



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

ANÁLISIS DE LA TRANSCRIPCIÓN DE LÍNEA GERMINAL (TLG) DEL
LOCUS IGH EN RESPUESTA A LA VACUNACIÓN Y SU RELACIÓN CON LA
MOVILIZACIÓN DE PLASMABLASTOS Y LA MEMORIA INMUNOLÓGICA

TESIS PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
HELENA KURI MAGAÑA

DIRECTOR DE TESIS
DR. JESÚS MARTÍNEZ BARNETCHE, INSP
COMIT. TUTOR
DRA. YVONNE JANE ROSENSTEIN AZOULAY, IBT, UNAM
DR. FERNANDO ROGER ESQUIVEL GUADARRAMA, FACULTAD
DE MEDICINA, UAEM

CUERNAVACA, MORELOS, DICIEMBRE, 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ÍNDICE

ABREVIATURAS.....	5
RESUMEN	6
ABSTRACT	7
IMPORTANCIA	8
INTRODUCCIÓN	10
ESTRUCTURA Y GENÉTICA DE LAS INMUNOGLOBULINAS	11
<i>Región variable</i>	13
<i>Región constante</i>	13
<i>Clases de anticuerpos y sus principales efectos biológicos</i>	14
ONTOGENIA DE LOS LINFOCITOS B	17
ACTIVACIÓN DE LINFOCITOS B MEDIADA POR EL ANTÍGENO T INDEPENDIENTE Y DEPENDIENTE	21
RECOMBINACIÓN DE CAMBIO DE CLASE.....	24
<i>Transcripción no codificante en la recombinación de cambio de clase</i>	25
LA RESPUESTA INMUNE A LA VACUNACIÓN	28
RESERVORIOS DE DATOS PÚBLICOS DE SECUENCIACIÓN MASIVA.....	30
<i>Sequence Read Archive (SRA)</i> :.....	31
<i>Genotype-Tissue Expression (GTEx)</i> :.....	31
<i>The Cancer Genomic Atlas (TCGA)</i> :	31
<i>recount2</i> :	32
PLANTEAMIENTO DEL PROBLEMA	32
OBJETIVOS	33
HIPÓTESIS	ERROR! BOOKMARK NOT DEFINED.
MATERIALES Y MÉTODOS	34
VACUNACIÓN DE VOLUNTARIOS SANOS	34
SEROLOGÍA	34
CUANTIFICACIÓN DE LOS PLASMABLASTOS POR FACS	34
ESTIMULACIÓN DE CÉLULAS B IN VITRO	35
qRT-PCR DE TRANSCRITOS DE RCCNC	35
DEFINICIÓN DE LÍMITES DE TRANSCRIPCIÓN RCCNC.....	36
CUANTIFICACIÓN DE TRANSCRIPCIÓN DE RCCNC Y CH	36
MAPEO DE ONTOLOGÍA DE MUESTRAS DE RNA-SEQ DE SRA	37
AGRUPAMIENTO DE PERFILES DE EXPRESIÓN	38
PRUEBAS DE ENRIQUECIMIENTO DE TEJIDO CON RESPECTO AL PERFIL TRANSCRIPCIONAL RCCNC	38
ANÁLISIS DE EXPRESIÓN DIFERENCIAL.....	38
RESULTADOS.....	39
MAPEO DE LA TRANSCRIPCIÓN DE RCCNC EN LINFOCITOS B HUMANOS AISLADOS:	39
CARACTERIZACIÓN DE LA TRANSCRIPCIÓN DE RCCNC IN VITRO Y EN RESPUESTA A LA VACUNACIÓN... ..	46
LA TRANSCRIPCIÓN RCCNC ES PREVALENTE EN EL SET DE RECOUNT2, Y PREDOMINA LA TRANSCRIPCIÓN DE I μ	50
EL SET DE RECOUNT2 SE AGRUPA EN PERFILES TRANSCRIPCIONALES BIEN DEFINIDOS	53
LA TRANSCRIPCIÓN RCCNC ESTÁ AMPLIAMENTE DISTRIBUIDA EN TEJIDOS SANOS Y SU PERFIL DEPENDE DEL TEJIDO	55
I μ SE EXPRESA EN PRECURSORES LINFOIDES TEMPRANOS E I γ 4 EN TEJIDOS FETALES	61
TRANSCRIPCIÓN DE RCCNC EN CANCER.....	63

UNA FIRMA TRANSCRIPCIONAL DE Iδ SE ASOCIA A LA ENFERMEDAD DE CROHN PEDIÁTRICA CON ULCERACIÓN PROFUNDA	68
DISCUSIÓN	72
CONCLUSIONES	79
REFERENCIAS	80
MATERIAL SUPLEMENTARIO	88

ÍNDICE DE FIGURAS

FIGURA 1.....	12
FIGURA 2.....	19
FIGURA 3.....	21
FIGURA 4.....	23
FIGURA 5.....	26
FIGURA 6.....	39
FIGURA 7.....	41
FIGURA 8.....	43
FIGURA 9.....	44
FIGURA 10.....	47
FIGURA 11.....	48
FIGURA 12.....	52
FIGURA 13.....	54
FIGURA 14.....	56
FIGURA 15.....	59
FIGURA 16.....	60
FIGURA 17.....	62
FIGURA 18.....	64
FIGURA 19.....	65
FIGURA 20.....	67
FIGURA 21.....	71

ÍNDICE DE TABLAS

TABLA 1.....	45
TABLA 2.....	51
TABLA 3.....	69

MATERIAL SUPLEMENTARIO

TABLA S1.....	90
FIGURA S1.....	91
TABLA S2.....	92
TABLA S3.....	93

Abreviaturas

AdCa: Adenocarcinoma

AID: Desaminasa de citidinas inducida por activación;

AML: Leucemia Mieloide Aguda

BCR: Receptor de células B

Bm: Células B de memoria;

Ca: Cáncer

CaSa: Carcinosarcoma

CG: Centro Germinal;

CH: Gen de la región constante de la cadena pesada;

cPLV: Células plasmáticas de larga vida;

DLBCL: Linfoma Difuso de Células B Grandes

EBV: Virus de Epstein Barr;

ELE: Estructuras Linfoides Ectópicas;

GTEEx: *Genotype-Tissue Expression Project*;

Ig: Inmunoglobulina;

I_H: exon-I;

MALT: Tejido linfoide asociado a mucosas;

MHC: Complejo mayor de histocompatibilidad

RCC: Recombinación de Cambio de Clase

RCCnc: Recombinación de Cambio de Clase no codificante;

RPKM: *Reads Per Kilobase (transcript) Per Million (reads)*;

SRA: *Sequence Read Archive*;

TCGA: *The Cancer Genome Atlas*;

TCR: Receptor de células T

TLG: Transcritos de línea germinal

Resumen

La recombinación de cambio de clase (RCC) en células B es un proceso que modifica la función efectora de los anticuerpos a través de la recombinación de ADN de la región constante de IgM por otro isotipo (IgG, IgA o IgE), para responder de la mejor manera ante un tipo de antígeno específico. La RCC depende de la transcripción no codificante (RCCnc) de exones I_H localizados 5' río arriba de cada gen codificante C_H . Desconocemos la importancia de la transcripción RCCnc en humanos sanos y en respuestas inmunológicas aberrantes. Para caracterizar la transcripción RCCnc en una gran variedad de condiciones de salud y enfermedad, mapeamos la transcripción de exones I_H humanos mediante el análisis de 70,603 muestras de RNA-seq públicamente disponibles, utilizando la herramienta informática *recount2*. Validamos la expresión de I_μ , $I_\gamma3$ e $I_\gamma1$ en sangre periférica de voluntarios sanos vacunados con Hepatitis B, Tétanos/Difteria, e influenza mediante qPCR. Caracterizamos, mediante minería de datos, la transcripción RCCnc en tejidos sanos y neoplásicos, así como en enfermedades infecciosas e inflamatorias. Identificamos la transcripción de 10 I_H en el locus *IGH* humano (chr 14q32.33).

La transcripción RCCnc está ampliamente distribuida en tejidos humanos sanos y depende de la abundancia de tejido linfoide secundario asociado. En la etapa fetal, $I_\gamma4$ es expresado de manera única en diferentes tejidos linfoides y no linfoides, sugiriendo un papel diferente a la RCC. La transcripción I_H en neoplasias tiende a la disminución con respecto a tejido sano, a excepción de cáncer mamario, pancreático, y testicular, sugiriendo cambios en el microambiente tumoral e inmunomodulación. Finalmente, identificamos “firma” transcripcional de I_δ and $I_\alpha2$ en enfermedad de Crohn, relacionada únicamente con ulceración profunda, sugiriendo un microambiente inflamatorio particular que promueve RCC a IgD+.

Nuestro análisis indica que la transcripción RCCnc es amplia y refleja el microambiente inmunológico en condiciones normales y su modificación como resultado del proceso patológico. Estos perfiles pueden servir como marcadores de la respuesta de células B en respuestas inmunes normales (vacunación) y en diversas patologías.

Abstract

Class switch recombination (CSR) in B cells is a process that modifies the effector function of antibodies through DNA recombination of the constant region of IgM to another isotype (IgG, IgA or IgE) to better respond to a specific type of antigen. CSR depends on non-coding transcription (CSRnc) of I_H exons located 5' upstream of each C_H coding gene. We ignore the importance of CSRnc transcription in healthy humans and in aberrant immune responses. To obtain deeper insight of CSRnc transcription in humans, we mapped human I_H exons by analyzing 70,603 RNA-seq publicly available samples, using the *recount2* software. We validated the expression of I_μ, I_{γ3} and I_{γ1} in peripheral blood of volunteers vaccinated with Hepatitis B, Tetanus/Diphtheria, and influenza by qPCR, and characterized, through data mining, CSRnc transcription in healthy and neoplastic tissues, as well as in infectious and inflammatory diseases.

Our approach allowed the identification of 10 I_H exons in the human *IGH* locus (chr 14q32.33). CSRnc transcription is widely distributed in healthy human tissues and depends on the abundance of associated secondary lymphoid tissue. In fetal stage, I_{γ4} is expressed uniquely in different lymphoid and non-lymphoid tissues, suggesting a different role than CSR. I_H transcription in neoplasms tends to decrease with respect to healthy tissue, with the exception of breast, pancreas, and testicular cancer, suggesting changes in the tumor microenvironment. Finally, we identified a transcriptional “footprint” of I_δ and I_{α2} in Crohn's disease, related exclusively to deep ulceration, suggesting a particular inflammatory microenvironment that promotes CSR to IgD+.

