGA-G4-6310-01A	Tumor	Estadio iii	mss	Si	--	1935
TCGA-G4-6311-01A	Tumor	Estadio iii	mss	Si	--	1199
TCGA-G4-6314-01A	Tumor	Estadio iv	mss	No	--	1093
TCGA-G4-6315-01A	Tumor	Estadio iv	mss	No	--	1883
TCGA-G4-6317-01A	Tumor	Estadio iii	msi-l	Si	--	1095
TCGA-G4-6320-01A	Tumor	Estadio iii	msi-h	No	--	804
TCGA-G4-6321-01A	Tumor	Estadio iii	msi-l	Si	--	672
TCGA-G4-6322-01A	Tumor	Estadio iii	msi-l	No	--	792
TCGA-G4-6323-01A	Tumor	Estadio i	mss	No	--	419
TCGA-G4-6586-01A	Tumor	Estadio ii	msi-h	No	--	1089
TCGA-G4-6588-01A	Tumor	Estadio ii	msi-h	Si	--	796
TCGA-G4-6625-01A	Tumor	Estadio ii	mss	No	--	2792
TCGA-G4-6626-01A	Tumor	Estadio ii	mss	No	1422	--
TCGA-G4-6627-01A	Tumor	Estadio ii	msi-l	No	--	2275
TCGA-G4-6628-01A	Tumor	Estadio i	msi-h	No	--	2424
TCGA-NH-A50V-01A	Tumor	Estadio iii	msi-l	No	--	588
TCGA-NH-A6GB-01A	Tumor	Estadio iii	mss	No	--	476
TCGA-NH-A8F8-01A	Tumor	Estadio iv	mss	No	511	167
TCGA-QG-A5YV-01A	Tumor	Estadio iii	mss	No	--	1301
TCGA-QG-A5YW-01A	Tumor	Estadio iii	msi-l	Si	--	896
TCGA-QG-A5YX-01A	Tumor	Estadio ii	msi-l	Si	--	1003
TCGA-QG-A5Z1-01A	Tumor	Estadio iii	msi-l	No	256	202
TCGA-QG-A5Z2-01A	Tumor	Estadio i	msi-h	No	--	952
TCGA-RU-A8FL-01A	Tumor	Estadio iii	mss	No	--	1177
TCGA-SS-A7HO-01A	Tumor	Estadio ii	mss	No	--	1829
TCGA-T9-A92H-01A	Tumor	Estadio ii	mss	No	--	362

ID Paciente	Grupo	Estadio	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-WS-AB45-01A	Tumor	Estadio ii	msi-h	Si	--	2130
TCGA-A6-2671-11A	Control					
TCGA-A6-2675-11A	Control					
TCGA-A6-2678-11A	Control					
TCGA-A6-2679-11A	Control					
TCGA-A6-2680-11A	Control					
TCGA-A6-2682-11A	Control					
TCGA-A6-2683-11A	Control					
TCGA-A6-2684-11A	Control					
TCGA-A6-2685-11A	Control					
TCGA-A6-2686-11A	Control					
TCGA-A6-5659-11A	Control					
TCGA-A6-5662-11A	Control					
TCGA-A6-5665-11A	Control					
TCGA-A6-5667-11A	Control					
TCGA-AA-3489-11A	Control					
TCGA-AA-3496-11A	Control					
TCGA-AA-3511-11A	Control					
TCGA-AA-3514-11A	Control					
TCGA-AA-3516-11A	Control					
TCGA-AA-3517-11A	Control					
TCGA-AA-3518-11A	Control					
TCGA-AA-3520-11A	Control					
TCGA-AA-3522-11A	Control					
TCGA-AA-3525-11A	Control					
TCGA-AA-3527-11A	Control					
TCGA-AA-3531-11A	Control					
TCGA-AA-3534-11A	Control					
TCGA-AA-3655-11A	Control					
TCGA-AA-3660-11A	Control					
TCGA-AA-3662-11A	Control					
TCGA-AA-3663-11A	Control					
TCGA-AA-3697-11A	Control					
TCGA-AA-3712-11A	Control					
TCGA-AA-3713-11A	Control					
TCGA-AZ-6598-11A	Control					
TCGA-AZ-6599-11A	Control					
TCGA-AZ-6600-11A	Control					

ID Paciente	Grupo	Estadío	MSI	dMMR	Días hasta la muerte	Días hasta el último seguimiento
TCGA-AZ-6601-11A	Control					
TCGA-AZ-6603-11A	Control					
TCGA-AZ-6605-11A	Control					
TCGA-F4-6704-11A	Control					

Anexo B: Listado de glicogenes en modelos experimentales murinos y pacientes de CCR.

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>4930568D16Rik</i>	
Glycosyltransferase	<i>4933406M09Rik</i>	
Glycosylhydrolase	<i>9530003J23Rik</i>	
Glycosyltransferase	<i>A3galt2</i>	<i>A3GALT2</i>
Glycosyltransferase	<i>A4galt</i>	<i>A4GALT</i>
Glycosyltransferase	<i>A4gnt</i>	<i>A4GNT</i>
Carbohydrate Esterase	<i>Aadac</i>	<i>AADAC</i>
Housekeeping Gene	<i>Abl1</i>	<i>ABL1</i>
Glycosyltransferase	<i>Abo</i>	<i>ABO</i>
Lectin/C-type/Proteoglycan	<i>Acan</i>	<i>ACAN</i>
Glycolipid Related	<i>Acer1</i>	<i>ACER1</i>
Carbohydrate Esterase	<i>Ache</i>	<i>ACHE</i>
Housekeeping Gene	<i>Actb</i>	<i>ACTB</i>
Housekeeping Gene	<i>Adarb2</i>	<i>ADARB2</i>
Housekeeping Gene	<i>Afap1</i>	<i>AFAP1</i>
Glycolipid Related	<i>Aga</i>	<i>AGA</i>
Glycosylhydrolase	<i>Agl</i>	<i>AGL</i>
Lectin/Misc. Glycan-binding/Agrin	<i>Agrn</i>	<i>AGRN</i>
Glycosylhydrolase	<i>AI464131</i>	
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Aldob</i>	<i>ALDOB</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Aldoc</i>	<i>ALDOC</i>
Glycosyltransferase	<i>Alg1</i>	<i>ALG1</i>
Glycosyltransferase	<i>Alg10b</i>	<i>ALG10B</i>
Glycosyltransferase	<i>Alg11</i>	<i>ALG11</i>
Glycosyltransferase	<i>Alg12</i>	<i>ALG12</i>
Glycosyltransferase	<i>Alg13</i>	<i>ALG13</i>
Glycosyltransferase	<i>Alg14</i>	<i>ALG14</i>
Glycosyltransferase	<i>Alg2</i>	<i>ALG2</i>
Glycosyltransferase	<i>Alg3</i>	<i>ALG3</i>
Glycosyltransferase	<i>Alg5</i>	<i>ALG5</i>
Glycosyltransferase	<i>Alg6</i>	<i>ALG6</i>
Glycosyltransferase	<i>Alg8</i>	<i>ALG8</i>
Glycosyltransferase	<i>Alg9</i>	<i>ALG9</i>
Carbohydrate Esterase	<i>Amdhd2</i>	<i>AMDHD2</i>
Glycosylhydrolase	<i>Amy1</i>	
Glycosylhydrolase	<i>Amy2a5</i>	

Grupo	Gen MM	Gen HS
Lectin/Misc. Glycan-binding/Annexin	<i>Anxa4</i>	<i>ANXA4</i>
Lectin/Misc. Glycan-binding/Annexin	<i>Anxa5</i>	<i>ANXA5</i>
Lectin/Misc. Glycan-binding/Annexin	<i>Anxa7</i>	<i>ANXA7</i>
Lectin/Misc. Glycan-binding/Annexin	<i>Anxa9</i>	<i>ANXA9</i>
Glycolipid Related	<i>Aoah</i>	<i>AOAH</i>
Carbohydrate Esterase	<i>Apeh</i>	<i>APEH</i>
Sulfate Related/Arylsulfatase	<i>Arsa</i>	<i>ARSA</i>
Sulfate Related/Arylsulfatase	<i>Arsb</i>	<i>ARSB</i>
Sulfate Related/Arylsulfatase	<i>Arsg</i>	<i>ARSG</i>
Sulfate Related/Arylsulfatase	<i>Arsi</i>	<i>ARSI</i>
Glycolipid Related	<i>Asah2</i>	<i>ASAH2</i>
Lectin/C-type/Type 2 Receptor	<i>Asgr1</i>	<i>ASGR1</i>
Lectin/C-type/Type 2 Receptor	<i>Asgr2</i>	<i>ASGR2</i>
Housekeeping Gene	<i>Atp5b</i>	
Lectin/C-type/Attractin	<i>Atrn</i>	<i>ATRNL1</i>
Lectin/C-type/Attractin	<i>Atrnl1</i>	<i>ATRNL1</i>
Housekeeping Gene	<i>B2m</i>	<i>B2M</i>
Glycosyltransferase	<i>B3galnt1</i>	<i>B3GALNT1</i>
Glycosyltransferase	<i>B3galnt2</i>	<i>B3GALNT2</i>
Glycosyltransferase	<i>B3galt1</i>	<i>B3GALT1</i>
Glycosyltransferase	<i>B3galt2</i>	<i>B3GALT2</i>
Glycosyltransferase	<i>B3galt4</i>	<i>B3GALT4</i>
Glycosyltransferase	<i>B3galt5</i>	<i>B3GALT5</i>
Glycosyltransferase	<i>B3galt6</i>	<i>B3GALT6</i>
Glycosyltransferase	<i>B3gat1</i>	<i>B3GAT1</i>
Glycosyltransferase	<i>B3gat2</i>	<i>B3GAT2</i>
Glycosyltransferase	<i>B3gat3</i>	<i>B3GAT3</i>
Glycosyltransferase	<i>B3glct</i>	<i>B3GLCT</i>
Glycosyltransferase	<i>B3gnt2</i>	<i>B3GNT2</i>
Glycosyltransferase	<i>B3gnt3</i>	<i>B3GNT3</i>
Glycosyltransferase	<i>B3gnt4</i>	<i>B3GNT4</i>
Glycosyltransferase	<i>B3gnt5</i>	<i>B3GNT5</i>
Glycosyltransferase	<i>B3gnt6</i>	<i>B3GNT6</i>
Glycosyltransferase	<i>B3gnt7</i>	<i>B3GNT7</i>
Glycosyltransferase	<i>B3gnt8</i>	<i>B3GNT8</i>
Glycosyltransferase	<i>B3gnt9</i>	<i>B3GNT9</i>
Glycosyltransferase	<i>B3gntl1</i>	<i>B3GNTL1</i>
Glycosyltransferase	<i>B4galnt1</i>	<i>B4GALNT1</i>
Glycosyltransferase	<i>B4galnt2</i>	<i>B4GALNT2</i>

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>B4galnt3</i>	<i>B4GALNT3</i>
Glycosyltransferase	<i>B4galnt4</i>	<i>B4GALNT4</i>
Glycosyltransferase	<i>B4galt1</i>	<i>B4GALT1</i>
Glycosyltransferase	<i>B4galt2</i>	<i>B4GALT2</i>
Glycosyltransferase	<i>B4galt3</i>	<i>B4GALT3</i>
Glycosyltransferase	<i>B4galt4</i>	<i>B4GALT4</i>
Glycosyltransferase	<i>B4galt5</i>	<i>B4GALT5</i>
Glycosyltransferase	<i>B4galt6</i>	<i>B4GALT6</i>
Glycosyltransferase	<i>B4galt7</i>	<i>B4GALT7</i>
Glycosyltransferase	<i>B4gat1</i>	<i>B4GATI</i>
Lectin/C-type/Proteoglycan	<i>Bcan</i>	<i>BCAN</i>
Carbohydrate Esterase	<i>Bche</i>	<i>BCHE</i>
Glycosaminoglycan Related	<i>Bgn</i>	<i>BGN</i>
Glycosyltransferase	<i>C1galt1</i>	<i>C1GALT1</i>
Glycosyltransferase	<i>C1galt1c1</i>	<i>C1GALT1C1</i>
Lectin/Calnexin-like	<i>Calr</i>	<i>CALR</i>
Lectin/Calnexin-like	<i>Canx</i>	<i>CANX</i>
Lectin/Misc. Glycan-binding/MHC	<i>Cd1d1</i>	
Lectin/C-type/Type 2 Receptor	<i>Cd207</i>	<i>CD207</i>
Lectin/C-type/Type 2 Receptor	<i>Cd209a</i>	
Lectin/C-type/Type 2 Receptor	<i>Cd209b</i>	
Lectin/C-type/Type 2 Receptor	<i>Cd209c</i>	
Lectin/C-type/Type 2 Receptor	<i>Cd209d</i>	
Lectin/C-type/Type 2 Receptor	<i>Cd209e</i>	
Lectin/I-type/Siglec	<i>Cd22</i>	<i>CD22</i>
Lectin/C-type/Endosialin	<i>Cd248</i>	<i>CD248</i>
Lectin/C-type/Acidic Neck CTLD	<i>Cd302</i>	<i>CD302</i>
Lectin/I-type/CD33-Related	<i>Cd33</i>	<i>CD33</i>
Glycosaminoglycan Related	<i>Cd34</i>	<i>CD34</i>
Glycosaminoglycan Related	<i>Cd44</i>	<i>CD44</i>
Lectin/C-type/NK Receptor	<i>Cd69</i>	<i>CD69</i>
Lectin/C-type/NK Receptor	<i>Cd72</i>	<i>CD72</i>
Glycosaminoglycan Related	<i>Cd74</i>	<i>CD74</i>
Lectin/I-type/Non-Siglec	<i>Cd83</i>	<i>CD83</i>
Lectin/C-type/Endosialin	<i>Cd93</i>	<i>CD93</i>
Carbohydrate Esterase	<i>Cel</i>	<i>CEL</i>
Glycosyltransferase	<i>Cercam</i>	<i>CERCAM</i>
Glycolipid Related	<i>Cerk</i>	<i>CERK</i>
Glycolipid Related	<i>Cers1</i>	<i>CERS1</i>

Grupo	Gen MM	Gen HS
Glycolipid Related	<i>Cers2</i>	<i>CERS2</i>
Glycolipid Related	<i>Cers4</i>	<i>CERS4</i>
Glycolipid Related	<i>Cers5</i>	<i>CERS5</i>
Glycolipid Related	<i>Cers6</i>	<i>CERS6</i>
Carbohydrate Esterase	<i>Ces1d</i>	
Carbohydrate Esterase	<i>Ces1f</i>	
Carbohydrate Esterase	<i>Ces1g</i>	
Carbohydrate Esterase	<i>Ces2a</i>	
Carbohydrate Esterase	<i>Ces2c</i>	
Carbohydrate Esterase	<i>Ces2g</i>	
Carbohydrate Esterase	<i>Ces5a</i>	<i>CESSA</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Chga</i>	<i>CHGA</i>
Glycosylhydrolase	<i>Chia1</i>	
Glycosylhydrolase	<i>Chil1</i>	
Glycosylhydrolase	<i>Chil3</i>	
Glycosylhydrolase	<i>Chil4</i>	
Glycosylhydrolase	<i>Chil6</i>	
Lectin/C-type/Type 1 Receptor	<i>Chodl</i>	<i>CHODL</i>
Glycosyltransferase	<i>Chpf</i>	<i>CHPF</i>
Glycosyltransferase	<i>Chpf2</i>	<i>CHPF2</i>
Sulfate Related/Sulfotransferase	<i>Chst1</i>	<i>CHST1</i>
Sulfate Related/Sulfotransferase	<i>Chst10</i>	<i>CHST10</i>
Sulfate Related/Sulfotransferase	<i>Chst11</i>	<i>CHST11</i>
Sulfate Related/Sulfotransferase	<i>Chst12</i>	<i>CHST12</i>
Sulfate Related/Sulfotransferase	<i>Chst14</i>	<i>CHST14</i>
Sulfate Related/Sulfotransferase	<i>Chst15</i>	<i>CHST15</i>
Sulfate Related/Sulfotransferase	<i>Chst2</i>	<i>CHST2</i>
Sulfate Related/Sulfotransferase	<i>Chst3</i>	<i>CHST3</i>
Sulfate Related/Sulfotransferase	<i>Chst4</i>	<i>CHST4</i>
Sulfate Related/Sulfotransferase	<i>Chst5</i>	<i>CHST5</i>
Sulfate Related/Sulfotransferase	<i>Chst7</i>	<i>CHST7</i>
Sulfate Related/Sulfotransferase	<i>Chst8</i>	<i>CHST8</i>
Sulfate Related/Sulfotransferase	<i>Chst9</i>	<i>CHST9</i>
Glycosyltransferase	<i>Chsy1</i>	<i>CHSY1</i>
Glycosyltransferase	<i>Chsy3</i>	<i>CHSY3</i>
Lectin/C-type/Type 2 Receptor	<i>Clec10a</i>	<i>CLEC10A</i>
Lectin/C-type/Tetranectin	<i>Clec11a</i>	<i>CLEC11A</i>
Lectin/C-type/Endosialin	<i>Clec14a</i>	<i>CLEC14A</i>
Lectin/C-type/Macrophage Mannose Receptor	<i>Clec18a</i>	<i>CLEC18A</i>

Grupo	Gen MM	Gen HS
Lectin/C-type/NK Receptor	<i>Clec1a</i>	<i>CLEC1A</i>
Lectin/C-type/NK Receptor	<i>Clec1b</i>	<i>CLEC1B</i>
Lectin/C-type/NK Receptor	<i>Clec2d</i>	<i>CLEC2D</i>
Lectin/C-type/NK Receptor	<i>Clec2h</i>	
Lectin/C-type/NK Receptor	<i>Clec2i</i>	
Lectin/C-type/Tetranectin	<i>Clec3a</i>	<i>CLEC3A</i>
Lectin/C-type/Tetranectin	<i>Clec3b</i>	<i>CLEC3B</i>
Lectin/C-type/Type 2 Receptor	<i>Clec4a2</i>	
Lectin/C-type/Type 2 Receptor	<i>Clec4a3</i>	
Lectin/C-type/Type 2 Receptor	<i>Clec4b1</i>	
Lectin/C-type/Type 2 Receptor	<i>Clec4d</i>	<i>CLEC4D</i>
Lectin/C-type/Type 2 Receptor	<i>Clec4e</i>	<i>CLEC4E</i>
Lectin/C-type/Type 2 Receptor	<i>Clec4f</i>	<i>CLEC4F</i>
Lectin/C-type/Type 2 Receptor	<i>Clec4g</i>	<i>CLEC4G</i>
Lectin/C-type/Type 2 Receptor	<i>Clec4n</i>	
Lectin/C-type/NK Receptor	<i>Clec5a</i>	<i>CLEC5A</i>
Lectin/C-type/NK Receptor	<i>Clec7a</i>	<i>CLEC7A</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Cmah</i>	
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Cmas</i>	<i>CMAS</i>
Glycosaminoglycan Related	<i>Col14a1</i>	<i>COL14A1</i>
Glycolipid Related	<i>Col4a3bp</i>	<i>COL4A3BP</i>
Glycosaminoglycan Related	<i>Col9a2</i>	<i>COL9A2</i>
Lectin/C-type/Collectin	<i>Colec10</i>	<i>COLEC10</i>
Lectin/C-type/Type 2 Receptor	<i>Colec12</i>	<i>COLEC12</i>
Glycosyltransferase	<i>Colgalt1</i>	<i>COLGALT1</i>
Glycosyltransferase	<i>Colgalt2</i>	<i>COLGALT2</i>
Carbohydrate Binding Module	<i>Crybg1</i>	<i>CRYBG1</i>
Carbohydrate Binding Module	<i>Crybg2</i>	<i>CRYBG2</i>
Carbohydrate Binding Module	<i>Crybg3</i>	<i>CRYBG3</i>
Glycosyltransferase	<i>Csgalnact1</i>	<i>CSGALNACT1</i>
Glycosyltransferase	<i>Csgalnact2</i>	<i>CSGALNACT2</i>
Glycosaminoglycan Related	<i>Cspg4</i>	<i>CSPG4</i>
Glycosaminoglycan Related	<i>Cspg5</i>	<i>CSPG5</i>
Glycosylhydrolase	<i>Ctbs</i>	<i>CTBS</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Ctns</i>	<i>CTNS</i>
Miscellaneous	<i>Ctsa</i>	<i>CTSA</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Cyb5r3</i>	<i>CYB5R3</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Dad1</i>	<i>DAD1</i>
Glycosaminoglycan Related	<i>Dag1</i>	<i>DAG1</i>

Grupo	Gen MM	Gen HS
Glycosaminoglycan Related	<i>Dcn</i>	<i>DCN</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Ddost</i>	<i>DDOST</i>
Glycolipid Related	<i>Degs1</i>	<i>DEGS1</i>
Glycolipid Related	<i>Degs2</i>	<i>DEGS2</i>
Lectin/C-type/IDD	<i>Dgcr2</i>	<i>DGCR2</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Dolk</i>	<i>DOLK</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Dpagt1</i>	<i>DPAGT1</i>
Glycosyltransferase	<i>Dpm1</i>	<i>DPM1</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Dpm2</i>	<i>DPM2</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Dpm3</i>	<i>DPM3</i>
Carbohydrate Esterase	<i>Dpp10</i>	<i>DPP10</i>
Carbohydrate Esterase	<i>Dpp4</i>	<i>DPP4</i>
Glycosaminoglycan Related	<i>Dse</i>	<i>DSE</i>
Glycosaminoglycan Related	<i>Dsel</i>	<i>DSEL</i>
Glycosylhydrolase	<i>Edem1</i>	<i>EDEM1</i>
Glycosylhydrolase	<i>Edem2</i>	<i>EDEM2</i>
Glycosylhydrolase	<i>Edem3</i>	<i>EDEM3</i>
Housekeeping Gene	<i>Eef1g</i>	<i>EEF1G</i>
Glycolipid Related	<i>Elovl1</i>	<i>ELOVL1</i>
Glycolipid Related	<i>Elovl2</i>	<i>ELOVL2</i>
Glycolipid Related	<i>Elovl3</i>	<i>ELOVL3</i>
Glycolipid Related	<i>Elovl4</i>	<i>ELOVL4</i>
Glycolipid Related	<i>Elovl5</i>	<i>ELOVL5</i>
Glycolipid Related	<i>Elovl6</i>	<i>ELOVL6</i>
Glycosylhydrolase	<i>Engase</i>	<i>ENGASE</i>
Glycosyltransferase	<i>Eogt</i>	<i>EOGT</i>
Carbohydrate Binding Module	<i>Epm2a</i>	<i>EPM2A</i>
Carbohydrate Esterase	<i>Esd</i>	<i>ESD</i>
Glycosyltransferase	<i>Ext1</i>	<i>EXT1</i>
Glycosyltransferase	<i>Ext2</i>	<i>EXT2</i>
Glycosyltransferase	<i>Extl1</i>	<i>EXTL1</i>
Glycosyltransferase	<i>Extl2</i>	<i>EXTL2</i>
Glycosyltransferase	<i>Extl3</i>	<i>EXTL3</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Fbp1</i>	<i>FBP1</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Fbp2</i>	<i>FBP2</i>
Lectin/F-box	<i>Fbxo2</i>	<i>FBXO2</i>
Lectin/C-type/Type 2 Receptor	<i>Fcer2a</i>	
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf1</i>	<i>FGF1</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf10</i>	<i>FGF10</i>

Grupo	Gen MM	Gen HS
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf11</i>	<i>FGF11</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf12</i>	<i>FGF12</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf13</i>	<i>FGF13</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf14</i>	<i>FGF14</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf15</i>	
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf16</i>	<i>FGF16</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf17</i>	<i>FGF17</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf18</i>	<i>FGF18</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf2</i>	<i>FGF2</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf20</i>	<i>FGF20</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf21</i>	<i>FGF21</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf23</i>	<i>FGF23</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf3</i>	<i>FGF3</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf4</i>	<i>FGF4</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf5</i>	<i>FGF5</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf7</i>	<i>FGF7</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf8</i>	<i>FGF8</i>
Glycosaminoglycan Related/Fibroblast Growth Factor	<i>Fgf9</i>	<i>FGF9</i>
Glycosaminoglycan Related/Fibroblast Growth Factor Receptor	<i>Fgfr1</i>	<i>FGFR1</i>
Glycosaminoglycan Related/Fibroblast Growth Factor Receptor	<i>Fgfr2</i>	<i>FGFR2</i>
Glycosaminoglycan Related/Fibroblast Growth Factor Receptor	<i>Fgfr3</i>	<i>FGFR3</i>
Glycosaminoglycan Related/Fibroblast Growth Factor Receptor	<i>Fgfr4</i>	<i>FGFR4</i>
Glycosaminoglycan Related	<i>Fkrp</i>	<i>FKRP</i>
Miscellaneous	<i>Fktn</i>	<i>FKTN</i>
Glycosaminoglycan Related	<i>Fmod</i>	<i>FMOD</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Fpgt</i>	<i>FPGT</i>
Glycosylhydrolase	<i>Fuca1</i>	<i>FUCA1</i>
Glycosylhydrolase	<i>Fuca2</i>	<i>FUCA2</i>
Glycosyltransferase	<i>Fut1</i>	<i>FUT1</i>
Glycosyltransferase	<i>Fut10</i>	<i>FUT10</i>
Glycosyltransferase	<i>Fut11</i>	<i>FUT11</i>
Glycosyltransferase	<i>Fut2</i>	<i>FUT2</i>
Glycosyltransferase	<i>Fut4</i>	<i>FUT4</i>
Glycosyltransferase	<i>Fut7</i>	<i>FUT7</i>
Glycosyltransferase	<i>Fut8</i>	<i>FUT8</i>
Glycosyltransferase	<i>Fut9</i>	<i>FUT9</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>G6pc</i>	<i>G6PC</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>G6pc2</i>	<i>G6PC2</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>G6pd2</i>	<i>G6PD2</i>

Grupo	Gen MM	Gen HS
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>G6pdx</i>	
Glycosylhydrolase	<i>Gaa</i>	<i>GAA</i>
Sulfate Related/Sulfotransferase	<i>Gal3st1</i>	<i>GAL3ST1</i>
Sulfate Related/Sulfotransferase	<i>Gal3st2</i>	<i>GAL3ST2</i>
Sulfate Related/Sulfotransferase	<i>Gal3st3</i>	<i>GAL3ST3</i>
Sulfate Related/Sulfotransferase	<i>Gal3st4</i>	<i>GAL3ST4</i>
Glycosylhydrolase	<i>Galc</i>	<i>GALC</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gale</i>	<i>GALE</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Galk1</i>	<i>GALK1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Galk2</i>	<i>GALK2</i>
Sulfate Related/Sulfatase	<i>Galns</i>	<i>GALNS</i>
Glycosyltransferase	<i>Galnt1</i>	<i>GALNT1</i>
Glycosyltransferase	<i>Galnt10</i>	<i>GALNT10</i>
Glycosyltransferase	<i>Galnt11</i>	<i>GALNT11</i>
Glycosyltransferase	<i>Galnt12</i>	<i>GALNT12</i>
Glycosyltransferase	<i>Galnt13</i>	<i>GALNT13</i>
Glycosyltransferase	<i>Galnt14</i>	<i>GALNT14</i>
Glycosyltransferase	<i>Galnt15</i>	<i>GALNT15</i>
Glycosyltransferase	<i>Galnt17</i>	<i>GALNT17</i>
Glycosyltransferase	<i>Galnt18</i>	<i>GALNT18</i>
Glycosyltransferase	<i>Galnt2</i>	<i>GALNT2</i>
Glycosyltransferase	<i>Galnt3</i>	<i>GALNT3</i>
Glycosyltransferase	<i>Galnt4</i>	<i>GALNT4</i>
Glycosyltransferase	<i>Galnt5</i>	<i>GALNT5</i>
Glycosyltransferase	<i>Galnt6</i>	<i>GALNT6</i>
Glycosyltransferase	<i>Galnt7</i>	<i>GALNT7</i>
Glycosyltransferase	<i>Galnt9</i>	<i>GALNT9</i>
Glycosyltransferase	<i>Galntl5</i>	<i>GALNTL5</i>
Glycosyltransferase	<i>Galntl6</i>	<i>GALNTL6</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Galt</i>	<i>GALT</i>
Glycosylhydrolase	<i>Ganab</i>	<i>GANAB</i>
Glycosylhydrolase	<i>Ganc</i>	<i>GANC</i>
Housekeeping Gene	<i>Gapdh</i>	<i>GAPDH</i>
Glycosylhydrolase	<i>Gba</i>	<i>GBA</i>
Glycosylhydrolase	<i>Gbe1</i>	<i>GBE1</i>
Glycosyltransferase	<i>Gbgt1</i>	<i>GBGT1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gck</i>	<i>GCK</i>
Glycosyltransferase	<i>Gcnt1</i>	<i>GCNT1</i>
Glycosyltransferase	<i>Gcnt2</i>	<i>GCNT2</i>

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>Gcnt3</i>	<i>GCNT3</i>
Glycosyltransferase	<i>Gcnt7</i>	<i>GCNT7</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gfpt1</i>	<i>GFPT1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gfpt2</i>	<i>GFPT2</i>
Glycosyltransferase	<i>Ggta1</i>	
Glycosylhydrolase	<i>Gla</i>	<i>GLA</i>
Glycosylhydrolase	<i>Glb1</i>	<i>GLB1</i>
Glycosylhydrolase	<i>Glb1l</i>	<i>GLB1L</i>
Glycosylhydrolase	<i>Glb1l2</i>	<i>GLB1L2</i>
Glycosylhydrolase	<i>Glb1l3</i>	<i>GLB1L3</i>
Glycosaminoglycan Related	<i>Glce</i>	<i>GLCE</i>
Glycosyltransferase	<i>Glt28d2</i>	
Glycosyltransferase	<i>Glt6d1</i>	<i>GLT6D1</i>
Glycosyltransferase	<i>Glt8d1</i>	<i>GLT8D1</i>
Glycosyltransferase	<i>Glt8d2</i>	<i>GLT8D2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Gltg</i>	<i>GLTG</i>
Glycosaminoglycan Related	<i>Glycam1</i>	<i>GLYCAM1</i>
Glycolipid Related	<i>Gm2a</i>	<i>GM2A</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gmds</i>	<i>GMDS</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gmppa</i>	<i>GMPPA</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gmppb</i>	<i>GMPPB</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gnb1</i>	<i>GNB1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gne</i>	<i>GNE</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gnpda1</i>	<i>GNPDA1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gnpnat1</i>	<i>GNPNAT1</i>
Miscellaneous	<i>Gnptab</i>	<i>GNPTAB</i>
Lectin/P-Type	<i>Gnptg</i>	<i>GNPTG</i>
GPI anchor biosynthesis Related	<i>Gpaa1</i>	<i>GPAA1</i>
Glycosaminoglycan Related	<i>Gpc1</i>	<i>GPC1</i>
Glycosaminoglycan Related	<i>Gpc2</i>	<i>GPC2</i>
Glycosaminoglycan Related	<i>Gpc3</i>	<i>GPC3</i>
Glycosaminoglycan Related	<i>Gpc4</i>	<i>GPC4</i>
Glycosaminoglycan Related	<i>Gpc5</i>	<i>GPC5</i>
Glycosaminoglycan Related	<i>Gpc6</i>	<i>GPC6</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Gpi1</i>	
GPI anchor biosynthesis Related	<i>Gpld1</i>	<i>GPLD1</i>
Lectin/Misc. Glycan-binding/GAG-binding	<i>Gpnmb</i>	<i>GNPMB</i>
Lectin/S-Type/Galectin	<i>Grifin</i>	<i>GRIFIN</i>
Glycosyltransferase	<i>Gtdc1</i>	<i>GTDC1</i>