Our data indicates that CSRnc transcription is broadly distributed in human tissues and reflects the immune microenvironment in normal conditions and its modification as a result of the pathological process. These profiles can serve as markers for B cell response in normal immune responses (vaccination) and in various pathologies.

Importancia

Las enfermedades infecciosas y las neoplasias son un gran reto para la salud pública, ya que en nuestro país se encuentran entre las principales causas de mortalidad general. La vacunación es, y seguirá siendo, una de las estrategias preventivas de mayor impacto en la salud pública global, principalmente en los primeros años de vida y en ciertos grupos vulnerables. Aunque el mayor impacto de la vacunación tradicional ha sido en la prevención de enfermedades infecciosas, es de esperar que nuevos enfoques tecnológico-científicos permitan en un futuro próximo extender sus beneficios en la prevención de enfermedades crónico-degenerativas (Black et al., 2015; Rappuoli et al., 2014).

La base biológica de la vacunación yace en la memoria inmunológica, una propiedad del sistema inmune adaptativo de los vertebrados que confiere una capacidad cuasi infinita de reconocer patrones estructurales antigénicos “ajenos” o extraños, y responder más rápida y eficazmente ante un reencuentro con la misma estructura. La generación de memoria implica la selección de clones de linfocitos B y T específicos contra el antígeno, así como su maduración a células efectoras o de memoria de larga vida. En el caso de los linfocitos B, las células productoras de anticuerpos, la maduración funcional implica cambios a nivel molecular (por ejemplo, mutaciones) en la estructura del receptor de antígeno, seleccionados con base en la ganancia de afinidad y el cambio de clase. Esto repercute en el mejoramiento de su función ante reencuentros subsecuentes con el mismo antígeno.

El presente trabajo explora el mecanismo de recombinación de cambio de clase (RCC), mediante el cual los linfocitos B cambian su receptor de antígeno de IgM a IgG, IgA o IgE, cuyas propiedades funcionales son distintas a las originales y esenciales para la memoria inmunológica. El mecanismo de RCC es complejo y requiere de un paso de transcripción no codificante en el locus de la cadena pesada (*IGH*) de las inmunoglobulinas, necesario para desencadenar una serie de mecanismos enzimáticos de reparación de daño al ADN que permiten que el gen *IGHM* sea reemplazado por cualquiera de los demás genes de inmunoglobulina.

Aunque el mecanismo de RCC se conoce, los aspectos fundamentales de la transcripción no codificante asociada a RCC, referida en adelante como transcripción RCCnc, son aún desconocidas; en el tejido linfoide y no linfoide en condiciones de salud y en condiciones patológicas. Siendo un eslabón molecular del proceso de generación de memoria inmunológica, la transcripción RCCnc podría ser un biomarcador potencial de la respuesta inmune normal y/o patológica.

La revolución tecnológica del siglo XXI es la ciencia de datos. La capacidad para adquirir, almacenar, analizar grandes volúmenes de datos está revolucionando la astronomía, las ciencias sociales y desde luego las ciencias de la vida, a través de la genómica y la bioinformática (Stephens et al., 2015). Para abordar el estudio de la transcripción RCCnc en condiciones normales y patológicas, hemos llevado a cabo un estudio basado en la minería de datos públicos de transcriptoma en 70,603 muestras de origen humano, algunas sanas o con patología (de origen infeccioso o tumoral), incluyendo muestras derivadas de los consorcios The Genotype Tissue-Expression (GTEx) (Consortium, 2013) y The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research et al., 2013). Nuestra investigación es un ejemplo de análisis de “Big Data” genómico que permitió la identificación de las coordenadas de la transcripción RCCnc. Esta identificación incluye nuevos elementos transcripcionales y la caracterización de su expresión en más de 50 tejidos sanos derivados de más de 9,000 individuos; su caracterización en 33 tipos de cáncer y la identificación por primera vez, de una firma transcripcional de un exón no codificante previamente desconocido, en el locus de *IGHD* asociado con enfermedad de Crohn pediátrica con ulceración profunda.

Además, este trabajo cobra relevancia porque se conformó un grupo multidisciplinario de investigadores: inmunólogos, biólogos moleculares, bioinformáticos y bioestadísticos, gracias al cual se combinó el análisis de un proceso complejo y altamente relevante en términos de biomedicina traslacional, como lo es la inmunidad. Para esto, hicimos uso de estrategias de análisis de datos masivos derivados del transcriptoma humano, utilizamos datos públicos. Se estima que el volumen de los datos de secuenciación masiva se duplica cada 7 meses

(Stephens et al., 2015), y con la llegada de nuevas tecnologías como la transcriptómica de células individuales (Soon et al., 2013), la demanda de científicos de datos en el área de la biomedicina será enorme. De igual forma será la recompensa en cuanto a generación de conocimiento aplicable a la población.

Introducción

El sistema inmune orquesta la defensa del organismo contra agentes patógenos. El desarrollo de la inmunidad se da a través de dos tipos de respuestas: innata y adaptativa. La respuesta inmune innata es la primera línea de acción, ya que se basa en el reconocimiento de patrones estructurales comunes a diferentes patógenos (*pathogen associated molecular patterns*) y por lo general es una respuesta rápida, de especificidad genérica, de corta duración y estereotipada ante retos subsecuentes. A diferencia de la respuesta innata, la respuesta inmune adaptativa se basa en el reconocimiento de patrones estructurales específicos a cada patógeno, mediados por receptores de antígeno clonalmente distribuidos en linfocitos T y B. La respuesta adaptativa requiere más tiempo para desarrollarse, pero tiende a incrementar su intensidad y mejorar su calidad ante la exposición repetida a un mismo antígeno. Además, la respuesta inmune adaptativa tiene la capacidad de generar memoria a largo plazo.

Los componentes principales de la respuesta adaptativa son los linfocitos T y B antígeno-específicos. Los linfocitos T son activados a través del receptor de células T (TCR, por sus siglas en inglés), que reconoce un péptido antigénico mediante el complejo principal de histocompatibilidad (MHC) en células presentadoras de antígeno. Los linfocitos T CD4+ tienen un papel de gran importancia en la respuesta adaptativa, ya que su activación permite el desarrollo de mecanismos inmunes celulares mediante la acción citotóxica de los linfocitos T CD8+. Además, los linfocitos T CD4+ amplifican la respuesta humoral mediante la activación de linfocitos B y la actividad fagocítica de macrófagos, entre otras células del sistema inmune. Otra acción que llevan a cabo los linfocitos T es la de regular o controlar la respuesta inmune mediante los linfocitos T reguladores o Tregs.

Por su parte, los linfocitos B se caracterizan por reconocer el antígeno en forma directa a través del receptor de antígeno de linfocitos B (BCR, por sus siglas en inglés). Una vez reconocido el antígeno, las células se activan y diferencian a células plasmáticas secretoras de anticuerpos, que secretan el BCR como anticuerpos solubles; o a linfocitos B de memoria. El BCR se compone de una región Variable (V) en el N-terminal, y una región constante en el C-terminal. La región V determina la especificidad del anticuerpo, mientras que la región constante determina funciones efectoras. El ser humano cuenta con nueve tipos de regiones constantes o clases. Dependiendo de la clase, los anticuerpos secretados tienen diferentes funciones y capacidades, entre los cuales están, la neutralización, la opsonización y la fijación de complemento, entre otros.

Una de las características fundamentales del sistema inmune adaptativo es la memoria inmunológica, que se entiende como la capacidad de responder más eficientemente ante un estímulo antigénico subsecuente (respuesta secundaria). La memoria inmunológica es el fundamento biológico de la vacunación ya que el éxito de las vacunas depende de la capacidad de inducir la producción de anticuerpos neutralizantes que perduran por largos periodos de tiempo, simulando una respuesta secundaria.

Estructura y genética de las inmunoglobulinas

Los anticuerpos o inmunoglobulinas son moléculas que son las formas secretadas del receptor de células B, producidas por linfocitos B y células plasmáticas, que tiene la capacidad de unirse al antígeno de manera específica. Las inmunoglobulinas son glicoproteínas compuestas por una o más unidades que contienen dos cadenas pesadas (H) idénticas y dos cadenas ligeras (L) idénticas, unidas por puentes disulfuro. Cada cadena consta de una región variable en el amino terminal y una región constante en el carboxilo terminal, la cadena ligera solo tiene un dominio constante, mientras que la cadena pesada consta de 3 o 4 dominios constantes además de una región de bisagra (*hinge*), una porción de cadena peptídica que se caracteriza por ser muy flexible, la cual sirve de unión entre las dos cadenas pesadas por medio de enlaces disulfuro (**Figura 1**).

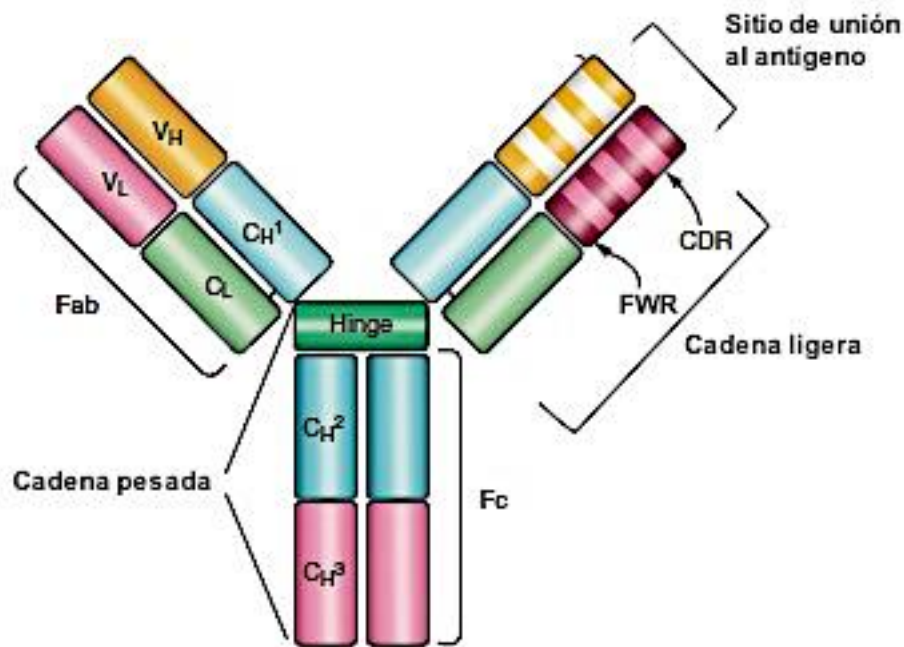


Figura 1. Estructura de la inmunoglobulina. La inmunoglobulina es una proteína que está compuesta por dos cadenas ligeras idénticas y dos cadenas pesadas idénticas. Cada cadena cuenta con una región variable en la cual se forma el sitio de unión a un antígeno específico. Las regiones variables tienen dominios hipervariables (CDR) así como regiones estructurales llamadas de andamio (FWR). Las cadenas ligeras contienen una región constante mientras que las cadenas pesadas tienen tres dominios. La región variable y el primer dominio constante de la cadena pesada y la cadena ligera conforman el fragmento Fab que tiene la capacidad de unión al antígeno. Los dominios CH₂ y CH₃ de la cadena pesada forman el fragmento Fc, el cual contiene la capacidad efectora de los anticuerpos (modificado de (Zou et al., 2017)).