Grupo	Gen MM	Gen HS
Glycosylhydrolase	<i>Gusb</i>	<i>GUSB</i>
Glycosyltransferase	<i>Gxylt2</i>	<i>GXYLT2</i>
Glycosyltransferase	<i>Gyg</i>	
Glycosyltransferase	<i>Gys1</i>	<i>GYS1</i>
Glycosyltransferase	<i>Gys2</i>	<i>GYS2</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>H6pd</i>	<i>H6PD</i>
Lectin/Misc. Glycan-binding/GAG-binding	<i>Habp4</i>	<i>HABP4</i>
Glycosyltransferase	<i>Has1</i>	<i>HAS1</i>
Glycosyltransferase	<i>Has2</i>	<i>HAS2</i>
Glycosyltransferase	<i>Has3</i>	<i>HAS3</i>
Glycosylhydrolase	<i>Hexa</i>	<i>HEXA</i>
Glycosylhydrolase	<i>Hexb</i>	<i>HEXB</i>
Glycosylhydrolase	<i>Hexdc</i>	
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Hk1</i>	<i>HK1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Hk2</i>	<i>HK2</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Hk3</i>	<i>HK3</i>
Housekeeping Gene	<i>Hmbs</i>	<i>HMBS</i>
Lectin/Misc. Glycan-binding/GAG-binding	<i>Hmmr</i>	<i>HMMR</i>
Housekeeping Gene	<i>Hprt</i>	
Glycosylhydrolase	<i>Hpse</i>	<i>HPSE</i>
Glycosylhydrolase	<i>Hpse2</i>	<i>HPSE2</i>
Sulfate Related/Sulfotransferase	<i>Hs2st1</i>	<i>HS2ST1</i>
Sulfate Related/Sulfotransferase	<i>Hs3st1</i>	<i>HS3ST1</i>
Sulfate Related/Sulfotransferase	<i>Hs3st2</i>	<i>HS3ST2</i>
Sulfate Related/Sulfotransferase	<i>Hs3st3a1</i>	<i>HS3ST3A1</i>
Sulfate Related/Sulfotransferase	<i>Hs3st3b1</i>	<i>HS3ST3B1</i>
Sulfate Related/Sulfotransferase	<i>Hs3st5</i>	<i>HS3ST5</i>
Sulfate Related/Sulfotransferase	<i>Hs3st6</i>	<i>HS3ST6</i>
Sulfate Related/Sulfotransferase	<i>Hs6st1</i>	<i>HS6ST1</i>
Sulfate Related/Sulfotransferase	<i>Hs6st2</i>	<i>HS6ST2</i>
Sulfate Related/Sulfotransferase	<i>Hs6st3</i>	<i>HS6ST3</i>
Glycosaminoglycan Related	<i>Hspg2</i>	<i>HSPG2</i>
Glycosylhydrolase	<i>Hyal1</i>	<i>HYAL1</i>
Glycosylhydrolase	<i>Hyal2</i>	<i>HYAL2</i>
Glycosylhydrolase	<i>Hyal3</i>	<i>HYAL3</i>
Glycosylhydrolase	<i>Hyal4</i>	<i>HYAL4</i>
Glycosylhydrolase	<i>Hyal5</i>	
Glycosylhydrolase	<i>Hyal6</i>	
Lectin/I-type/Non-Siglec	<i>Icam1</i>	<i>ICAM1</i>

Grupo	Gen MM	Gen HS
Lectin/I-type/Non-Siglec	<i>Icam2</i>	<i>ICAM2</i>
Lectin/I-type/Non-Siglec	<i>Icam5</i>	<i>ICAM5</i>
Sulfate Related/Sulfatase	<i>Ids</i>	<i>IDS</i>
Glycosylhydrolase	<i>Idua</i>	<i>IDUA</i>
Glycosyltransferase	<i>Itga3</i>	<i>ITGA3</i>
Housekeeping Gene	<i>Junb</i>	<i>JUNB</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Khk</i>	<i>KHK</i>
Glycosylhydrolase	<i>Kl</i>	<i>KL</i>
Glycosylhydrolase	<i>Klb</i>	<i>KLB</i>
Lectin/C-type/NK Receptor	<i>Klra10</i>	
Lectin/C-type/NK Receptor	<i>Klra17</i>	
Lectin/C-type/NK Receptor	<i>Klra2</i>	
Lectin/C-type/NK Receptor	<i>Klra3</i>	
Lectin/C-type/NK Receptor	<i>Klra4</i>	
Lectin/C-type/NK Receptor	<i>Klra8</i>	
Lectin/C-type/NK Receptor	<i>Klrb1f</i>	
Lectin/C-type/NK Receptor	<i>Klrc1</i>	<i>KLRC1</i>
Lectin/C-type/NK Receptor	<i>Klrc2</i>	<i>KLRC2</i>
Lectin/C-type/NK Receptor	<i>Klrd1</i>	<i>KLRD1</i>
Lectin/C-type/NK Receptor	<i>Klrg1</i>	<i>KLRG1</i>
Lectin/C-type/NK Receptor	<i>Klrk1</i>	<i>KLRK1</i>
Lectin/I-type/Non-Siglec	<i>L1cam</i>	<i>L1CAM</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Lamp1</i>	<i>LAMP1</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Lamp2</i>	<i>LAMP2</i>
Glycosyltransferase	<i>Large1</i>	<i>LARGE1</i>
Glycosyltransferase	<i>Large2</i>	<i>LARGE2</i>
Lectin/C-type/Type 1 Receptor	<i>Layn</i>	<i>LAYN</i>
Glycosylhydrolase	<i>Lct</i>	<i>LCT</i>
Glycosylhydrolase	<i>Lctl</i>	<i>LCTL</i>
Glycosyltransferase	<i>Lfng</i>	<i>LFNG</i>
Lectin/S-Type/Galectin	<i>Lgals1</i>	<i>LGALS1</i>
Lectin/S-Type/Galectin	<i>Lgals12</i>	<i>LGALS12</i>
Lectin/S-Type/Galectin	<i>Lgals2</i>	<i>LGALS2</i>
Lectin/S-Type/Galectin	<i>Lgals3</i>	<i>LGALS3</i>
Lectin/S-Type/Galectin	<i>Lgals4</i>	<i>LGALS4</i>
Lectin/S-Type/Galectin	<i>Lgals7</i>	<i>LGALS7</i>
Lectin/S-Type/Galectin	<i>Lgals8</i>	<i>LGALS8</i>
Lectin/S-Type/Galectin	<i>Lgals9</i>	<i>LGALS9</i>
Lectin/S-Type/Galectin	<i>Lgalsl</i>	<i>LGALS</i>

Grupo	Gen MM	Gen HS
Glycolipid Related	<i>Lipa</i>	<i>LIPA</i>
Lectin/L-Type	<i>Lman1</i>	<i>LMAN1</i>
Lectin/L-Type	<i>Lman1l</i>	<i>LMAN1L</i>
Lectin/L-Type	<i>Lman2</i>	<i>LMAN2</i>
Lectin/L-Type	<i>Lman2l</i>	<i>LMAN2L</i>
Glycosaminoglycan Related	<i>Lum</i>	<i>LUM</i>
Lectin/C-type/Macrophage Mannose Receptor	<i>Ly75</i>	<i>LY75</i>
Glycosylhydrolase	<i>Lyg1</i>	<i>LYG1</i>
Glycosylhydrolase	<i>Lyg2</i>	<i>LYG2</i>
Glycosylhydrolase	<i>Lyz1</i>	<i>LYZ1</i>
Glycosylhydrolase	<i>Lyz2</i>	
Glycosylhydrolase	<i>Lyz1l</i>	
Glycosylhydrolase	<i>Lyz14</i>	<i>LYZ14</i>
Glycosylhydrolase	<i>Lyz16</i>	<i>LYZ16</i>
Lectin/P-Type	<i>M6pr</i>	<i>M6PR</i>
Glycosaminoglycan Related	<i>Madcam1</i>	<i>MADCAM1</i>
Lectin/I-type/Siglec	<i>Mag</i>	<i>MAG</i>
Glycosylhydrolase	<i>Man1a</i>	
Glycosylhydrolase	<i>Man1a2</i>	<i>MAN1A2</i>
Glycosylhydrolase	<i>Man1b1</i>	<i>MAN1B1</i>
Glycosylhydrolase	<i>Man1c1</i>	<i>MAN1C1</i>
Glycosylhydrolase	<i>Man2a1</i>	<i>MAN2A1</i>
Glycosylhydrolase	<i>Man2a2</i>	<i>MAN2A2</i>
Glycosylhydrolase	<i>Man2b1</i>	<i>MAN2B1</i>
Glycosylhydrolase	<i>Man2b2</i>	<i>MAN2B2</i>
Glycosylhydrolase	<i>Man2c1</i>	<i>MAN2C1</i>
Glycosylhydrolase	<i>Manba</i>	<i>MANBA</i>
Glycosylhydrolase	<i>Manea</i>	<i>MANEA</i>
Glycosylhydrolase	<i>Maneal</i>	<i>MANEAL</i>
Lectin/Lectin-Associated Proteins/MASP	<i>Masp1</i>	<i>MASP1</i>
Lectin/Lectin-Associated Proteins/MASP	<i>Masp2</i>	<i>MASP2</i>
Lectin/C-type/Collectin	<i>Mbl1</i>	
Lectin/C-type/Collectin	<i>Mbl2</i>	<i>MBL2</i>
Lectin/Misc. Glycan-binding/GAG-binding	<i>Mdk</i>	<i>MDK</i>
Glycosyltransferase	<i>Mfng</i>	<i>MFNG</i>
Glycosyltransferase	<i>Mgat1</i>	<i>MGAT1</i>
Glycosyltransferase	<i>Mgat2</i>	<i>MGAT2</i>
Glycosyltransferase	<i>Mgat3</i>	<i>MGAT3</i>
Glycosyltransferase	<i>Mgat4a</i>	<i>MGAT4A</i>

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>Mgat4b</i>	<i>MGAT4B</i>
Glycosyltransferase	<i>Mgat4c</i>	<i>MGAT4C</i>
Glycosyltransferase	<i>Mgat4d</i>	<i>MGAT4D</i>
Glycosyltransferase	<i>Mgat5</i>	<i>MGAT5</i>
Glycosyltransferase	<i>Mgat5b</i>	<i>MGAT5B</i>
Glycosylhydrolase	<i>Mgea5</i>	
Lectin/C-type/Type 2 Receptor	<i>Mgl2</i>	
Glycosylhydrolase	<i>Mogs</i>	<i>MOGS</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Mpdu1</i>	<i>MPDU1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Mpi</i>	<i>MPI</i>
Lectin/C-type/Macrophage Mannose Receptor	<i>Mrc1</i>	<i>MRC1</i>
Lectin/C-type/Macrophage Mannose Receptor	<i>Mrc2</i>	<i>MRC2</i>
Glycolipid Related	<i>Naaa</i>	<i>NAAA</i>
Glycosylhydrolase	<i>Naga</i>	<i>NAGA</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Nagk</i>	<i>NAGK</i>
Glycosylhydrolase	<i>Naglu</i>	<i>NAGLU</i>
Miscellaneous	<i>Nagpa</i>	<i>NAGPA</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Nans</i>	<i>NANS</i>
Lectin/I-type/Non-Siglec	<i>Ncam1</i>	<i>NCAM1</i>
Lectin/C-type/Proteoglycan	<i>Ncan</i>	<i>NCAN</i>
Carbohydrate Esterase	<i>Nceh1</i>	<i>NCEH1</i>
Sulfate Related/Sulfotransferase	<i>Ndst1</i>	<i>NDST1</i>
Sulfate Related/Sulfotransferase	<i>Ndst2</i>	<i>NDST2</i>
Sulfate Related/Sulfotransferase	<i>Ndst3</i>	<i>NDST3</i>
Sulfate Related/Sulfotransferase	<i>Ndst4</i>	<i>NDST4</i>
Glycosylhydrolase	<i>Neu1</i>	<i>NEU1</i>
Glycosylhydrolase	<i>Neu2</i>	<i>NEU2</i>
Glycosylhydrolase	<i>Neu3</i>	<i>NEU3</i>
Glycosylhydrolase	<i>Neu4</i>	<i>NEU4</i>
Miscellaneous	<i>Ngly1</i>	<i>NGLY1</i>
Carbohydrate Esterase	<i>Nlgn2</i>	<i>NLGN2</i>
Carbohydrate Esterase	<i>Nlgn3</i>	<i>NLGN3</i>
Carbohydrate Esterase	<i>Notum</i>	<i>NOTUM</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Npl</i>	<i>NPL</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Nt5m</i>	<i>NT5M</i>
Glycosyltransferase	<i>Ogt</i>	<i>OGT</i>
Lectin/C-type/NK Receptor	<i>Olr1</i>	<i>OLR1</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Ost4</i>	<i>OST4</i>
Glycosylhydrolase	<i>Ovgp1</i>	<i>OVGP1</i>