A partir de los primeros experimentos realizados para estudiar la estructura de los anticuerpos, tratando IgG de conejo con enzimas proteolíticas, papaína o pepsina, se pudieron identificar regiones con diferentes propiedades funcionales y estructurales. Al digerir con papaína, que rompe los enlaces de la región de bisagra, se obtuvieron dos fragmentos idénticos individuales, formados por una cadena ligera completa VL-CL asociados a un fragmento de la cadena pesada VH- CH₁, responsables del reconocimiento y unión al antígeno, por lo que se conocen como fragmentos Fab. Cuando se digirió con pepsina se mantuvieron unidos ambos fragmentos Fab y se identificó como fragmento F(ab')₂. Por otro lado, existe otro fragmento que no tiene capacidad para reconocer el antígeno, pero tiene función

protectora y que está formado por dos péptidos idénticos que contienen dominios de la región constante de la cadena pesada C_{H2} y C_{H3} . Este fragmento tiene la capacidad de cristalizarse por lo que se le conoce como Fc (**Figura1**). La cadena pesada y la cadena ligera están codificadas por familia multigénicas, es decir en varios genes de cromosomas diferentes. La cadena pesada está codificada en el cromosoma 14, la cadena ligera kappa se encuentra en el cromosoma 2 y la lambda en el cromosoma 22, del genoma humano.

Región variable

La región variable forma parte de la cadena pesada y la cadena ligera que en conjunto forman la región de unión y reconocimiento del antígeno. La gran diversidad en la secuencia de aminoácidos que se encuentra en esta región es la que mayoritariamente permite generar el gran repertorio de anticuerpos que pueden reconocer una amplia gama de antígenos. La diversidad de la región variable está principalmente concentrada en tres segmentos hipervariables, conocidos también como regiones determinantes de complementariedad CDR1, CDR2 y CDR3, las cuales forman múltiples contactos con el antígeno. La región CDR3 es la que tiene mayor contacto con el antígeno y es la región más hipervariable. Además, se encuentran las regiones “*framework*” o andamio (FR1, FR2, FR3 y FR4) las cuales son menos variables en su secuencia y le dan estructura al dominio.

Región constante

La región constante es la porción que conforma los dominios de inmunoglobulina, cuya secuencia no varía o se encuentra muy conservada en diferentes clonotipos, además no participa en el reconocimiento del antígeno. La región constante de las cadenas ligeras consta de dos tipos de dominios diferentes que definen la cadena kappa (C_{κ}) y la cadena lambda (C_{λ}), que no presentan una función efectora específica más que la unión estructural con la región variable y tampoco parece haber diferencia de función entre ambas. En cambio, en la región constante de la cadena pesada se define la función efectora del anticuerpo. Existen 5 principales isotipos o clases de inmunoglobulinas IgM, IgD, IgG, IgE e IgA. Además de las principales clases de inmunoglobulinas, existen subclases que

varían entre especies. En humanos, IgG se subdivide en IgG1, IgG2, IgG3 e IgG4, mientras que IgA se subdivide en IgA1 e IgA2. Cada subclase está codificada por un gen independiente.

En el locus de la cadena pesada, los diferentes isotipos se expresan en genes separados localizados río debajo de la región variable. En las células B vírgenes se expresan únicamente IgM e IgD en membrana. IgM e IgD representan la unidad de reconocimiento antigénico que media la respuesta primaria del linfocito B contra el antígeno. Posteriormente dependiendo del tipo de antígeno que activa al linfocito B, se puede llevar a cabo el cambio de clase o isotipo de la inmunoglobulina a clases secundarias como son IgG, IgA o IgE. Una vez que los linfocitos han sido activados por el antígeno, aquellos que se diferencian a células plasmáticas, producen las formas secretadas de los anticuerpos.

Clases de anticuerpos y sus principales efectos biológicos

Los linfocitos B, como se ha mencionado anteriormente, son el elemento principal de la respuesta inmune humoral, ya que producen anticuerpos altamente específicos contra una gran diversidad de antígenos. Los anticuerpos se presentan de dos formas: los anticuerpos unidos a la membrana del linfocito B, el cual le permite el reconocimiento y la unión a un antígeno específico y, los anticuerpos secretados los cuales llevan a cabo diferentes mecanismos efectores. Los anticuerpos secretados contribuyen a la protección del organismo a través de tres acciones diferentes. La primera es la neutralización, donde los anticuerpos al unirse a ciertas estructuras del patógeno le impiden la interacción con su receptor celular y por lo tanto la infección de células blanco. La segunda acción es a través de la opsonización, en la que al unirse anticuerpos a la superficie de un patógeno promueven que sea fagocitado al unirse el receptor Fc de los fagocitos con la región constante del anticuerpo. Los anticuerpos también tienen la capacidad de activar a las proteínas del complemento, lo cual promueve la fagocitosis del organismo patógeno y, también pueden dirigir la lisis de células formando poros en la membrana. Además, los anticuerpos pueden mediar la acción citotóxica de las células asesinas naturales (NK) al recubrir la membrana de células tumorales o

infectadas, así como helmintos, las NK interactúan con los anticuerpos a través de la región Fc (Murphy et al., 2012).

Cada clase tiene una función determinada sobre la respuesta efectora. La IgD se encuentra en muy bajas concentraciones en suero (0.2% aproximadamente) y, se expresan junto a IgM en la superficie de las células B vírgenes. Su función en membrana parece estar relacionada con la activación de mecanismos tolerogénicos o señales apoptóticas (Mak, 2014). También se puede encontrar IgD secretada, la cual parece tener una función en la sangre y en las secreciones mucosas, así como en la superficie de células efectoras del sistema inmune como basófilos (Chen and Cerutti, 2011). La expresión de IgD se da por dos mecanismos: transcripción bicistronica seguida del splicing alternativo, o bien por recombinación. El cambio de clase a IgD mediante recombinación es un proceso poco común poco entendido. Se han detectado células B que han cambiado de clase a IgD en mucosas respiratorias, incluyendo amígdalas, cavidades nasales y glándulas salivales. La abundancia de células IgD+ en el tracto respiratorio superior y la capacidad de la IgD secretada de unir factores virulentos microbianos, además de bacterias y virus que afectan el sistema respiratorio, sugiere que IgD promueve la inmunidad en mucosas (Chen and Cerutti, 2011).

IgM es la primera inmunoglobulina que se expresa durante el desarrollo de las células B. Es el primer anticuerpo que se sintetiza en la respuesta primaria a un antígeno. Presenta una baja afinidad por el antígeno, lo cual le permite responder rápidamente a una gran variedad de antígenos. Se expresa en la superficie de linfocitos B vírgenes en forma monomérica. Representa entre el 5 y 10 % de las inmunoglobulinas en suero, en donde se agrupa en pentámeros generalmente. Tiene la capacidad de aglutinación y de neutralizar al antígeno, además de que activa al complemento por la vía clásica.

IgG es el isotipo o la clase predominante en suero (75 a 80%). Estas inmunoglobulinas se expresan como resultado de recombinación de cambio de clase durante la respuesta secundaria frente al antígeno. Se difunden con facilidad por lo que se pueden transportar a través de la placenta y también se encuentran

en el líquido extracelular. Son inmunoglobulinas muy eficientes en la neutralización de virus y toxinas.

Existen 4 subclases de IgG, IgG1, IgG2, IgG3 e IgG4. Las subclases se determinan por diferencias en su función, que se deben a variaciones estructurales en la región Fc así como en la región de bisagra. Cada subclase presenta una actividad efectora específica y puede o no llevar a cabo la activación del complemento. IgG1 e IgG3 responden principalmente a antígenos protéicos solubles o provenientes de membrana siendo IgG1 la subclase que se encuentra con más frecuencia en suero. IgG2 responde a antígenos de tipo polisacárido, por lo que son de gran importancia en infecciones bacterianas. IgG3 es el isotipo más eficiente en la activación de las funciones efectoras como la activación del complemento seguido de IgG1 e IgG2, siendo IgG4 la única subclase que carece de esta función (Vidarsson et al., 2014) . IgG4 es activada generalmente por alérgenos además de estar presente después de largas o repetidas exposiciones a antígenos no infecciosos se le relaciona con la activación de mecanismos de tolerancia ya que no se conoce que tenga función efectora mediada por Fc (Aalberse et al., 2009).

La inmunoglobulina de tipo IgA es la segunda más abundante en suero y la más abundante en secreciones corporales como la saliva, lágrimas, leche materna y moco del tracto genitourinario, gastrointestinal y respiratorio; ya que es la primera línea de defensa contra la invasión de patógenos ingeridos o inhalados. Debido a la gran concentración de IgA en calostro y leche, sugiere que tiene un papel importante en la inmunidad pasiva del recién nacido. Además, tiene un papel esencial en el mantenimiento del equilibrio de la microbiota comensal y neutralizan las toxinas. La IgA tiene una capacidad muy limitada para activar el complemento y la opsonización lo cual previene una respuesta inflamatoria exagerada a la microbiota en mucosas (Martinez-Barnetche, 2016). La IgA se subdivide en IgA1 que se encuentra en mayor concentración en suero y la IgA2 que tiene mayor concentración en leche. La regulación de ambas subclases se da de manera independiente.

Los anticuerpos IgE se encuentran principalmente en mucosas, por lo que junto a los mastocitos sirven para la defensa contra la infección por parásitos, especialmente helmintos. Además, la IgE está relacionada con procesos alérgicos, ya que puede activarse en la presencia de alérgenos sirviendo de unión entre la respuesta innata y la respuesta adaptativa, al interactuar con los mastocitos. Esto resulta en la liberación de histamina y citocinas al exterior como respuesta al alérgeno (Amarasekera, 2011).

Ontogenia de los Linfocitos B

Al igual que la mayoría de las células involucradas en la respuesta inmune, los linfocitos B derivan de células troncales hematopoyéticas (HSC) que se encuentran en la médula ósea, las cuales evolucionan y se diferencian en cinco principales estados de desarrollo: progenitores multipotentes (LMPP), que posteriormente pasarán a, progenitores linfoides comunes (CLP) que dan origen a las células NK, linfocitos T y linfocitos B. El microambiente de la médula ósea es de gran importancia para proveer de las citocinas necesarias para el desarrollo de los linfocitos. La presencia de estas citocinas así como los marcadores específicos de superficie, revelan la etapa de desarrollo de las células B. La producción de IL-7 por parte del estroma permite el desarrollo y diferenciación de las células B, en particular el proceso de generación de diversidad del receptor de antígeno (BCR) mediante el proceso de recombinación VDJ que ocurre en diferentes fases progresivas:

Linfocitos pro-B: Los linfocitos pro-B son la primera célula de médula ósea comprometida al linaje B, la cual expresa CD34, CD19 y CD10 en superficie. Aunque en esta etapa aún no se expresa el receptor de células B (BCR), comienza el reordenamiento de los segmentos VDJ para su posterior ensamblaje (Dorshkind and Rawlings, 2018).

Generación de repertorio pre-inmune: Durante esta etapa de desarrollo comienza la generación de la diversidad estructural en la región variable de los linfocitos B (repertorio de los linfocitos B), se da en médula ósea en linfocitos pro-B y pre-B. La región codificante de la región variable no está codificada en la línea

germinal, sino que se genera por recombinación somática de segmentos V, D y J o recombinación VDJ (**Figura 2**). Para llevar a cabo la recombinación VDJ se requiere un complejo de enzimas denominadas RAG1 y RAG2 (***Recombination Activating Gene***). RAG1 y 2 tienen actividad de endonucleasa, haciendo cortes de doble cadena en secuencias conservadas de 7 pb separadas por una secuencia menos conservada de 9 pb. Estas secuencias de reconocimiento son conocidas como las secuencias de señal de recombinación (RSS). Además, estas mismas secuencias favorecen el reclutamiento de enzimas necesarias para la reparación del ADN, la unión de los fragmentos y el rearrreglo final de los fragmentos que conforman a la región variable (Schatz and Ji, 2011).

Este proceso genera una inmensa diversidad combinatoria de los linfocitos ya que para la cadena pesada existen 39 genes funcionales V_H , 23 genes D_H y 6 genes para el segmento J_H , existen 40 genes V_L para la cadena ligera lambda, 32 para la cadena ligera kappa, 5 genes J para la ligera *kappa* y 4 en la cadena ligera *lambda*. Las cadenas ligeras carecen de segmentos D. En adición a la gran variedad de combinaciones entre los diferentes segmentos, la deoxinucleotidil transferasa terminal (TdT), que se expresa únicamente en linfocitos, puede agregar nucleótidos conocidos como nucleótidos N, que tampoco están codificados en la línea germinal, a los extremos de la cadena sencilla en la unión de los segmentos, lo cual incrementa aún más la diversidad del repertorio de los linfocitos B (Schroeder and Cavacini, 2010). El rearrreglo de la inmunoglobulina tiene jerarquía. El proceso se inicia en las células pro-B con la unión de los segmentos D_H - J_H , posteriormente se hace el rearrreglo de los segmentos V_H - DJ_H en las primeras etapas de las células pre-B, y se concluye con la unión de segmentos V_L - J_L (cadena ligera) en las últimas etapas de desarrollo de las células pre-B.

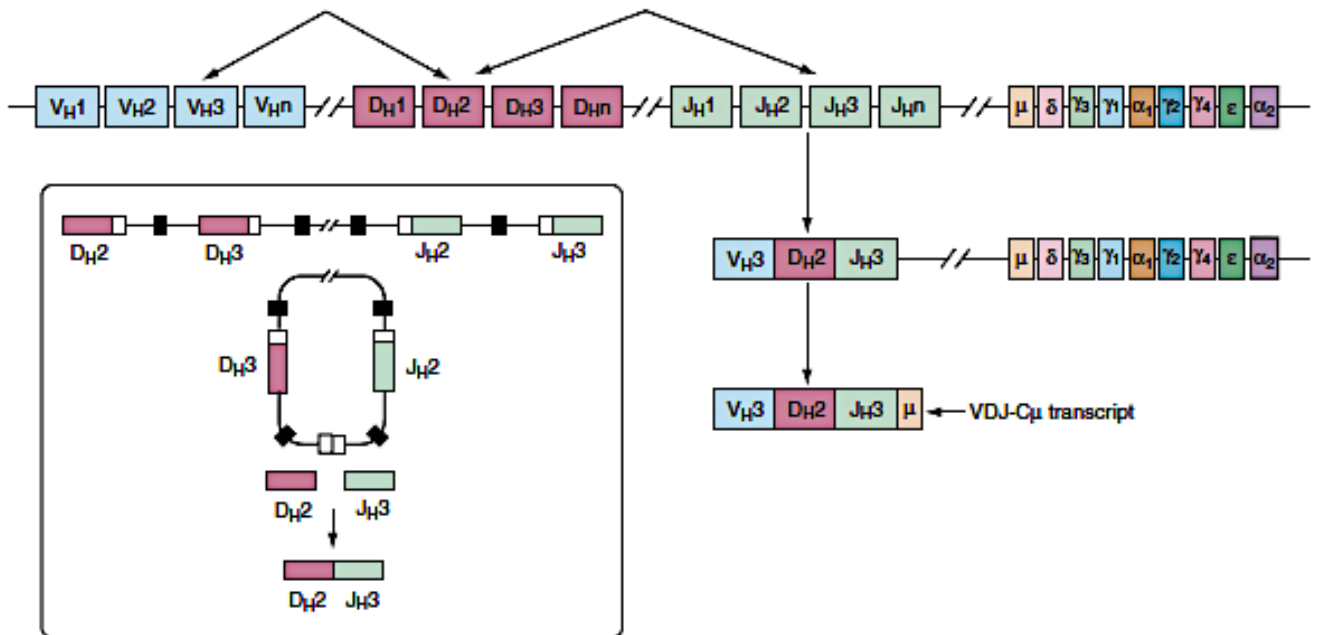


Figura 2. Segmentos V, D, J y la generación de la región variable codificante. Esquema simplificado de cómo se encuentran varios genes para los distintos segmentos V,D y J en la región variable de la cadena pesada y como se forma la región variable codificante al unir los genes de cada dominio con la región constante (Zou et al., 2017).

Linfocitos pre-B: En esta etapa de desarrollo los linfocitos han perdido la expresión de CD34. Las células pre-B tempranas expresan la cadena pesada que ha sido recientemente rearrreglada en sus segmentos VDJ y se determina si es funcional. La cadena pesada se ensambla junto con la cadena ligera subrogada para probar su activación y correcto funcionamiento, lo cual permite el ensamblaje del receptor de células pre-B (pre-BCR). Si el ensamblaje es funcional, el pre-BCR señala intracelularmente, al interactuar con ligandos de células estromales de la médula ósea, rescatando a la célula pre-B de morir por apoptosis lo que permite la continuación de su desarrollo y por lo tanto que el rearreglo de segmento V-J de la cadena ligera kappa se lleve a cabo. Cuando este es exitoso, se expresa el BCR clase IgM de superficie y las células se consideran entonces células B inmaduras. Si el ensamblaje de la cadena ligera kappa no es funcional, pueden ocurrir nuevos

rearreglos en el otro alelo de la cadena kappa, y en caso de no obtener un receptor funcional se procede al rearreglo de la cadena lambda (Melchers, 2015).

Linfocitos B inmaduros: Expresan la proteína IgM completa en membrana y son IgD -, no proliferan ni se diferencian en respuesta a un antígeno. Si reconocen con gran avidez algún antígeno en médula ósea ocurre el proceso de selección negativa, que culmina en la eliminación de clonas autoreactivas por apoptosis (tolerancia central), o bien se promueve la edición del receptor para evitar el reconocimiento de autoantígenos o para perder la capacidad de reacción frente al antígeno (anergia). Una vez que se ha llevado a cabo la selección negativa, las células B inmaduras migran como células de transición principalmente hacia bazo u otros órganos linfoides secundarios donde terminan su proceso de maduración.

Linfocitos B maduros: Los linfocitos maduros se pueden diferenciar en linfocitos B foliculares o linfocitos de zona marginal, dependiendo de las señales proporcionadas por el microambiente del bazo. Los linfocitos de zona marginal se encuentran en la zona marginal del bazo, pero también en el seno marginal de los ganglios linfáticos. Responden principalmente a antígenos de forma T-independiente. Producen grandes cantidades de IgM, son especialmente sensibles a los lipopolisacáridos ya que tienen una alta expresión del receptor tipo Toll CD180 (Ly64). Además, expresan el receptor CD1 lo que favorece la interacción con el sistema inmune innato, y el receptor del complemento CD21. Debido a esto, se les considera una primera línea de defensa que actúa entre la respuesta inmune innata y la respuesta inmune adaptativa (Baumgarth, 2004).

Por otro lado, los linfocitos foliculares se encuentran dentro de los folículos en los órganos linfoides secundarios como bazo y nódulos linfáticos y se asocian a las células dendríticas foliculares. Tienen la capacidad de recircular por los órganos linfoides secundarios para encontrarse con el antígeno y poder ser activados por los linfocitos T CD4+, lo que les permite convertirse en células B de memoria productoras de anticuerpo. Se caracterizan por expresar altos niveles de IgM e IgD además de CD23. Tienen una baja expresión de CD21 y no expresan CD1 (Allman and Pillai, 2008).