Grupo	Gen MM	Gen HS
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Papss1</i>	<i>PAPSS1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Papss2</i>	<i>PAPSS2</i>
Housekeeping Gene	<i>Pcx</i>	
Lectin/Lectin-Associated Proteins/PDI	<i>Pdia3</i>	<i>PDIA3</i>
Lectin/I-type/Non-Siglec	<i>Pecam1</i>	<i>PECAM1</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Pfkfb2</i>	<i>PFKFB2</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Pfkfb3</i>	<i>PFKFB3</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Pfkl</i>	<i>PFKL</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Pfkm</i>	<i>PFKM</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Pfkp</i>	<i>PFKP</i>
Glycosylhydrolase	<i>Pgghg</i>	<i>PGGHG</i>
Housekeeping Gene	<i>Pgk1</i>	
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Pgm1</i>	<i>PGM1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Pgm3</i>	<i>PGM3</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Pgm5</i>	<i>PGM5</i>
Sulfate Related/Sulfotransferase	<i>Phactr1</i>	<i>PHACTR1</i>
Glycosyltransferase	<i>Piga</i>	<i>PIGA</i>
Glycosyltransferase	<i>Pigb</i>	<i>PIGB</i>
GPI anchor biosynthesis Related	<i>Pigc</i>	<i>PIGC</i>
GPI anchor biosynthesis Related	<i>Pigf</i>	<i>PIGF</i>
GPI anchor biosynthesis Related	<i>Pigg</i>	<i>PIGG</i>
GPI anchor biosynthesis Related	<i>Pigh</i>	<i>PIGH</i>
GPI anchor biosynthesis Related	<i>Pigk</i>	<i>PIGK</i>
GPI anchor biosynthesis Related	<i>Pigl</i>	<i>PIGL</i>
Glycosyltransferase	<i>Pigm</i>	<i>PIGM</i>
GPI anchor biosynthesis Related	<i>Pign</i>	<i>PIGN</i>
GPI anchor biosynthesis Related	<i>Pigo</i>	<i>PIGO</i>
GPI anchor biosynthesis Related	<i>Pigp</i>	<i>PIGP</i>
GPI anchor biosynthesis Related	<i>Pigq</i>	<i>PIGQ</i>
GPI anchor biosynthesis Related	<i>Pigs</i>	<i>PIGS</i>
GPI anchor biosynthesis Related	<i>Pigt</i>	<i>PIGT</i>
GPI anchor biosynthesis Related	<i>Pigu</i>	<i>PIGU</i>
Glycosyltransferase	<i>Pigv</i>	<i>PIGV</i>
GPI anchor biosynthesis Related	<i>Pigw</i>	<i>PIGW</i>
GPI anchor biosynthesis Related	<i>Pigx</i>	<i>PIGX</i>
Glycosyltransferase	<i>Pigz</i>	<i>PIGZ</i>
Lectin/C-type/Polycystin	<i>Pkd1</i>	<i>PKD1</i>
Lectin/C-type/Polycystin	<i>Pkd1l2</i>	<i>PKD1L2</i>
Lectin/C-type/Polycystin	<i>Pkd2</i>	<i>PKD2</i>

Grupo	Gen MM	Gen HS
Lectin/C-type/Macrophage Mannose Receptor	<i>Pla2r1</i>	<i>PLA2R1</i>
Glycosaminoglycan Related	<i>Plaur</i>	<i>PLAUR</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Pmm1</i>	<i>PMM1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Pmm2</i>	<i>PMM2</i>
Glycosyltransferase	<i>Pofut1</i>	<i>POFUT1</i>
Glycosyltransferase	<i>Pofut2</i>	<i>POFUT2</i>
Glycosyltransferase	<i>Pomgnt1</i>	<i>POMGNT1</i>
Glycosyltransferase	<i>Pomgnt2</i>	<i>POMGNT2</i>
Glycosyltransferase	<i>Pomt1</i>	<i>POMT1</i>
Glycosyltransferase	<i>Pomt2</i>	<i>POMT2</i>
Housekeeping Gene	<i>Ppia</i>	<i>PPIA</i>
Carbohydrate Binding Module	<i>Ppp1r3a</i>	<i>PPP1R3A</i>
Carbohydrate Binding Module	<i>Ppp1r3b</i>	<i>PPP1R3B</i>
Carbohydrate Binding Module	<i>Ppp1r3c</i>	<i>PPP1R3C</i>
Carbohydrate Binding Module	<i>Ppp1r3f</i>	<i>PPP1R3F</i>
Lectin/C-type/Acidic Neck CTLD	<i>Prg2</i>	<i>PRG2</i>
Lectin/C-type/Acidic Neck CTLD	<i>Prg3</i>	<i>PRG3</i>
Miscellaneous	<i>Prkcsh</i>	<i>PRKCSH</i>
Housekeeping Gene	<i>Psm10</i>	<i>PSMD10</i>
Glycosaminoglycan Related	<i>Ptn</i>	<i>PTN</i>
Glycosyltransferase	<i>Pygb</i>	<i>PYGB</i>
Glycosyltransferase	<i>Pygl</i>	<i>PYGL</i>
Glycosyltransferase	<i>Pygm</i>	<i>PYGM</i>
GPI anchor biosynthesis Related	<i>Pyurf</i>	<i>PYURF</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Rack1</i>	<i>RACK1</i>
Lectin/C-type/Free CTLDs	<i>Reg1</i>	
Lectin/C-type/Free CTLDs	<i>Reg2</i>	
Lectin/C-type/Free CTLDs	<i>Reg3b</i>	
Lectin/C-type/Free CTLDs	<i>Reg3d</i>	
Lectin/C-type/Free CTLDs	<i>Reg4</i>	<i>REG4</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Renbp</i>	<i>RENBP</i>
Glycosyltransferase	<i>Rfng</i>	<i>RFNG</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Rft1</i>	<i>RFT1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Rnf123</i>	<i>RNF123</i>
Housekeeping Gene	<i>Rpl13a</i>	<i>RPL13A</i>
Housekeeping Gene	<i>Rpl4</i>	<i>RPL4</i>
Housekeeping Gene	<i>Rpl9</i>	<i>RPL9</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Rpn1</i>	<i>RPN1</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Rpn2</i>	<i>RPN2</i>

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>Rtf2</i>	<i>RTF2</i>
Glycosaminoglycan Related	<i>Sdc1</i>	<i>SDC1</i>
Glycosaminoglycan Related	<i>Sdc2</i>	<i>SDC2</i>
Glycosaminoglycan Related	<i>Sdc3</i>	<i>SDC3</i>
Glycosaminoglycan Related	<i>Sdc4</i>	<i>SDC4</i>
Lectin/Misc. Glycan-binding/GAG-binding	<i>Sdcbp</i>	<i>SDCBP</i>
Lectin/Misc. Glycan-binding/GAG-binding	<i>Sdcbp2</i>	<i>SDCBP2</i>
Housekeeping Gene	<i>Sdha</i>	<i>SDHA</i>
Glycosyltransferase	<i>Sec1</i>	
Lectin/C-type/Selectin	<i>Sele</i>	<i>SELE</i>
Lectin/C-type/Selectin	<i>Sell</i>	<i>SELL</i>
Lectin/C-type/Selectin	<i>Selp</i>	<i>SELP</i>
Glycosaminoglycan Related	<i>Selplg</i>	<i>SELPLG</i>
Glycosaminoglycan Related	<i>Serpinc1</i>	<i>SERPINC1</i>
Lectin/C-type/Collectin	<i>Sftpa1</i>	<i>SFTPA1</i>
Lectin/C-type/Collectin	<i>Sftpb</i>	<i>SFTPB</i>
Lectin/C-type/Collectin	<i>Sftpc</i>	<i>SFTPC</i>
Lectin/C-type/Collectin	<i>Sftpd</i>	<i>SFTPD</i>
Glycolipid Related	<i>Sgms1</i>	<i>SGMS1</i>
Glycolipid Related	<i>Sgpl1</i>	<i>SGPL1</i>
Sulfate Related/Sulfohydrolase	<i>Sgsh</i>	<i>SGSH</i>
Glycosaminoglycan Related	<i>Shh</i>	<i>SHH</i>
Miscellaneous	<i>Siae</i>	<i>SIAE</i>
Lectin/I-type/Siglec	<i>Siglec1</i>	<i>SIGLEC1</i>
Lectin/I-type/CD33-Related	<i>Siglece</i>	
Lectin/I-type/CD33-Related	<i>Siglecf</i>	
Lectin/I-type/CD33-Related	<i>Siglecg</i>	
Lectin/I-type/CD33-Related	<i>Siglech</i>	
Glycosylhydrolase	<i>Sis</i>	
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc17a5</i>	<i>SLC17A5</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc26a2</i>	<i>SLC26A2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a1</i>	<i>SLC2A1</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a10</i>	<i>SLC2A10</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a12</i>	<i>SLC2A12</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a2</i>	<i>SLC2A2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a3</i>	<i>SLC2A3</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a4</i>	<i>SLC2A4</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a5</i>	<i>SLC2A5</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc2a8</i>	<i>SLC2A8</i>

Grupo	Gen MM	Gen HS
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35a1</i>	<i>SLC35A1</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35a2</i>	<i>SLC35A2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35a3</i>	<i>SLC35A3</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35b1</i>	<i>SLC35B1</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35b2</i>	<i>SLC35B2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35b3</i>	<i>SLC35B3</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35b4</i>	<i>SLC35B4</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35c1</i>	<i>SLC35C1</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35c2</i>	<i>SLC35C2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35d1</i>	<i>SLC35D1</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35d2</i>	<i>SLC35D2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc35d3</i>	<i>SLC35D3</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc37a4</i>	<i>SLC37A4</i>
Glycosylhydrolase	<i>Slc3a1</i>	<i>SLC3A1</i>
Glycosylhydrolase	<i>Slc3a2</i>	<i>SLC3A2</i>
Carbohydrate Synthesis and Metabolism/Transporter	<i>Slc5a1</i>	<i>SLC5A1</i>
Glycosaminoglycan Related	<i>Smc3</i>	<i>SMC3</i>
Glycolipid Related	<i>Smpd1</i>	<i>SMPD1</i>
Glycolipid Related	<i>Smpd3</i>	<i>SMPD3</i>
Glycosylhydrolase	<i>Spam1</i>	<i>SPAM1</i>
Glycolipid Related	<i>Sphk1</i>	<i>SPHK1</i>
Glycolipid Related	<i>Sphk2</i>	<i>SPHK2</i>
Glycolipid Related	<i>Sptlc1</i>	<i>SPTLC1</i>
Glycolipid Related	<i>Sptlc2</i>	<i>SPTLC2</i>
Glycosaminoglycan Related	<i>Srgn</i>	<i>SRGN</i>
Glycosyltransferase	<i>St3gal1</i>	<i>ST3GAL1</i>
Glycosyltransferase	<i>St3gal2</i>	<i>ST3GAL2</i>
Glycosyltransferase	<i>St3gal3</i>	<i>ST3GAL3</i>
Glycosyltransferase	<i>St3gal4</i>	<i>ST3GAL4</i>
Glycosyltransferase	<i>St3gal5</i>	<i>ST3GAL5</i>
Glycosyltransferase	<i>St3gal6</i>	<i>ST3GAL6</i>
Glycosyltransferase	<i>St6gal1</i>	<i>ST6GAL1</i>
Glycosyltransferase	<i>St6gal2</i>	<i>ST6GAL2</i>
Glycosyltransferase	<i>St6galnac1</i>	<i>ST6GALNAC1</i>
Glycosyltransferase	<i>St6galnac2</i>	<i>ST6GALNAC2</i>
Glycosyltransferase	<i>St6galnac3</i>	<i>ST6GALNAC3</i>
Glycosyltransferase	<i>St6galnac4</i>	<i>ST6GALNAC4</i>
Glycosyltransferase	<i>St6galnac5</i>	<i>ST6GALNAC5</i>
Glycosyltransferase	<i>St6galnac6</i>	<i>ST6GALNAC6</i>

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>St8sia1</i>	<i>ST8SIA1</i>
Glycosyltransferase	<i>St8sia2</i>	<i>ST8SIA2</i>
Glycosyltransferase	<i>St8sia3</i>	<i>ST8SIA3</i>
Glycosyltransferase	<i>St8sia4</i>	<i>ST8SIA4</i>
Glycosyltransferase	<i>St8sia5</i>	<i>ST8SIA5</i>
Glycosyltransferase	<i>St8sia6</i>	<i>ST8SIA6</i>
Carbohydrate Binding Module	<i>Stbd1</i>	<i>STBD1</i>
Glycosyltransferase	<i>Stt3a</i>	<i>STT3A</i>
Glycosyltransferase	<i>Stt3b</i>	<i>STT3B</i>
Sulfate Related/Sulfatase	<i>Sulf1</i>	<i>SULF1</i>
Sulfate Related/Sulfatase	<i>Sulf2</i>	<i>SULF2</i>
Housekeeping Gene	<i>Tbp</i>	<i>TBP</i>
Carbohydrate Synthesis and Metabolism/Metabolic Enzymes	<i>Tgds</i>	<i>TGDS</i>
Lectin/C-type/Endosialin	<i>Thbd</i>	<i>THBD</i>
Glycolipid Related	<i>Thnsl2</i>	<i>THNSL2</i>
Sulfate Related/Sulfotransferase	<i>Tpst1</i>	<i>TPST1</i>
Sulfate Related/Sulfotransferase	<i>Tpst2</i>	<i>TPST2</i>
Glycolipid Related	<i>Tram2</i>	<i>TRAM2</i>
Glycosylhydrolase	<i>Treh</i>	<i>TREH</i>
Lipid-Linked Oligosaccharide Pathway Related	<i>Tusc3</i>	<i>TUSC3</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Uap1</i>	<i>UAP1</i>
Glycosyltransferase	<i>Ugcg</i>	<i>UGCG</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Ugdh</i>	<i>UGDH</i>
Glycosyltransferase	<i>Ugg1</i>	<i>UGGT1</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Ugp2</i>	<i>UGP2</i>
Glycosyltransferase	<i>Ugt1a1</i>	<i>UGT1A1</i>
Glycosyltransferase	<i>Ugt1a10</i>	<i>UGT1A10</i>
Glycosyltransferase	<i>Ugt1a2</i>	
Glycosyltransferase	<i>Ugt1a5</i>	<i>UGT1A5</i>
Glycosyltransferase	<i>Ugt1a6a</i>	
Glycosyltransferase	<i>Ugt1a6b</i>	
Glycosyltransferase	<i>Ugt1a7c</i>	
Glycosyltransferase	<i>Ugt1a9</i>	<i>UGT1A9</i>
Glycosyltransferase	<i>Ugt2a1</i>	<i>UGT2A1</i>
Glycosyltransferase	<i>Ugt2a2</i>	<i>UGT2A2</i>
Glycosyltransferase	<i>Ugt2a3</i>	<i>UGT2A3</i>
Glycosyltransferase	<i>Ugt2b1</i>	
Glycosyltransferase	<i>Ugt2b34</i>	
Glycosyltransferase	<i>Ugt2b35</i>	

Grupo	Gen MM	Gen HS
Glycosyltransferase	<i>Ugt2b36</i>	
Glycosyltransferase	<i>Ugt2b37</i>	
Glycosyltransferase	<i>Ugt2b38</i>	
Glycosyltransferase	<i>Ugt2b5</i>	
Glycosyltransferase	<i>Ugt3a1</i>	<i>UGT3A1</i>
Glycosyltransferase	<i>Ugt3a2</i>	<i>UGT3A2</i>
Glycosyltransferase	<i>Ugt8a</i>	
Sulfate Related/Sulfotransferase	<i>Ust</i>	<i>UST</i>
Carbohydrate Synthesis and Metabolism/Nucleotide Synthesis	<i>Uxs1</i>	<i>UXS1</i>
Lectin/I-type/Non-Siglec	<i>Vcam1</i>	<i>VCAM1</i>
Glycosyltransferase	<i>Xxylt1</i>	<i>XXYLT1</i>
Glycosyltransferase	<i>Xylt1</i>	<i>XYLT1</i>
Glycosyltransferase	<i>Xylt2</i>	<i>XYLT2</i>
Housekeeping Gene	<i>Ywhab</i>	<i>YWHAB</i>

Anexo C: Resultados de la comparación por grupos del infiltrado inmune en modelos experimentales murinos y pacientes de CCR.