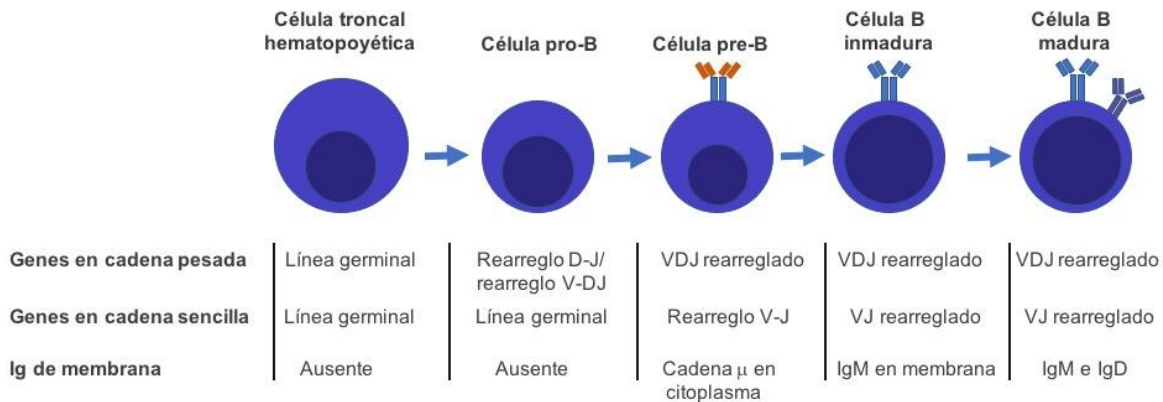


Figura 3. Ontogenia de células B. Los linfocitos B provienen de células troncales hematopoyéticas y pasan por varias etapas de maduración (Modificado de (Murphy et al., 2012)).

Activación de linfocitos B mediada por el antígeno T independiente y dependiente

Los linfocitos B maduros son activados a través de la unión de su receptor BCR con el antígeno y dependiendo del origen del antígeno la activación del linfocito B puede ser independiente o dependiente de la cooperación con células T. La activación tipo T-independiente se da en presencia de antígenos multivalentes, generalmente de tipo polisacáridos ya que, al presentar epítopos repetidos a lo largo de la membrana, tienen una alta capacidad de entrecruzamiento del BCR y generan las señales suficientemente fuertes para la activación de los linfocitos B, sin requerir de señales provenientes de las células T para activarse, lo que permite una respuesta relativamente rápida. Este tipo de activación se da afuera de los folículos linfoides principalmente en células B de zona marginal o B1. Algunas características de la respuesta T-independiente es la ausencia de memoria inmunológica, la falta de maduración de afinidad mediante la hipermutación somática y ausencia de cambio de clase. Sin embargo, se ha identificado recientemente la generación de memoria bajo ciertas circunstancias (Defrance et al., 2011; Obukhanych and Nussenzweig, 2006)

La respuesta T-dependiente se inicia cuando linfocitos B y linfocitos T interactúan en los límites del folículo y la zona de células T. El linfocito B es activado por el reconocimiento de antígenos de generalmente de tipo proteico, con secuencias de aminoácidos y estructuras únicas. El antígeno es endocitado y posteriormente degradado y así se facilita la presentación moléculas del antígeno por medio del complejo mayor de histocompatibilidad de tipo II (MHCII) a las células T cooperadoras que reconocen específicamente al antígeno. A su vez, los linfocitos T sintetizan moléculas unidas a membrana, así como secretadas que promueven la activación de los linfocitos B. El ligando de CD40 expresado en las células T, es uno de los elementos más importantes para la activación a través del receptor de CD40 de las células B. Además, los linfocitos T producen citocinas que promueven la proliferación y la activación del linfocito B como IL-4, IL-5, IL6, IL-21, IFN-g y TNF, entre otros.

En la fase temprana de la respuesta timo dependiente, los linfocitos B vírgenes IgM+/IgD+ recirculan entre los órganos linfoides secundarios donde tienen un primer encuentro con el antígeno e interactúan con células T CD4+, las cuales los estimulan para que se lleve a cabo una primera expansión clonal y la formación de células B plasmáticas secretoras de IgM. Sin embargo, algunos de estos linfocitos B pueden continuar su expansión clonal, generando así centros germinales (CG) dentro de los folículos de linfocitos B en los órganos linfoides secundarios. En el CG, conformado por una zona clara y una zona oscura, tiene un microambiente único que permite la generación de alteraciones en los anticuerpos que modifican la afinidad por el antígeno (hipermutación somática) y la función que llevan a cabo (recombinación de cambio de clase). Tras la estimulación antigénica dependiente de T, las células B vírgenes IgM+ se relocalizan a los folículos de células B, donde experimentan una expansión clonal dentro del CG dando lugar a la zona oscura. Uno de los procesos principales dentro del CG es la hipermutación somática, con la que se busca aumentar la afinidad del anticuerpo por el antígeno, a través de mutaciones, inserciones o pequeñas deleciones que se dan en motivos de la región variable de la inmunoglobulina. Los anticuerpos con alta afinidad por el antígeno son seleccionados, mientras los de baja afinidad entran a un proceso

conocido como maduración de afinidad donde vuelve a pasar por hipermutación somática frente a repetidas exposiciones al antígeno.

Otro de los procesos de gran importancia que se llevan a cabo dentro del centro germinal es la recombinación de cambio de clase de anticuerpos IgM a IgG, IgA o IgE, que favorece la diversificación de la función efectora de los anticuerpos, así como su localización anatómica particular y el cual será detallado a continuación.

La reacción de CG es un proceso iterativo de selección de mutaciones que, conduce a la generación de células B de memoria (Bm) antígeno-específicas de alta afinidad y células plasmáticas de vida larga secretoras de anticuerpos (cPLV) (revisado en (Victoria and Nussenzweig, 2012)) (**Figura 4**). Tanto las células Bm como cPLV son determinantes críticos en la eficacia de las vacunas (Yoshida et al., 2010).

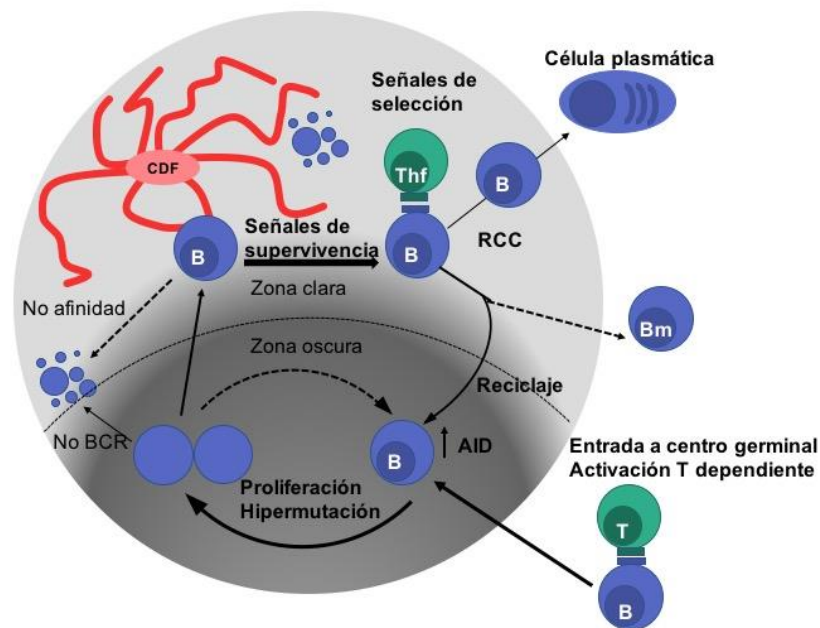


Figura 4. Generación de centro germinal. El centro germinal se forma por medio de la activación de linfocitos B, T-dependiente. Los linfocitos B proliferan y generan una zona oscura donde se transcribe la AID y se lleva a cabo la hipermutación somática para aumentar la afinidad por antígeno. Posteriormente interactúa en la zona clara, con las células dendríticas foliculares (CDF) donde se prueba la afinidad por el antígeno y si es exitosa hay señales de supervivencia y de selección que le permitirán convertirse en célula

plasmática o célula de memoria, según sea el caso.(Modificada de (Victoria and Nussenzweig, 2012)).

Recombinación de Cambio de Clase

El sello distintivo de la respuesta inmune adaptativa humoral es la producción de anticuerpos con alta afinidad por el antígeno y con cambio de clase a IgG, IgA o IgE, a partir de precursores vírgenes IgM⁺ de baja afinidad, lo cual se logra a través de maduración de afinidad y la recombinación de cambio de clase (RCC), lo que hace que este proceso sea central en la respuesta de la maduración de los anticuerpos frente a una infección natural o por vacunación.

La *desaminasa de citidina inducida por activación* (AID) es una enzima de vital importancia para la hipermutación somática y la recombinación de cambio de clase. Esta enzima se activa dentro del CG y es altamente mutagénica, por lo que está relacionada con el desarrollo de linfomas de células B y, se encuentra altamente regulada. AID se expresa en células B activadas de centro germinal principalmente, aunque también puede estar presente en células B foliculares. Como su nombre lo indica, tiene la capacidad de desaminar citidinas transformándolas en uracilo, una señal captada por el mecanismo de reparación de ADN de la célula y dependiendo de las enzimas reclutadas puede resultar en la hipermutación somática o el cambio de clase (Muramatsu et al., 2000).

Cada isotipo tiene una región constante diferente que le confiere una función efectora específica, así como la capacidad de reclutar efectores moleculares y celulares que llevan a cabo la respuesta específica al antígeno. El cambio de clase se realiza como respuesta al antígeno con la activación de células B. Al cambiar el isotipo del anticuerpo también cambia la función efectora, lo que permite mejorar la eliminación del patógeno que originó la activación.

La RCC requiere de la activación de linfocitos B por medio de BCR, CD40 y TLR, como estímulo primario; el cual induce la expresión de AID además de regular la expresión de factores de transcripción, como NFκB. Además, se requieren de estímulos secundarios, que, aunque no inducen la expresión de AID, dirigen y

promueven la transcripción no codificante relacionada a la recombinación de cambio de clase como IL-4 y TGF β . La señalización de TLR parece ser importante en respuestas T-dependiente y T-independiente (Xu et al., 2012).

La RCC implica un proceso de delección y recombinación que se lleva a cabo entre segmentos repetidos de ADN llamados regiones de “*switch*”, que se encuentran río arriba de cada uno de los ocho genes codificantes C_H en humano, excepto para IgD donde se encuentra una región de secuencias repetidas poco caracterizada denominada $\sigma\delta$ localizada entre C μ y C δ . La delección se lleva a cabo desde un segmento del intervalo codificante IgM / IgD (C μ -C δ) al flanco río arriba de alguno de los genes *IGHG*, *IGHA* o *IGHE* en la región telomérica del cromosoma humano 14 (**Figura 5**). Las células B vírgenes tienen la capacidad de cambiar a cualquier isotipo. La especificidad del isotipo al que va a cambiar depende de la inducción de la transcripción que comienza en la región del promotor y donde su activación es dependiente de las citocinas que promueven la expresión de distintos factores de transcripción que se unen a los promotores iniciando así la transcripción dirigida.

Transcripción no codificante en la recombinación de cambio de clase

La mayoría de los estudios y el conocimiento generado sobre los mecanismos y los transcritos no codificantes relacionados a la RCC, se han realizado principalmente en modelos murinos, lo que deja ciertas incógnitas sobre el mecanismo en humanos. Uno de los elementos necesarios para iniciar la RCC es el ADN de cadena sencilla, el cual es el sustrato de la AID. Por lo tanto, el inicio de la RCC depende de la transcripción no codificante que inicia en los exones I_H conocidos como transcritos de línea germinal (TLG) o "estériles" (en lo sucesivo, en este escrito se nombrarán como transcritos RCCnc). Los exones I_H están ubicados en la región 5' de cada gen C_H y comprenden solo un fragmento de lo que son los transcritos RCCnc ya que la transcripción no codificante se extiende de los exones I_H a la región S y C_H, se acopla a la remodelación de la cromatina y es dependiente de *splicing* (Chowdhury et al., 2008; Lorenz et al., 1995).

La transcripción no codificante de la región donadora ($I\mu$ - $C\mu$) se encuentra constitutivamente activa. La transcripción de línea germinal inicia de manera específica en el promotor de la región aceptora, el cual es activado por distintas citocinas que activan la transcripción de factores de transcripción que dirigen la RCC a un isotipo específico, dependiendo del origen del antígeno que activa la respuesta. Un ejemplo de esto es la IL-4 que activa al factor de transcripción STAT6 (*signal transducer and activator of transcripton 6*), el cual se une a los promotores de $I\gamma 1$ e $I\epsilon$ promoviendo así el cambio de clase a IgG1 e IgE. Esta transcripción permite la formación de estructuras híbridas de ARN:ADN como los R-loops, lo cual permite que la región de switch quede expuesta como ADN de cadena sencilla (revisado en (Yewdell and Chaudhuri, 2017)).

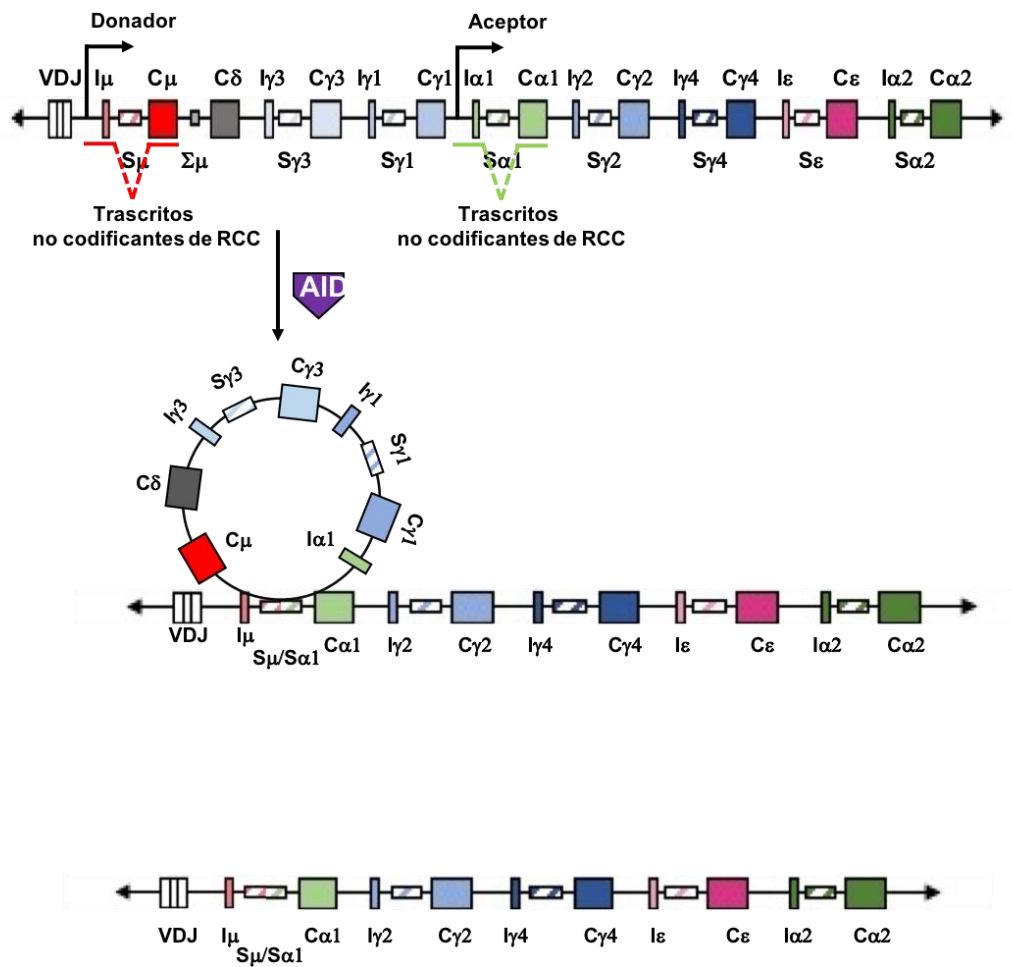


Figura 5. Recombinación de cambio de clase. Esquema que muestra el locus de región constante de la cadena pesada en humanos. Se muestran los dominios presentes en cada isotipo, exón I (I_H), región de switch (S_H) y región codificante (C_H). La AID permite el inicio de la recombinación de cambio de clase a través de la deaminación de la citidina en regiones repetidas CG en la región de switch, iniciando así la recombinación del locus de la cadena pesada y la escisión de una parte del genoma (modificada de (Xu et al., 2012).

Para que el proceso de delección y recombinación se pueda llevar a cabo se requiere de la expresión de AID, la cual desamina las citidinas en las regiones de *switch* (S) ricas en G situadas río arriba de cada gen codificante de la cadena constante de inmunoglobulina (C_H). La AID reconoce motivos WGCW que se encuentran en gran densidad dentro de las regiones de S y se sabe que, por su secuencia y estructura únicas, las regiones de switch sirven como blancos para el proceso de recombinación de cambio de clase. La desaminación de citidinas genera desoxiuracilos los cuales son reconocidos y removidos por la endonucleasa de uracilo de ADN UNG, que es parte importante de la vía de reparación de ADN por escisión de base (BER). Estos sitios donde se ha eliminado la base se escinden por medio de la endonucleasa apurinica /apirimidinica (APE), lo que finalmente conduce a cortes del ADN de cadena sencilla tanto en el donador (S_μ) como en la región S aceptora correspondiente. Los cortes de doble cadena se pueden dar por medio de dos mecanismos: i) uno puede ser debido a la proximidad con la cadena complementaria o bien ii) un mecanismo dependiente de MSH2, MSH6 y exonucleasa 1. Los cortes de doble cadena se resuelven al reunir los extremos cromosómicos mientras que el segmento C_μ - C_δ eliminado se recirculariza en un episoma no replicante mediante, religación de extremos no homólogos (NHEJ) (Xu et al., 2012).

La actividad de AID fuera del objetivo (*off-target*) está implicada en la generación de neoplasias de células B, pero el mecanismo preciso por el cual la actividad de AID se limita al locus *IGH* es aún desconocida (Alt et al., 2013; Yewdell and Chaudhuri, 2017).

La RCC, es un proceso celular complejo que ocurre en microambientes especializados, en los órganos linfoides secundarios y terciarios. La elección celular sobre cual I_H se transcribirá y, en consecuencia, la clase de Ig a la que va a cambiar, está influenciada por la disponibilidad de ciertas citocinas, como IL-4, IFN_γ , TGF_β y PAMP's, entre otras moléculas de señalización. Tales señales ambientales activan vías intracelulares que promueven la transcripción selectiva de un exón I_H dado, guiando la RCC de acuerdo con un microambiente particular o un insulto patogénico (Xu et al., 2012).

La RCC, es un proceso que se da como respuesta a la presencia de un antígeno, que activa a las células B y posteriormente genera células B de memoria. La generación de la memoria inmunológica, que resulta en una respuesta más rápida y eficiente ante un mismo antígeno es lo que se busca con la vacunación. Debido a esto, la RCC es un proceso que permite aumentar las probabilidades de una buena respuesta y por lo tanto de una mejor protección ante el antígeno deseado y una vacunación más exitosa.

La respuesta inmune a la vacunación

La vacunación ha demostrado ser una intervención eficaz para la prevención de enfermedades infecciosas. Vacunas contra viruela, varicela, sarampión, tétanos, poliomielitis, etc., han resultado tener un efecto protector para la población, conteniendo la expansión de estas enfermedades y en el caso de la viruela, lograr erradicarla por completo. Lo que se busca con la vacunación es que, frente a una exposición subsecuente a un mismo antígeno introducido en la vacuna, se presente una respuesta inmune más rápida y eficiente, lo que le confiere al individuo protección contra el patógeno. Además del linfocito B antígeno-específico, el desarrollo de memoria inmunológica mediada por anticuerpos es dependiente de la co-estimulación por parte de linfocitos T antígeno-específicos.