1. GSE57533

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Linfocitos B	CCRAC	CONTROL	1,000	1,000	Wilcoxon
Linfocitos B	CCRAC	EII	0,057	0,086	Wilcoxon
Linfocitos B	CONTROL	EII	0,057	0,086	Wilcoxon
Linfocitos T CD4 ⁺	CCRAC	CONTROL	0,507	0,760	Wilcoxon
Linfocitos T CD4 ⁺	CCRAC	EII	1,000	1,000	Wilcoxon
Linfocitos T CD4 ⁺	CONTROL	EII	0,359	0,760	Wilcoxon
Linfocitos T CD8 ⁺	CCRAC	CONTROL	1,000	1,000	Wilcoxon
Linfocitos T CD8 ⁺	CCRAC	EII	0,199	0,600	Wilcoxon
Linfocitos T CD8 ⁺	CONTROL	EII	0,582	0,870	Wilcoxon
Células Dendríticas	CCRAC	CONTROL	1,000	1,000	Wilcoxon
Células Dendríticas	CCRAC	EII	0,629	1,000	Wilcoxon
Células Dendríticas	CONTROL	EII	0,721	1,000	Wilcoxon
Eosinófilos	CCRAC	CONTROL	0,643	0,960	Wilcoxon
Eosinófilos	CCRAC	EII	0,359	0,960	Wilcoxon
Eosinófilos	CONTROL	EII	1,000	1,000	Wilcoxon
Macrófagos	CCRAC	CONTROL	0,184	0,340	Wilcoxon
Macrófagos	CCRAC	EII	0,229	0,340	Wilcoxon
Macrófagos	CONTROL	EII	0,359	0,360	Wilcoxon
Mastocitos	CCRAC	CONTROL	1,000	1,000	Wilcoxon
Mastocitos	CCRAC	EII	0,629	1,000	Wilcoxon
Mastocitos	CONTROL	EII	1,000	1,000	Wilcoxon
Monocitos	CCRAC	CONTROL	1,000	1,000	Wilcoxon
Monocitos	CCRAC	EII	0,721	1,000	Wilcoxon
Monocitos	CONTROL	EII	0,229	0,690	Wilcoxon
Neutrófilos	CCRAC	CONTROL	0,505	0,760	Wilcoxon
Neutrófilos	CCRAC	EII	1,000	1,000	Wilcoxon
Neutrófilos	CONTROL	EII	0,270	0,760	Wilcoxon
Asesinos Naturales	CCRAC	CONTROL	0,643	0,640	Wilcoxon
Asesinos Naturales	CCRAC	EII	0,229	0,340	Wilcoxon
Asesinos Naturales	CONTROL	EII	0,050	0,150	Wilcoxon

Grupo	Resultado Test Shapiro Multivariable
Control	No hay muestras suficientes para correr el test
EII	No hay muestras suficientes para correr el test
CCRAC	No hay muestras suficientes para correr el test

2. GSE64658

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Mastocitos	CCRAC Distal temprano	CCRAC Distal tardío	0,381	0,380	Wilcoxon
Mastocitos	CCRAC Distal temprano	Control Distal	0,048	0,071	Wilcoxon
Mastocitos	CCRAC Distal tardío	Control Distal	0,009	0,026	Wilcoxon
Neutrófilos	CCRAC Distal temprano	CCRAC Distal tardío	0,795	0,800	Wilcoxon
Neutrófilos	CCRAC Distal temprano	Control Distal	0,052	0,078	Wilcoxon
Neutrófilos	CCRAC Distal tardío	Control Distal	0,010	0,029	Wilcoxon
Eosinófilos	CCRAC Distal temprano	CCRAC Distal tardío	0,637	0,640	Wilcoxon
Eosinófilos	CCRAC Distal tardío	Control Distal	0,405	0,640	Wilcoxon
Linfocitos B de memoria	CCRAC Distal temprano	CCRAC Distal tardío	0,243	0,240	Wilcoxon
Linfocitos B de memoria	CCRAC Distal temprano	Control Distal	0,052	0,097	Wilcoxon
Linfocitos B de memoria	CCRAC Distal tardío	Control Distal	0,065	0,097	Wilcoxon
Linfocitos B virgen	CCRAC Distal temprano	CCRAC Distal tardío	0,052	0,160	Wilcoxon
Linfocitos B virgen	CCRAC Distal temprano	Control Distal	0,220	0,330	Wilcoxon
Linfocitos B virgen	CCRAC Distal tardío	Control Distal	0,405	0,400	Wilcoxon
Plasmática	CCRAC Distal temprano	CCRAC Distal tardío	1,000	1,000	Wilcoxon
Plasmática	CCRAC Distal temprano	Control Distal	0,048	0,071	Wilcoxon
Plasmática	CCRAC Distal tardío	Control Distal	0,009	0,026	Wilcoxon
Linfocitos T CD8 ⁺ activado	CCRAC Distal temprano	CCRAC Distal tardío	0,028	0,083	Wilcoxon
Linfocitos T CD8 ⁺ activado	CCRAC Distal temprano	Control Distal	0,364	0,360	Wilcoxon
Linfocitos T CD8 ⁺ activado	CCRAC Distal tardío	Control Distal	0,222	0,330	Wilcoxon
Linfocitos T CD8 ⁺ virgen	CCRAC Distal temprano	CCRAC Distal tardío	0,167	0,170	Wilcoxon
Linfocitos T CD8 ⁺ virgen	CCRAC Distal temprano	Control Distal	0,167	0,170	Wilcoxon
Linfocitos T CD8 ⁺ virgen	CCRAC Distal tardío	Control Distal	0,004	0,013	Wilcoxon
Macrófagos M0	CCRAC Distal temprano	CCRAC Distal tardío	0,694	0,870	Wilcoxon
Macrófagos M0	CCRAC Distal temprano	Control Distal	0,511	0,870	Wilcoxon
Macrófagos M0	CCRAC Distal tardío	Control Distal	0,873	0,870	Wilcoxon
Macrófagos M1	CCRAC Distal temprano	CCRAC Distal tardío	0,167	0,250	Wilcoxon
Macrófagos M1	CCRAC Distal temprano	Control Distal	0,095	0,250	Wilcoxon
Macrófagos M1	CCRAC Distal tardío	Control Distal	0,485	0,480	Wilcoxon

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Macrófagos M2	CCRAC Distal temprano	CCRAC Distal tardío	0,357	0,690	Wilcoxon
Macrófagos M2	CCRAC Distal temprano	Control Distal	0,888	0,890	Wilcoxon
Macrófagos M2	CCRAC Distal tardío	Control Distal	0,462	0,690	Wilcoxon
Linfocitos T Reguladores	CCRAC Distal temprano	CCRAC Distal tardío	0,714	0,710	Wilcoxon
Linfocitos T Reguladores	CCRAC Distal temprano	Control Distal	0,048	0,071	Wilcoxon
Linfocitos T Reguladores	CCRAC Distal tardío	Control Distal	0,026	0,071	Wilcoxon
Linfocitos T CD4 ⁺ de memoria	CCRAC Distal temprano	CCRAC Distal tardío	0,138	0,210	Wilcoxon
Linfocitos T CD4 ⁺ de memoria	CCRAC Distal temprano	Control Distal	1,000	1,000	Wilcoxon
Linfocitos T CD4 ⁺ de memoria	CCRAC Distal tardío	Control Distal	0,115	0,210	Wilcoxon
Linfocitos T CD4 ⁺ virgen	CCRAC Distal temprano	CCRAC Distal tardío	0,671	0,670	Wilcoxon
Linfocitos T CD4 ⁺ virgen	CCRAC Distal temprano	Control Distal	0,225	0,670	Wilcoxon
Linfocitos T CD4 ⁺ virgen	CCRAC Distal tardío	Control Distal	0,455	0,670	Wilcoxon
Linfocitos T CD4 ⁺ folicular	CCRAC Distal temprano	CCRAC Distal tardío	0,381	0,380	Wilcoxon
Linfocitos T CD4 ⁺ folicular	CCRAC Distal temprano	Control Distal	0,009	0,014	Wilcoxon
Linfocitos T CD4 ⁺ folicular	CCRAC Distal tardío	Control Distal	0,003	0,008	Wilcoxon
T cooperador 17	CCRAC Distal temprano	CCRAC Distal tardío	0,686	0,690	Wilcoxon
T cooperador 17	CCRAC Distal temprano	Control Distal	0,088	0,260	Wilcoxon
T cooperador 17	CCRAC Distal tardío	Control Distal	0,226	0,340	Wilcoxon
T cooperador 2	CCRAC Distal temprano	CCRAC Distal tardío	0,381	0,380	Wilcoxon
T cooperador 2	CCRAC Distal temprano	Control Distal	0,043	0,130	Wilcoxon
T cooperador 2	CCRAC Distal tardío	Control Distal	0,087	0,130	Wilcoxon
Monocitos	CCRAC Distal temprano	CCRAC Distal tardío	0,052	0,078	Wilcoxon
Monocitos	CCRAC Distal temprano	Control Distal	1,000	1,000	Wilcoxon
Monocitos	CCRAC Distal tardío	Control Distal	0,020	0,060	Wilcoxon
T Gamma Delta	CCRAC Distal temprano	CCRAC Distal tardío	0,714	0,710	Wilcoxon
T Gamma Delta	CCRAC Distal temprano	Control Distal	0,381	0,570	Wilcoxon
T Gamma Delta	CCRAC Distal tardío	Control Distal	0,240	0,570	Wilcoxon
Asesinos naturales descansando	CCRAC Distal temprano	CCRAC Distal tardío	0,364	0,360	Wilcoxon
Asesinos naturales descansando	CCRAC Distal temprano	Control Distal	0,024	0,046	Wilcoxon
Asesinos naturales descansando	CCRAC Distal tardío	Control Distal	0,030	0,046	Wilcoxon
Asesinos naturales activados	CCRAC Distal temprano	CCRAC Distal tardío	0,120	0,120	Wilcoxon

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Asesinos naturales activados	CCRAC Distal temprano	Control Distal	0,052	0,078	Wilcoxon
Asesinos naturales activados	CCRAC Distal tardío	Control Distal	0,010	0,029	Wilcoxon
Dendríticas activadas	CCRAC Distal temprano	CCRAC Distal tardío	0,381	0,380	Wilcoxon
Dendríticas activadas	CCRAC Distal temprano	Control Distal	0,009	0,014	Wilcoxon
Dendríticas activadas	CCRAC Distal tardío	Control Distal	0,003	0,008	Wilcoxon
Dendríticas inmaduras	CCRAC Distal temprano	CCRAC Distal tardío	0,596	0,690	Wilcoxon
Dendríticas inmaduras	CCRAC Distal temprano	Control Distal	0,686	0,690	Wilcoxon
Dendríticas inmaduras	CCRAC Distal tardío	Control Distal	0,061	0,180	Wilcoxon

Grupo	Resultado Test Shapiro Multivariable
CCRAC Distal temprano	No hay muestras suficientes para correr el test
CCRAC Distal tardío	No hay muestras suficientes para correr el test
Control Distal	No hay muestras suficientes para correr el test

3. Pacientes CCR - Cibersort

3.1 xCell

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Dendríticas activadas	Tumor	Control	0,000	0,000	Wilcoxon
Adipocitos	Tumor	Control	0,000	0,000	Wilcoxon
Astroцитos	Tumor	Control	0,155	0,160	Wilcoxon
Linfocitos B	Tumor	Control	0,000	0,000	Wilcoxon
Basófilos	Tumor	Control	0,735	0,740	Wilcoxon
Linfocitos T CD4 ⁺ de memoria	Tumor	Control	0,960	0,960	Wilcoxon
Linfocitos T CD4 ⁺ virgen	Tumor	Control	0,671	0,670	Wilcoxon
Linfocitos T CD4 ⁺	Tumor	Control	0,558	0,560	Wilcoxon
Linfocitos T CD4 ⁺ de memoria central	Tumor	Control	0,965	0,970	Wilcoxon
T CD4 ⁺ de memoria efectora	Tumor	Control	0,001	0,001	Wilcoxon
T CD8 ⁺ virgen	Tumor	Control	0,008	0,008	Wilcoxon
Linfocitos T CD8 ⁺	Tumor	Control	0,034	0,034	Wilcoxon
Linfocitos T CD8 ⁺ de memoria central	Tumor	Control	0,001	0,001	Wilcoxon
Linfocitos T CD8 ⁺ de memoria efectora	Tumor	Control	0,422	0,420	Wilcoxon
Dendríticas convencionales	Tumor	Control	0,000	0,000	Wilcoxon
B de memoria conmutada de clase	Tumor	Control	0,000	0,000	Wilcoxon
Dendríticas	Tumor	Control	0,000	0,000	Wilcoxon

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Eosinófilos	Tumor	Control	0,000	0,000	Wilcoxon
Dendríticas inmaduras	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos M1	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos M2	Tumor	Control	0,003	0,003	Wilcoxon
Mastocitos	Tumor	Control	0,000	0,000	Wilcoxon
Linfocitos B de memoria	Tumor	Control	0,000	0,000	Wilcoxon
Monocitos	Tumor	Control	0,002	0,002	Wilcoxon
Linfocitos B virgen	Tumor	Control	0,000	0,000	Wilcoxon
Neutrófilos	Tumor	Control	0,000	0,000	Wilcoxon
Asesinos Naturales	Tumor	Control	0,243	0,240	Wilcoxon
T Asesinos naturales	Tumor	Control	0,000	0,000	Wilcoxon
Plasmática	Tumor	Control	0,000	0,000	Wilcoxon
Linfocitos B pro	Tumor	Control	0,000	0,000	Wilcoxon
T Gamma Delta	Tumor	Control	0,028	0,028	Wilcoxon
T cooperador 1	Tumor	Control	0,000	0,000	Wilcoxon
T cooperador 2	Tumor	Control	0,027	0,027	Wilcoxon
T regulador	Tumor	Control	0,004	0,004	Wilcoxon

Grupo	Resultado Test Shapiro Multivariable
Tumor	No hay muestras suficientes para correr el test
Control	No hay muestras suficientes para correr el test

3.2 Cibersort

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Linfocitos B virgen	Tumor	Control	0,691	0,690	Wilcoxon
Linfocitos B de memoria	Tumor	Control	0,004	0,004	Wilcoxon
Plasmática	Tumor	Control	0,000	0,000	Wilcoxon
T CD8 ⁺	Tumor	Control	0,046	0,046	Wilcoxon
T CD4 ⁺ de memoria descansando	Tumor	Control	0,087	0,087	Wilcoxon
T CD4 ⁺ de memoria activado	Tumor	Control	0,000	0,000	Wilcoxon
T regulador ayudante folicular	Tumor	Control	0,937	0,940	Wilcoxon
T regulador	Tumor	Control	0,004	0,004	Wilcoxon
T Gamma Delta	Tumor	Control	0,000	0,000	Wilcoxon

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Asesinos naturales descansando	Tumor	Control	0,000	0,000	Wilcoxon
Asesinos naturales activados	Tumor	Control	0,000	0,000	Wilcoxon
Monocitos	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos M0	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos M1	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos M2	Tumor	Control	0,000	0,000	Wilcoxon
Dendríticas descansando	Tumor	Control	0,000	0,000	Wilcoxon
Dendríticas activadas	Tumor	Control	0,763	0,760	Wilcoxon
Mastocitos descansando	Tumor	Control	0,000	0,000	Wilcoxon
Mastocitos activados	Tumor	Control	0,000	0,000	Wilcoxon
Eosinófilos	Tumor	Control	0,000	0,000	Wilcoxon
Neutrófilos	Tumor	Control	0,001	0,001	Wilcoxon