La base de la vacunación es introducir patógenos atenuados, inactivados o fragmentos de patógenos, en un individuo; para así estimular su sistema inmune. La primera respuesta ante este estímulo se da a través del sistema inmune innato, el cual se encarga de marcar, procesar y presentar al antígeno a los linfocitos por

medio de macrófagos y células dendríticas. Dependiendo de su origen, el antígeno desata una respuesta celular, la cual elimina a las células infectadas a través de células CD8+ o NK's o bien, promueve una respuesta de anticuerpos, resultante de que células B específicas identifican al antígeno y montan una respuesta que implica una expansión clonal de las células B que a su vez se convertirán en células de memoria o en células plasmáticas productoras de anticuerpos específicos contra el antígeno. En el caso de algunas vacunas, la protección depende de la inducción de la producción y persistencia de anticuerpos específicos contra la vacuna, así como del desarrollo de memoria inmunológica, de tal forma que linfocitos B antígeno-específicos puedan ser reactivados efectiva y rápidamente ante una exposición frente al agente patógeno (Pulendran and Ahmed, 2011).

A pesar de que las vacunas representan una solución exitosa para el control de enfermedades infecciosas, la gran mayoría de las vacunas han sido desarrolladas de manera empírica y sin conocimiento claro acerca de los mecanismos inmunológicos por medio de los cuales confieren protección. Hasta el momento no ha sido posible desarrollar vacunas eficaces para combatir a las infecciones por VIH, tuberculosis y malaria. Actualmente, el desarrollo de nuevas vacunas enfrenta grandes retos ya la producción requiere de una infraestructura robusta que cumpla con las exigencias cada vez mayores de la normatividad sanitaria, para proporcionar productos efectivos y seguros. La efectividad de las vacunas es probada únicamente a través de ensayos clínicos, lo que ocasiona que el proceso dure varios años y sea muy costoso. Otro punto limitante en el desarrollo de vacunas es la falta de una identificación temprana de aquellas vacunas que podrían ser candidatos prometedores. Por lo tanto, la identificación de “marcadores” moleculares tempranos con carácter predictivo de éxito de las vacunas podría acelerar la producción e incorporación de nuevas vacunas eficaces y seguras en la población podría acelerarse mediante. (Li et al., 2014; Pulendran and Ahmed, 2011).

El reciente avance en las ciencias genómicas y ciencias de la complejidad ha permitido el desarrollo de nuevos enfoques para entender procesos biológicos inherentemente complejos. Los enfoques de biología de sistemas han sido

utilizados para entender la respuesta inmune protectora y así diferenciarla de la no protectora, revelando las bases moleculares y celulares de tales diferencias, y proporcionando información para el desarrollo de nuevas vacunas (Clem, 2011; Plaut et al., 2009). Así, por ejemplo, mediante el análisis de citometría de flujo multiparamétrica, el análisis transcripcional a escala genómica y la cuantificación de citocinas séricas en individuos vacunados contra fiebre amarilla se identificaron ciertas “firmas” transcripcionales al día 7 post-vacunación, asociados a la función de linfocitos B y que correlacionan con la intensidad de la respuesta a la vacuna. En función de estos resultados se propuso que la cantidad de plasmablastos en sangre periférica en respuesta a una vacuna, puede ser una medida para evaluar la capacidad de la vacuna para inducir una respuesta inmune de memoria adecuada (Wrammert et al., 2012; Wrammert et al., 2008).

Para poder comprender y analizar “firmas” moleculares en diferentes tipos de vacunación es necesario tener cohortes suficientemente grandes para poder identificar cambios relevantes. Esto no siempre es factible debido a los recursos y la logística necesarios para llevar a cabo proyectos tan grandes. Sin embargo, en la actualidad, el desarrollo de nuevas herramientas informáticas y tecnologías permite acceder y hacer uso de las bases de datos generados por grandes consorcios abriendo la posibilidad de analizar datos de distintas cohortes y de tener un mejor panorama para comprender procesos e identificar marcas que permitan predecir la respuesta a la vacunación.

Reservorios de datos públicos de secuenciación masiva

La reciente explosión en la generación de datos genómicos públicos, y en particular la descripción del transcriptoma por medio de secuenciación de ARN (RNA-seq) y otras fuentes de datos biológicos a escala genómica, proporciona una oportunidad única para explorar características previamente no anotadas en el genoma humano. Existen grandes consorcios que han llevado a cabo proyectos muy importantes que han generado grandes cantidades de datos como el consorcio del proyecto para obtener muestras de tejido humano sano GTEx, el proyecto para

la secuenciación de tejido tumoral TCGA, además del reservorio de datos de secuenciación de distintos orígenes SRA, los cuales detallo a continuación:

Sequence Read Archive (SRA): Es una base de datos bioinformáticos que sirve de reservorio de secuencias crudas que provienen de tecnologías de secuenciación de nueva generación como Illumina, 454, IonTorrent, PacBio, Complete Genomics y Oxford Nanopore. Este archivo se estableció dentro del Centro Nacional para la Información Biotecnológica (*National Center for Biotechnology Information*, NCBI), además es parte de la Colaboración Internacional para la Base de Datos de Secuencia de Nucleótidos (*International Nucleotide Sequence Database Collaboration*, INSDC). Dentro del SRA se encuentran proyectos de gran escala, así como proyectos pequeños sobre diferentes tipos de muestras, condiciones, origen y tratamiento.

Genotype-Tissue Expression (GTEx): Uno de los grandes proyectos de secuencias de expresión en humanos es GTEx (*Genotype-Tissue Expression* (Consortium, 2017; Consortium, 2013)). Es un proyecto que ha buscado generar un recurso público para estudiar la expresión y la regulación génica específica para cada tejido. Se han recolectado muestras de 53 tejidos sanos, de casi 1000 individuos, para realizar ensayos de secuenciación de genoma completo (WGS), secuenciación del exoma completo (WES) y secuenciación a ARN (RNA-Seq). La base de datos que contiene los resultados de este estudio es una herramienta invaluable para estudios comparativos de expresión génica entre patologías y tejidos sanos. Así, una de las publicaciones recientes obtenida a partir de estos datos, analiza las variaciones en enfermedades complejas en diferentes tejidos (Consortium et al., 2017; Gamazon et al., 2018).

The Cancer Genomic Atlas (TCGA): El Instituto Nacional de Cáncer (NCI) y el Instituto Nacional de Investigación Genómica Humana (NHGRI), en Estados Unidos de América, realizaron una colaboración para crear un reservorio de datos genómicos de tejidos tumorales creando el Atlas Genómico del Cáncer o TCGA (Cancer Genome Atlas Research et al., 2013; Hoadley et al., 2018). Esta base de

datos tiene mapas multidimensionales sobre cambios genómicos claves en 33 tipos de cáncer diferentes. Estos datos han servido para determinar el paisaje inmunológico en cáncer (Thorsson et al., 2018), lo cual es de gran utilidad para comprender el papel del sistema inmune y la relación con el microambiente tumoral.

recount2: Uno de los grandes retos que existen frente a la gran cantidad de información y datos en los distintos reservorios de secuenciación, es el análisis y la minería de datos. El poder escanear y buscar dentro de todo el universo de datos puede ser una tarea complicada. Además, el poder homologar datos de diferentes fuentes es de gran importancia a la hora de realizar estudios de expresión diferencial. Por ello, las herramientas bioinformáticas para el análisis de datos son cada vez más importantes y necesarias. Una de las herramientas que ha sido de gran utilidad y ha permitido el análisis transcripcional de genes es *recount2*: un recurso bioinformático al que se puede acceder vía internet por medio de su página web o a través **Bioconductor** (<https://www.bioconductor.org>). Esta herramienta cuenta con más de 70,000 muestras de RNA-seq procesadas de manera uniforme que se encuentran agrupadas por proyecto provenientes de SRA, de los datos de los consorcios de GTEx y TCGA. Aquí se encuentran cuentas de genes y exones, además de archivos de cobertura tipo *bigWig*. *recount2* provee datos de expresión resumidos y organizados a diferentes niveles, lo que permite análisis entre estudios. Para facilitar los análisis estadísticos, para cada estudio se crea una tabla de cuentas a nivel de gen y exón y también permite la extracción de datos de fenotipo. Una de las características más importantes de esta herramienta es que además de las cuentas en genes anotados permite el conteo de expresión en regiones que no se encuentran anotadas en el genoma, lo cual es de gran utilidad para nuestro proyecto.

Planteamiento del problema

La reacción del centro germinal es fundamental para el desarrollo de memoria inmunológica y la respuesta inmune adaptativa. Debido a que la recombinación de cambio de clase no codificante es un evento molecular asociado

al cambio de clase, el cual a su vez es una característica de los linfocitos B que pasan por centro germinal, nos preguntamos si el análisis cuantitativo de la transcripción no codificante producto de la recombinación de cambio de clase refleja la respuesta adaptativa en condiciones de salud y de enfermedad. Debido a que nuestro conocimiento sobre el papel de estos transcritos no codificantes de recombinación de cambio de clase en el humano es escaso, en este trabajo proponemos caracterizar los patrones de transcripción de recombinación de cambio de clase no codificante y su asociación con determinantes inmunológicos distintivos, como la vacunación, la localización anatómica y señales microambientales, durante la respuesta inmune humana normal y patológica, con la finalidad de identificar nuevos mecanismos patogénicos y firmas transcripcionales con potencial valor clínico.

Hipótesis

Debido a que la recombinación de cambio de clase es un proceso derivado de la activación y la respuesta de los linfocitos B ante un estímulo antigénico, y que los transcritos no codificantes asociados a recombinación de cambio de clase preceden este mecanismo, los transcritos no codificantes reflejan el estado funcional de la respuesta de linfocitos B.

Objetivos

- Definir los límites de expresión de los transcritos no codificantes asociados a cambio de clase de los nueve isotipos de inmunoglobulina en humanos.
- Cuantificar los transcritos no codificantes $I\mu$, $I\gamma1$ e $I\gamma3$ *in vitro* y *ex vivo*
- Identificar y caracterizar la expresión de los transcritos no codificantes asociados a cambio de clase de los nueve tipos de inmunoglobulina en tejido sano.
- Identificar y caracterizar la expresión de los transcritos no codificantes en condiciones patológicas de infección o tumorales.