Grupo	Resultado Test Shapiro Multivariable
Tumor	No hay muestras suficientes para correr el test
Control	Normalidad FALSO

- Prueba de Shapiro-Wilk generalizada para normalidad multivariante de Villasenor-Alva y Gonzalez-Estrada.
- Datos: `as.matrix(data[data$Factor == levels(data$Factor)[i], 1:numbersOfImmneCells])`
- MVW = 0.8713, p-value < 2.2e-16

3.3 quantIsec

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Linfocitos B	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos M1	Tumor	Control	0,002	0,002	Wilcoxon
Macrófagos M2	Tumor	Control	0,000	0,000	Wilcoxon
Monocitos	Tumor	Control	0,410	0,410	Wilcoxon
Neutrófilos	Tumor	Control	0,000	0,000	Wilcoxon
Asesinos Naturales	Tumor	Control	0,001	0,001	Wilcoxon
Linfocitos T CD4 ⁺	Tumor	Control	0,000	0,000	Wilcoxon
Linfocitos T CD8 ⁺	Tumor	Control	0,602	0,600	Wilcoxon
T regulador	Tumor	Control	0,000	0,000	Wilcoxon
Dendríticas	Tumor	Control	0,994	0,990	Wilcoxon

- Prueba de Shapiro-Wilk generalizada para normalidad multivariante de Villasenor-Alva y Gonzalez-Estrada.
- Datos: `as.matrix(data[data$Factor == levels(data$Factor)[i], 1:numbersOfImmneCells])`.
- $MVW = 0.87262$, $p\text{-value} < 2.2e-16$

3.4 Epic

Población Inmune	Grupo 1	Grupo 2	p	p ajustado	Método
Linfocitos B	Tumor	Control	0,000	0,000	Wilcoxon
CAFs	Tumor	Control	0,000	0,000	Wilcoxon
T CD4 ⁺	Tumor	Control	0,000	0,000	Wilcoxon
T CD8 ⁺	Tumor	Control	0,091	0,091	Wilcoxon
Endoteliales	Tumor	Control	0,000	0,000	Wilcoxon
Macrófagos	Tumor	Control	0,000	0,000	Wilcoxon
Asesinos Naturales	Tumor	Control	0,414	0,410	Wilcoxon
Otras células	Tumor	Control	0,000	0,000	Wilcoxon

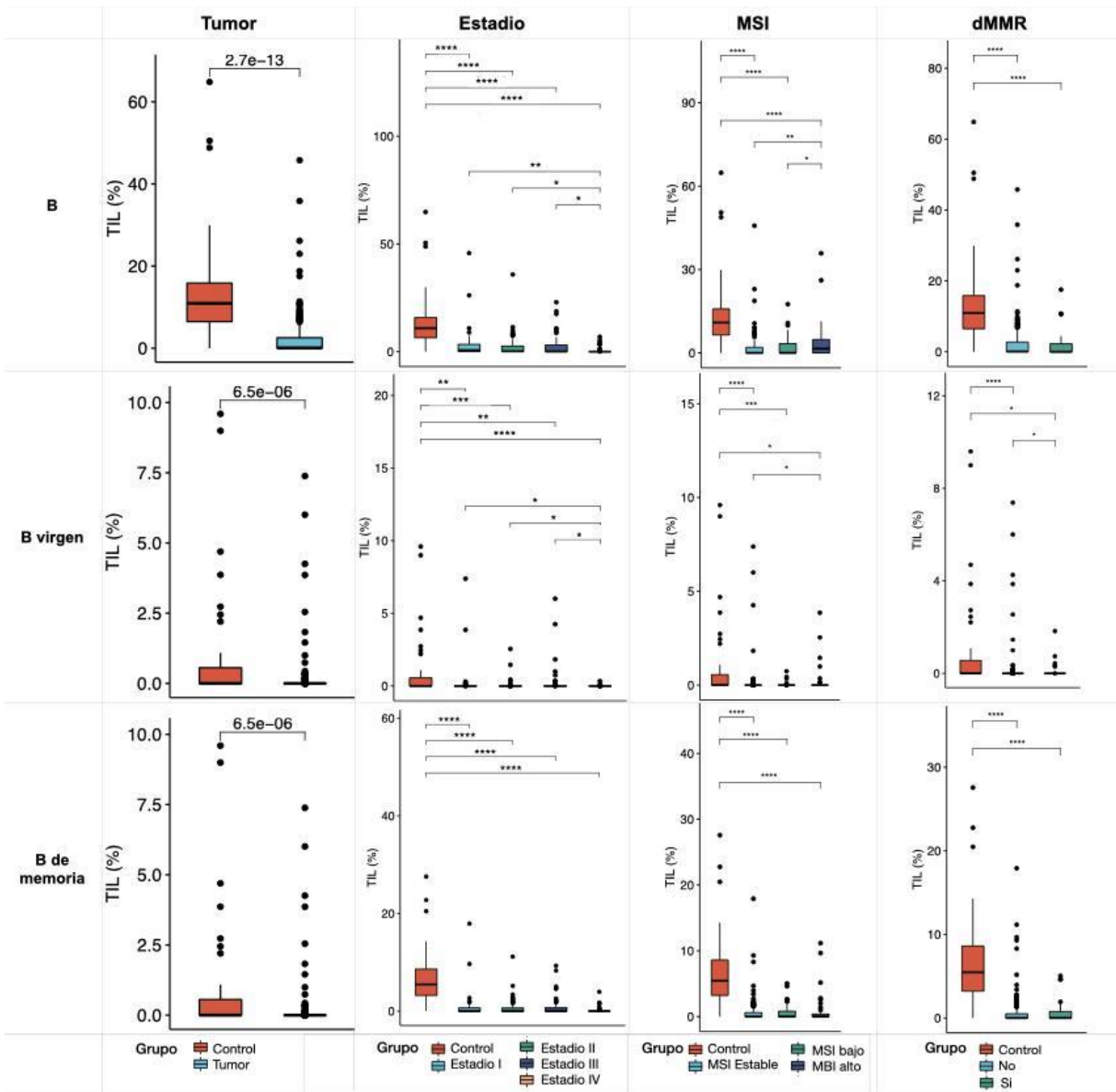
Grupo	Resultado Test Shapiro Multivariable
Tumor	Normalidad FALSO

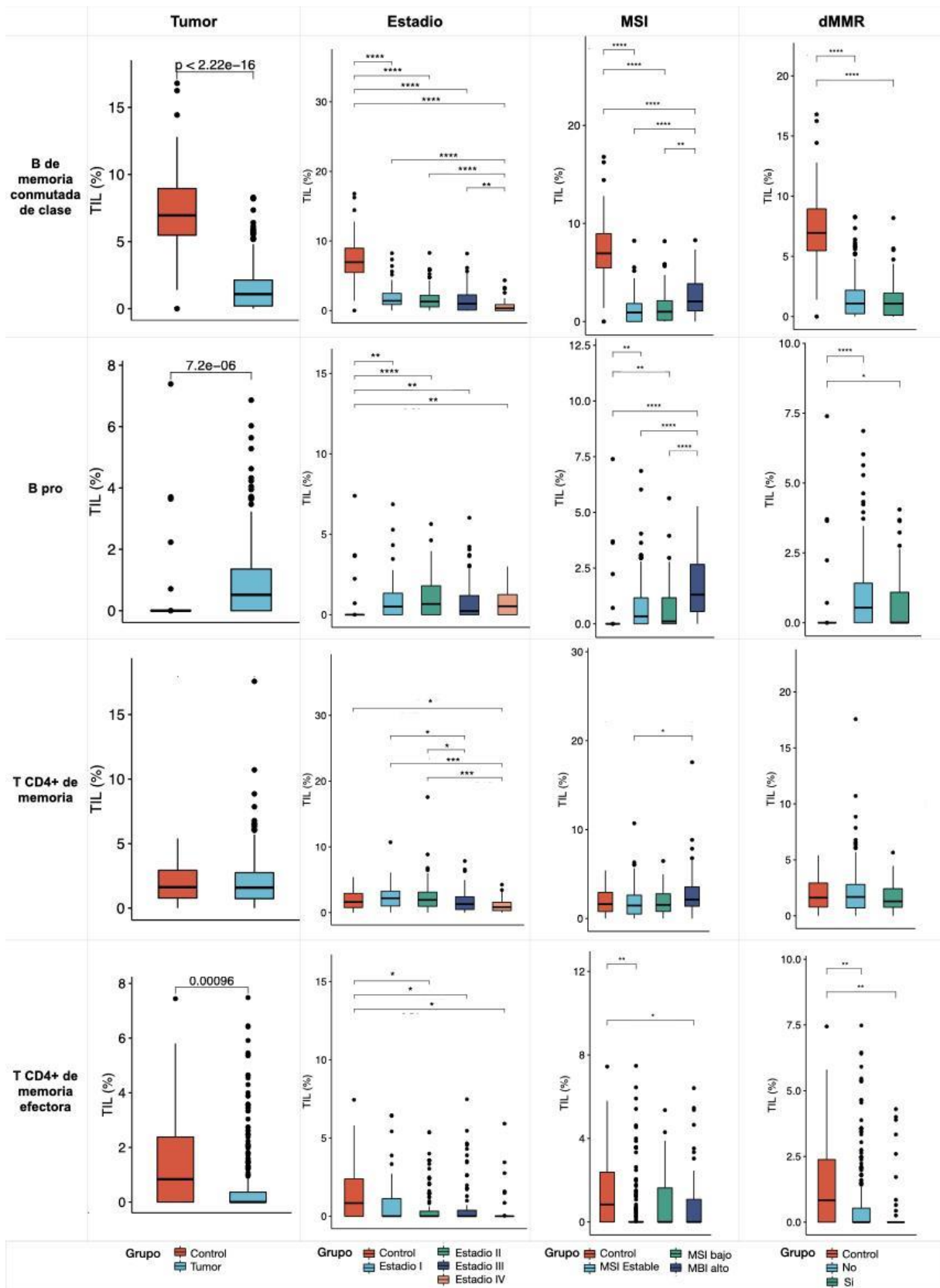
- Prueba de Shapiro-Wilk generalizada para normalidad multivariante de Villasenor-Alva y Gonzalez-Estrada.
- Datos: `as.matrix(data[data$Factor == levels(data$Factor)[i], 1:numbersOfImmneCells])`
- $MVW = 0.75983$, $p\text{-value} < 2.2e-16$

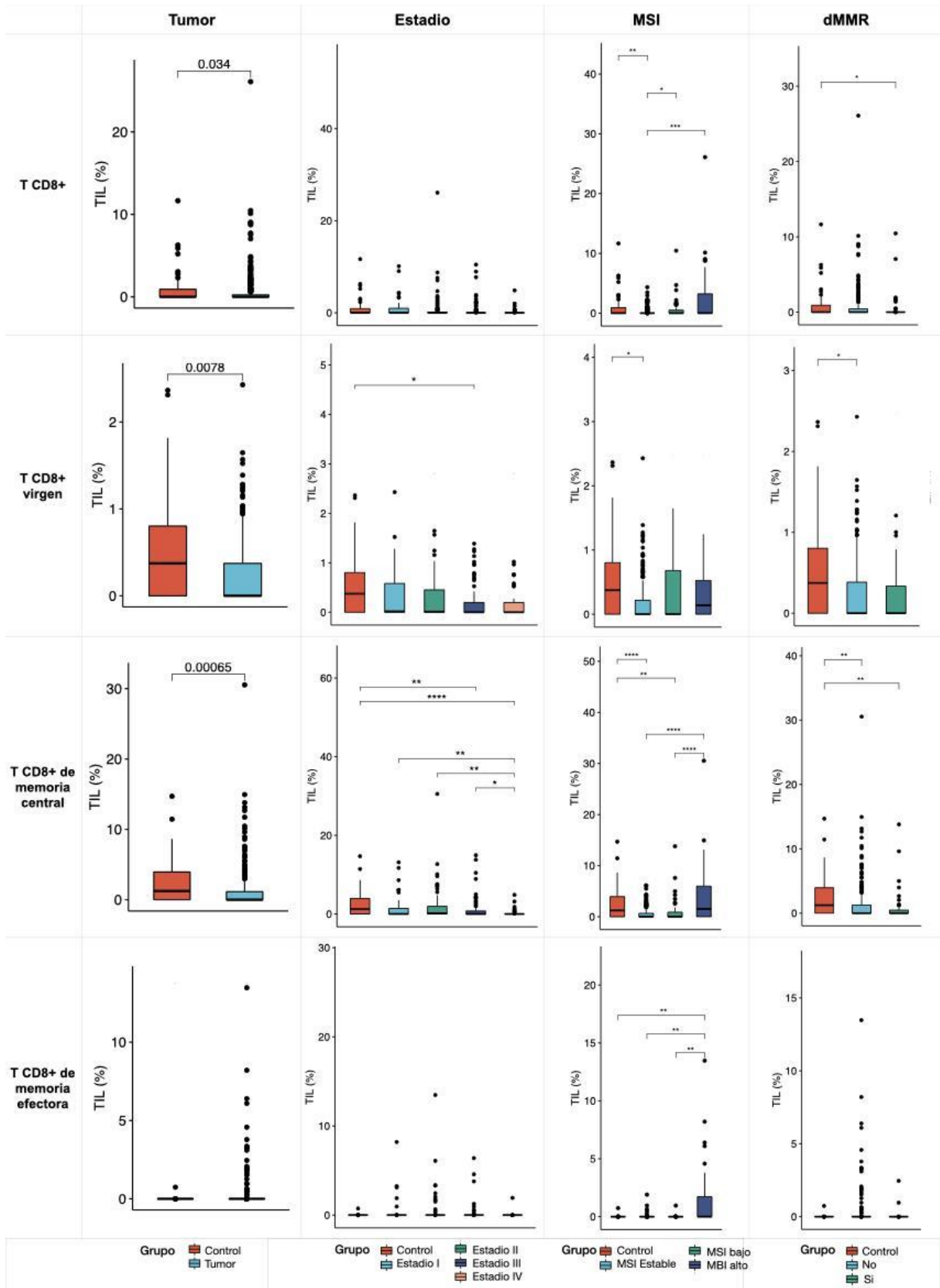
Grupo	Resultado Test Shapiro Multivariable
Control	Normalidad FALSO

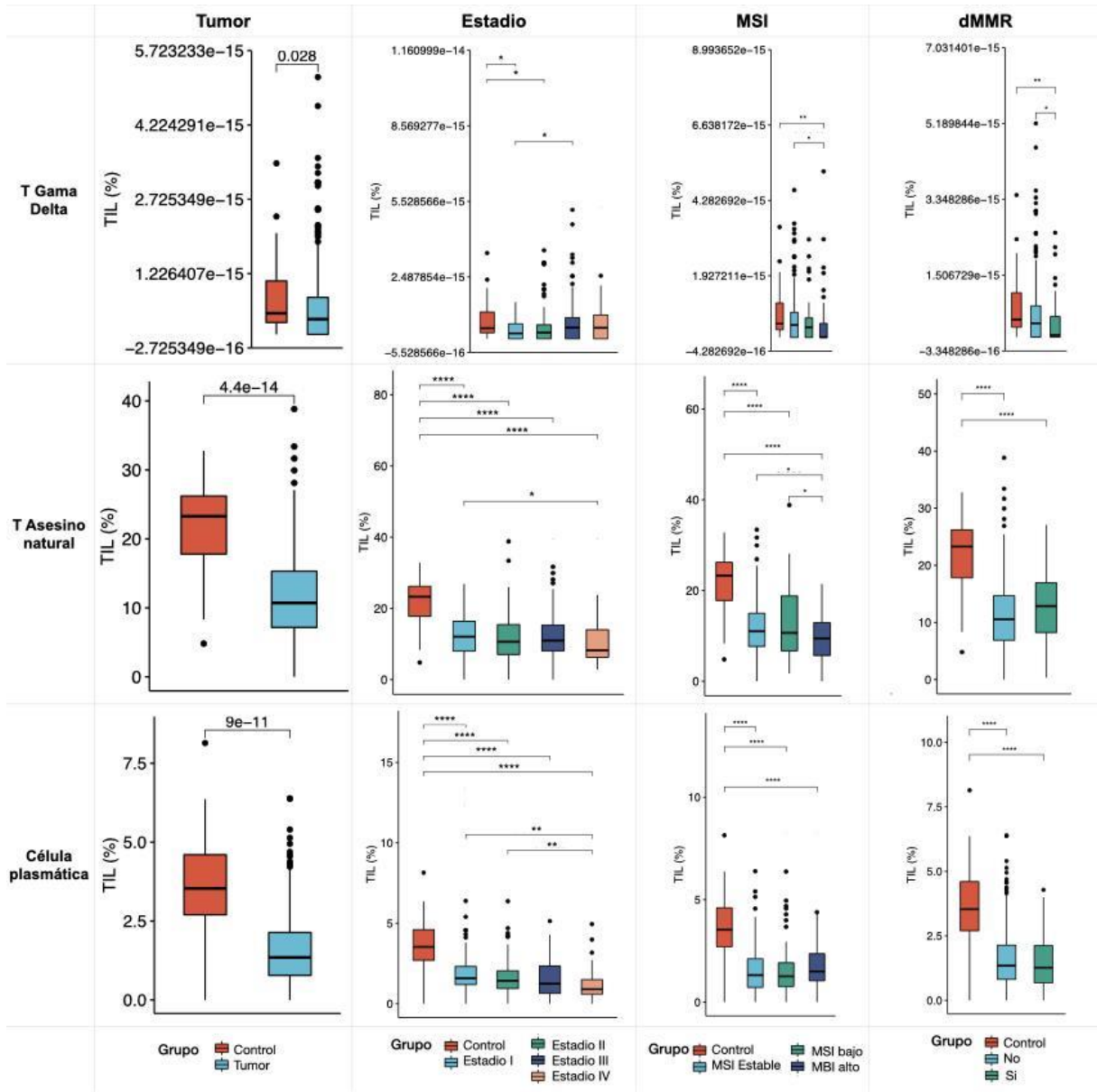
- Prueba de Shapiro-Wilk generalizada para normalidad multivariante de Villasenor-Alva y Gonzalez-Estrada.
- Datos: `as.matrix(data[data$Factor == levels(data$Factor)[i], 1:numbersOfImmneCells])`
- $MVW = 0.86978$, $p\text{-value} < 2.2e-16$

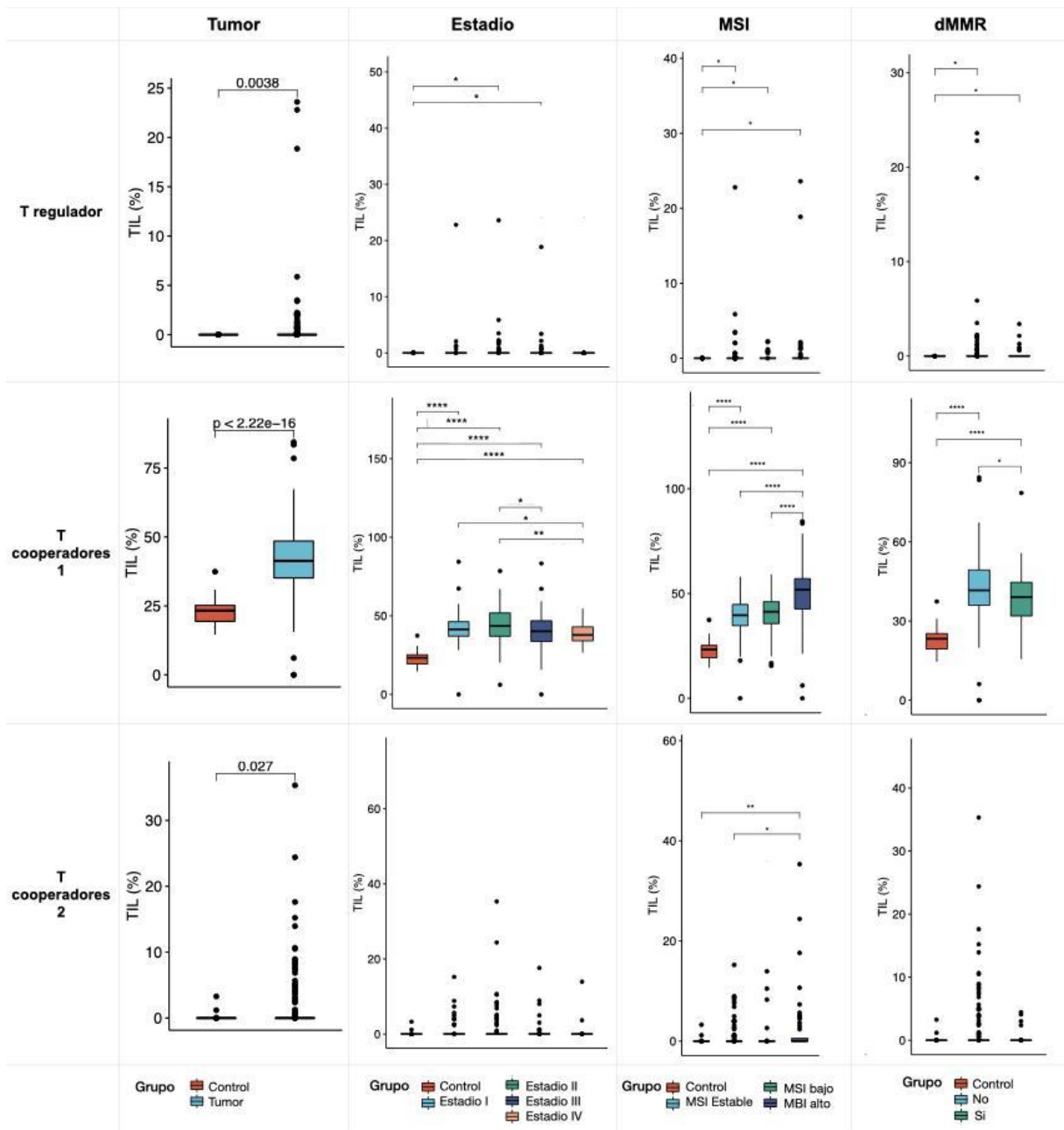
Anexo D: Caracterización del infiltrado tumoral en pacientes con CCR utilizando xCell para los grupos: tumor vs. control, estadio vs. control, MSI vs. control y dMMR vs. control.

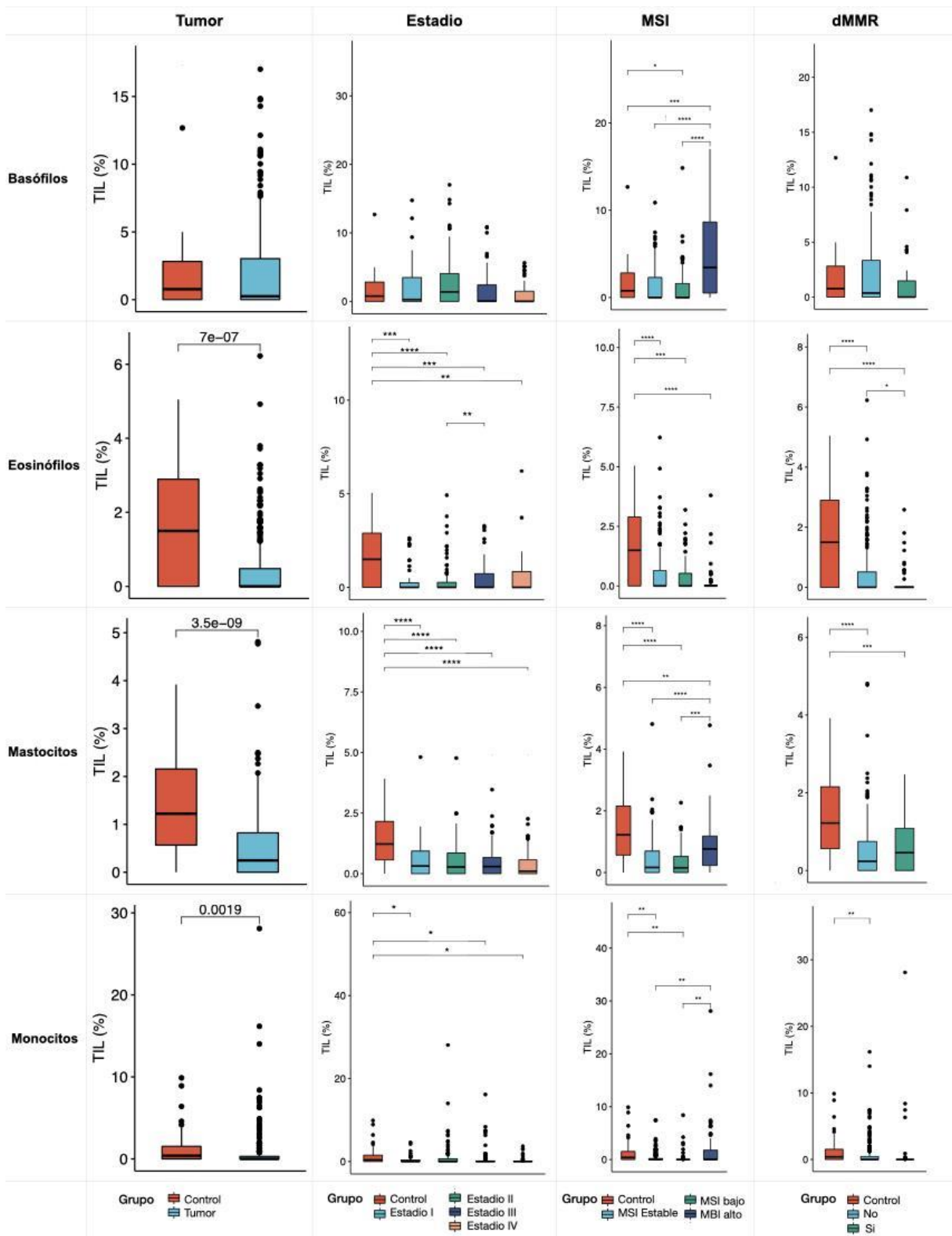


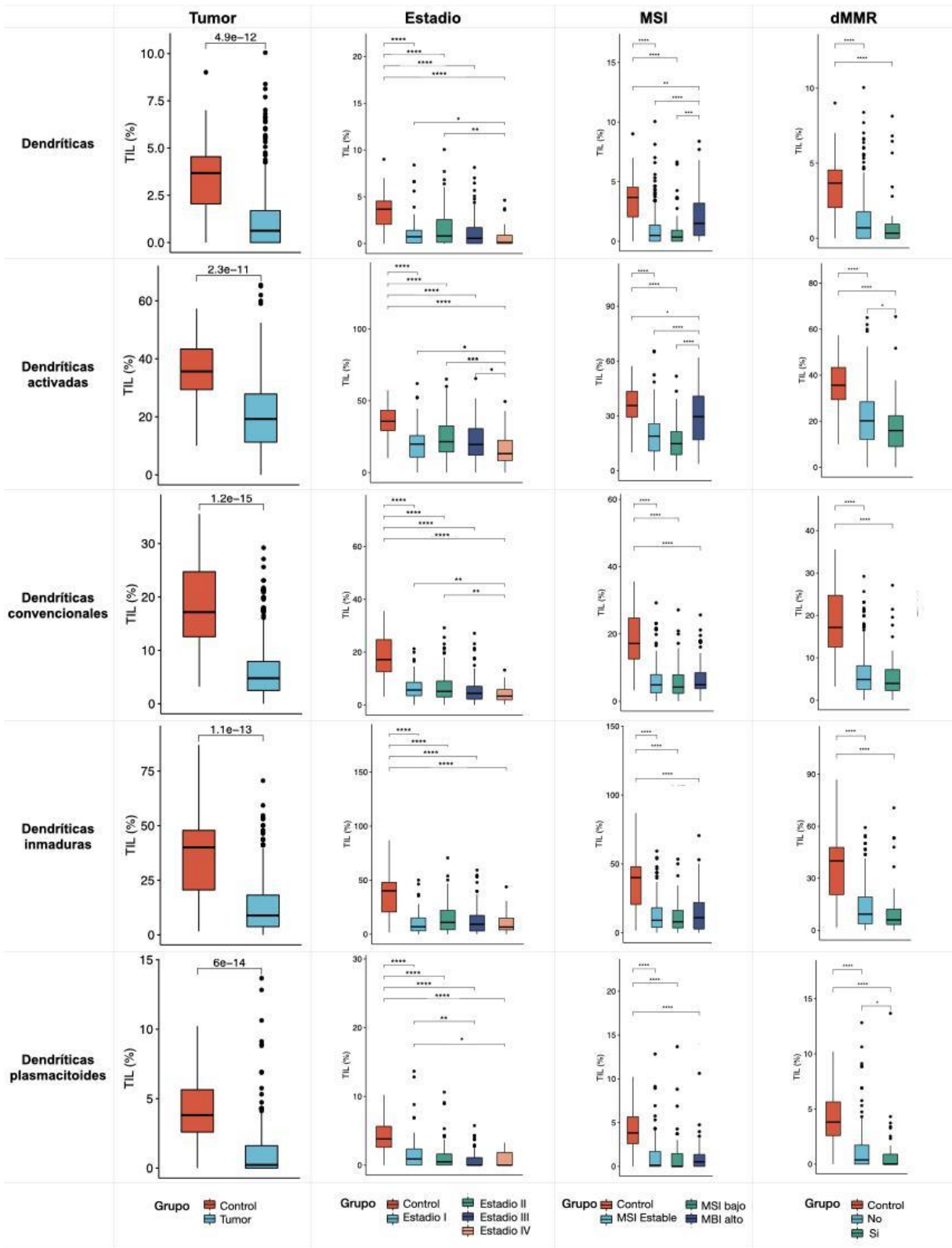


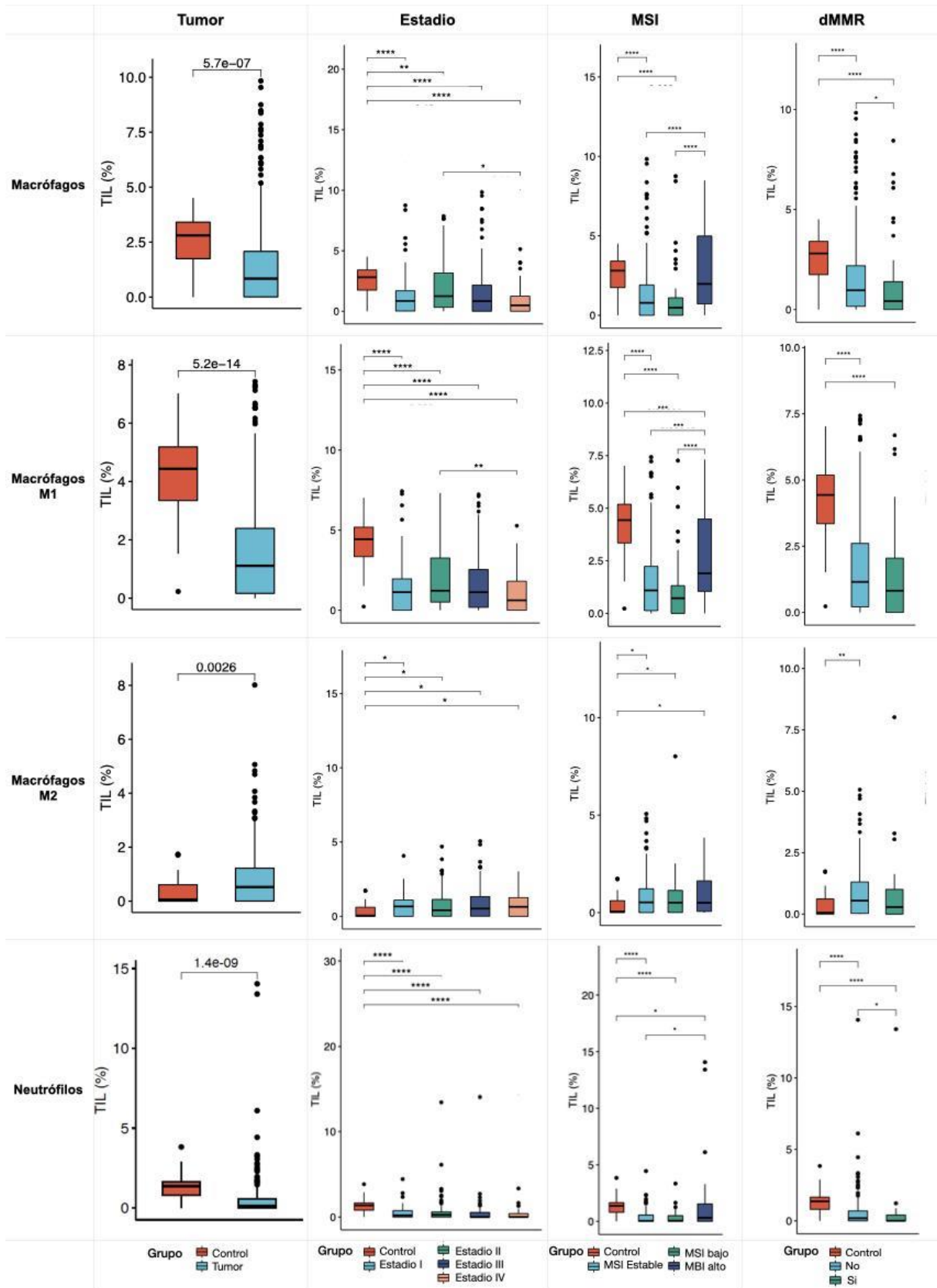






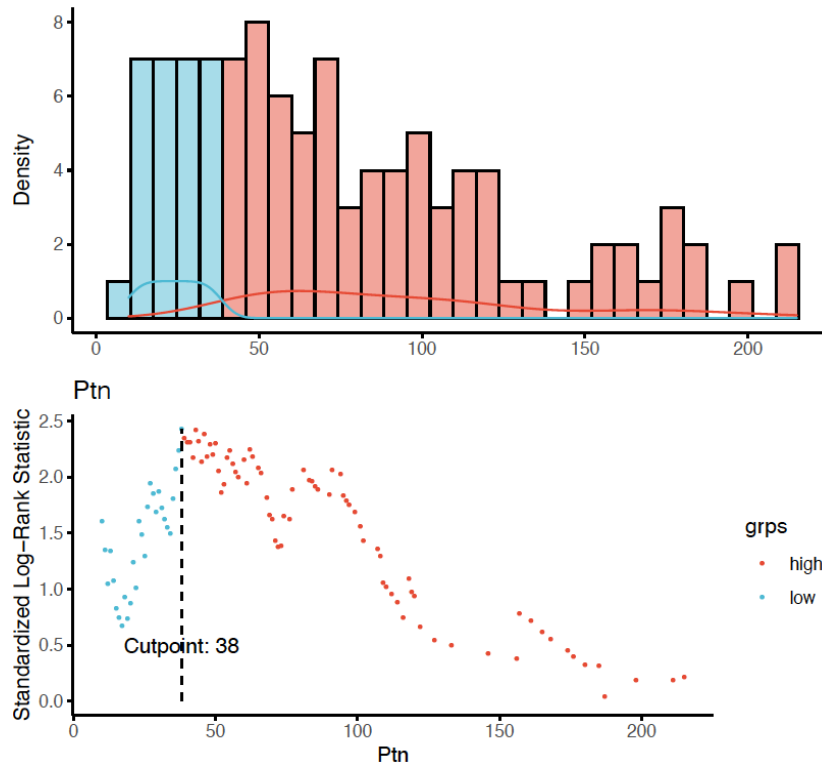




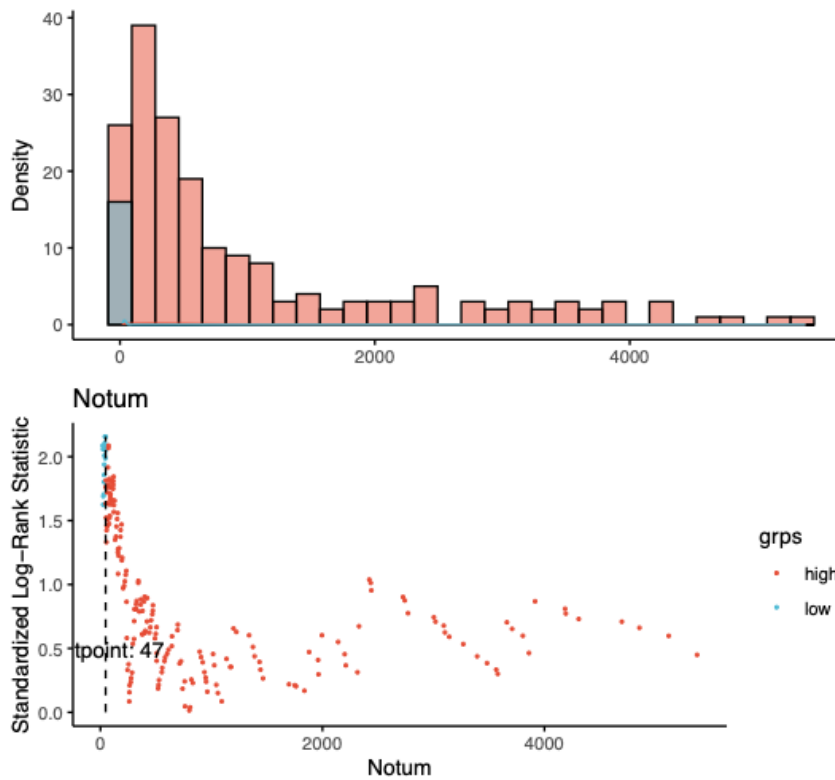


Anexo E: Conversión de variable cuantitativa a categórica.

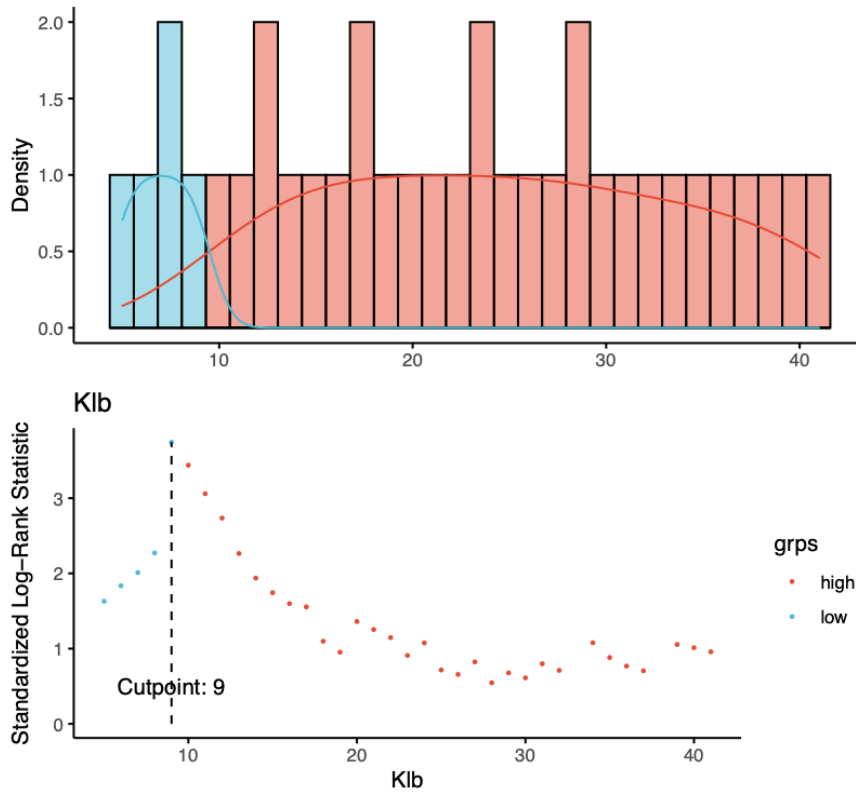
1. *PTN*



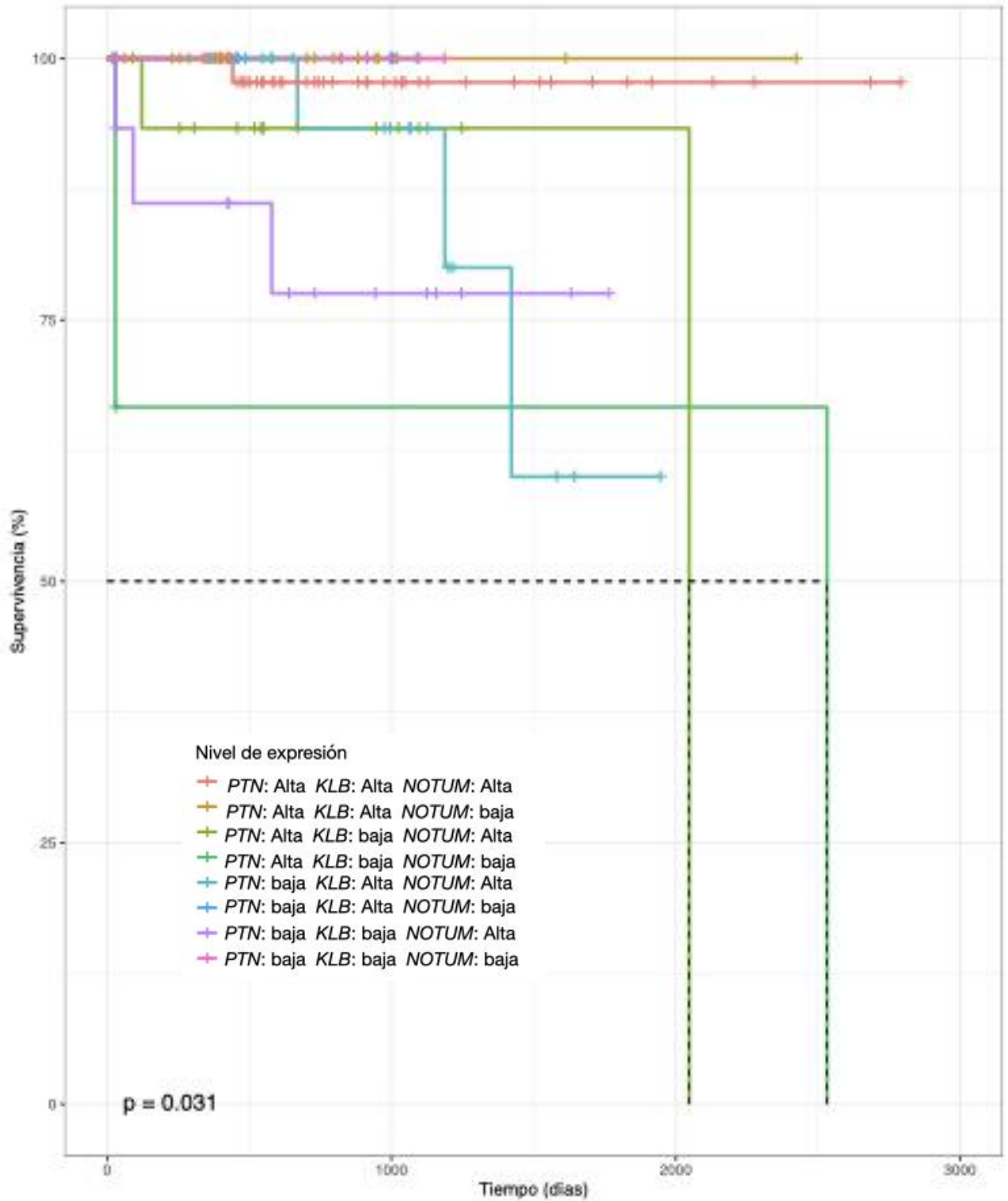
2. *NOTUM*



3. *KLB*



Anexo F: Curva de supervivencia en pacientes con CCR filtrado por estadio I y II para los genes *NOTUM*, *PTN* y *KLB*.



Gen	coef	exp(coef)	se(coef)	z	Pr(> z)	exp(-coef)	lower .95	upper .95
Ptn: alto Klb: alto Notum: bajo	-1.734e+01	2.948e-08	1.184e+04	-	0.9988	3.392e+07	0.0000	Inf
Ptn: alto Klb: bajo Notum: alto	2.350e+00	1.049e+01	1.235e+00	1.904	0.0570	9.533e-02	0.9328	117.97
Ptn: alto Klb: bajo Notum: bajo	2.848e+00	1.725e+01	1.359e+00	2.095	0.0361 *	5.799e-02	1.2020	247.44
Ptn: bajo Klb: alto Notum: alto	2.179e+00	8.838e+00	1.183e+00	1.842	0.0655	1.131e-01	0.8697	89.82
Ptn: bajo Klb: alto Notum: bajo	-1.737e+01	2.854e-08	2.173e+04	-	0.9994	3.504e+07	0.0000	Inf
Ptn: bajo Klb: bajo Notum: alto	2.808e+00	1.658e+01	1.183e+00	2.374	0.0176 *	6.033e-02	1.6324	168.34
Ptn: bajo Klb: bajo Notum: bajo	-1.737e+01	2.845e-08	2.105e+04	-	0.9993	3.514e+07	0.0000	Inf

- Concordance = 0.842 (se = 0.067)
- Likelihood ratio test = 15.38 on 7 df, p=0.03
- Wald test = 6.63 on 7 df, p=0.5
- Score (logrank) test = 15.44 on 7 df, p=0.03