Materiales y Métodos

Vacunación de voluntarios sanos: Todos los procedimientos en sujetos humanos se realizaron después de la aprobación de la comisión de Investigación, comisión de Ética y comisión de Bioseguridad del Instituto Nacional de Salud Pública (CI: 971 / 82-6684). Posterior a firma de consentimiento informado de cada voluntario, se obtuvieron muestras de sangre periférica pre-inmune (día 0), y post- vacunación (día 7, 15, 30 y 180). Se obtuvieron aproximadamente 18 mL, mediante venopunción en tubos Vacutainer® CPT™ de 2 x 8 mL de voluntarios sanos vacunados con Hepatitis B y/o Toxoide tetánico/Difteria, o bien con la vacuna trivalente contra la influenza durante la temporada 2013-2014 (A/California/7/2009 (H1N1) pdm09; A (H3N2) A/Victoria/361/2011; B/Massachusetts/2/2012). Se aisló plasma y células mononucleares de sangre periférica (PBMC) por centrifugación en gradiente, y almacenaron a -80°C y N₂, respectivamente. El ARN total se extrajo de PBMC con TRIzol Reagent (Cat. No. 15596026, ThermoFisher) y se almacenó a -80°C hasta su uso.

Serología: A los días 0, 28 y 180 post-vacunación en tubos BD Vacutainer® para Suero con Activador de Coagulación, se obtuvo entre 1 y 2.5 mL de suero de sangre periférica y se utilizaron para cuantificar anticuerpos contra el antígeno de superficie de hepatitis B, así como anticuerpos contra el antígeno toxoide diftérico, mediante ensayos de ELISA. Estos ensayos se realizaron con los siguientes kits: Diphtheria Human Bioassay™ ELISA Kit (Cat. No. D8065-49D, USBiological) y Hepatitis B Surface Antigen (HbsAg) Bioassay™ ELISA Kit (Cat. No. H1909-09, USBiological). Se leyó la densidad óptica a 450nm mediante el lector de ELISA.

Cuantificación de los plasmablastos por FACS: PBMC se descongelaron en medio RPMI SFB10%, se fijaron con paraformaldehído al 1% y se tiñeron con el siguiente cóctel de anticuerpos (Biolegend): anti-CD3 PerCP / Cy5.5 (clona SKY) dilución 1/100, anti-CD19 FITC (clona HIB19) dilución 1/50, anti-CD20 PE Cy7 (clona 2H7) dilución 1/50, anti-CD27 APC (clona O323) dilución 1/100 y anti-CD38

PE (clona HIT2) dilución 1/20. El análisis de citometría de flujo se realizó en un FACS Aria II (BD Biosciences). Los dobletes y CD3⁺ fueron excluidos. Los plasmablastos se definieron como CD3⁻ / CD19⁺ / CD20⁻ / CD27⁺ / CD38⁺. Se adquirieron entre 500-1000 plasmablastos por muestra. Todo el análisis se realizó utilizando el software Flowjo (TreeStar).

Estimulación de células B *in vitro*: Se aislaron PBMC con Ficoll-Paque™ (Cat. No. 17-1440-02, GEHealthcare) por gradiente de densidad a partir de paquetes globulares obtenidos del banco de sangre. Las células B se enriquecieron mediante selección negativa utilizando el kit de para el aislamiento de células B, B cell Isolation Kit II (Cat No. 130-091-151, MACS, Miltenyi). Se sembraron 1x10⁶ células B por pozo en placas de 6 pozos incubadas en medio RPMI suplementado con 10% de SFB, estreptomicina y penicilina (100X) a 37°C con 5% de CO₂. Se establecieron dos condiciones de activación en diferentes momentos (3 y 6 días después de la activación): **1.- Activación tipo centro germinal (similar a CG):** CD40 antihumano 1µg / mL (G28.5, obsequio del laboratorio del Dr. Leopoldo Santos), 5µg / mL CpG ODN 2006 (Cat. No. tlr-2006-1, Invivogen) y 25 ng / mL de IL-21 recombinante humana (Cat. No. 14-8219-62, eBiosciences). **2.- Activación T-independiente:** 5µg / mL CpG ODN 2006 0,05% de proteína A soluble de *S. aureus* Cowan (Cat. No. 82526-1MG, Pansorbin; Calbiochem), 5 ng / mL de lectina (mitógeno) de fitolaca (*Phytolacca americana*) (PWM, Cat. No. L9379-5MG, Sigma).

qRT-PCR de transcritos de RCCnc: El ARN total de las PBMC se extrajo usando TRIzol (Invitrogen). La integridad del ARN se evaluó con un bioanalizador de ARN, Agilent RNA 6000 Nano, utilizando el programa predeterminado del aparato para detección de ARN. Se realizó RT-PCR en un solo paso con SuperScript™ III Platinum™ One-Step qRT-PCR Kit (Cat. No. 11732088, Invitrogen) para transcripción inversa y amplificación. La PCR cuantitativa de los transcritos RCCnc para los *IGHM*, *IGHG1*, *IGHG3* y *AICDA* (AID) se realizó usando oligonucleótidos diseñados específicamente y sondas TaqMan (IDT). Los oligonucleótidos y las sondas utilizadas para cuantificar los transcritos RCCnc se detallan en la tabla al final de este párrafo. Se utilizaron sondas de Universal Probe Library #61 para el

transcrito de I μ y #69 para el transcrito de la AID, mientras que se diseñaron sondas específicas para el transcrito de I γ 1 e I γ 3 de IDT. La amplificación de HPRT se realizó con PrimeTime® Predesigned qPCR Assay (Cat. No. 1055772, IDT) como el gen de referencia. Se utilizaron células B enriquecidas en reposo como calibrador y células no B como control negativo. *HPRT1* se usó como normalizador para cada condición. El aumento de la transcripción (*fold difference*) se calculó usando el método $2^{\Delta\Delta CT}$.

Oligonucleotidos utilizados para qRT-PCR de I α , I γ 1 and I γ 3			
	Forward	Reverse	Sonda
I γ 1	CAAGCCCCTCCGTTAC	ACGAGGAACATGACTGGATG	5-F/AGGAGGCAG/Z/CAGAGCGAGG/3IB
I γ 3	CCTGTTGTGGCGAGGTACA	TTAGTGTTTGCAGCGTGGAG	5-F/CTGTCAGCT/Z/GCCACTTGCTTCCT/3IB
I α	CCAGGTGTTGTTTTGCTCAGT	CACCTTCTGGTTGTGAAGAGGTG	UPL_ROCHE#61
AID	TGGACACCACTATGGACAGC	GCGGACATTTTTGAATTGGT	UPL_ROCHE#69
F= FAM; Z = ZEN; IB = Iowa Black			

Definición de límites de transcripción RCCnc: Los transcritos RCCnc y las regiones de *switch* (S) no se encuentran mapeadas en la versión actual del genoma humano (GRCh38). Se conoce que los I_H están río arriba de la región de *switch* correspondiente (S_H), por lo que primero mapeamos regiones S en función de la distribución de la frecuencia del motivo AGCT en intervalos de 500 bp a lo largo de todo el locus *IGH* humano en el cromosoma 14 (Xu et al., 2012). Utilizamos los metadatos de los proyectos de SRA contenidos en *recount2*, así como información obtenida a través del motor de búsqueda de SRA-Run Selector (<https://www.ncbi.nlm.nih.gov/Traces/study/>) para identificar muestras de RNA-seq provenientes de células B purificadas, para asignar los I_H. Los archivos *bigWig* correspondientes se descargaron utilizando el paquete *recount2* de Bioconductor y las cuentas de lecturas mapeadas se inspeccionaron con Integrative Genomics Viewer (Thorvaldsdottir et al., 2013), lo cual permitió la anotación de las diferentes regiones basado en la visualización de la unanimidad en la expresión de linfocitos B purificados provenientes de diferentes proyectos.

Cuantificación de transcripción de RCCnc y CH: Por medio de *recount2* extrajimos las cuentas de lectura además del RPKM de cada uno de los nueve

genes codificantes de la región constante de C_H (*IGHM*, *IGHD*, *IGHG3*, *IGHG1*, *IGHA1*, *IGHG2*, *IGHG4*, *IGHE* e *IGHA2*), así como las coordenadas de las regiones de I_H obtenidas en nuestro mapeo correspondiente a cada región constante. Estas cuentas se concentraron en un objetos tipo *GRanges* (Lawrence et al., 2013), en el cual se concentraron las muestras por proyecto, las cuentas y el RPKM para cada uno de las regiones de interés. El RPKM transformado a \log_2 de C_H (codificante) y el promedio del RPKM transformado a \log_2 de *RCCnc* /muestra se utilizó como medida de la abundancia de la transcripción.

Debido a que detectamos una cantidad considerable de proyectos con cuentas por debajo de 1 RPKM, además de que en algunos casos no había cuentas a lo largo de toda la región sino solo en algunos pares de bases, decidimos clasificar las muestras con cuentas para diferenciarlas entre significativas (altas) o no significativas (bajas), por medio de un corte arbitrario basado en el promedio de RPKM. Para la transcripción *RCCnc*, el promedio de RPKM \log_2 /muestra adoptó una distribución cuasi-normal, con una media de 2.65 RPKM \log_2 ($DS \pm 3.79$), que corresponde a 6.29 RPKM. Como exploración inicial a qué tejidos y en qué enfermedades se produce la transcripción *RCCnc*, se utilizó el promedio de RPKM \log_2 como punto de corte para definir “alta” expresión ($> 2.65 \log_2$ RPKM) o “baja” ($< 2.65 \log_2$ RPKM). El promedio de la transcripción de C_H (codificante) RPKM \log_2 fue 7,82 ($DE \pm 5,16$). Dada la diferencia entre la transcripción C_H y *RCCnc*, ambos valores RPKM \log_2 de cada muestra para cada gen *IGH* se estandarizaron por transformación a valor Z , lo cual permite comparar su expresión relativa.

Mapeo de ontología de muestras de RNA-seq de SRA: Aunque todas las muestras de RNA-seq en TCGA y GTEx siguen una categorización ontológica homogénea, los metadatos asociados a los proyectos de SRA son ampliamente heterogéneos y comúnmente insuficientes. Para obtener una categorización más homogénea de casi la mitad de nuestro conjunto de datos, utilizamos las anotaciones de enfermedades (Kibbe et al., 2015) obtenidas de MetaSRA, versión 1.2 (Bernstein et al., 2017).

Agrupamiento de perfiles de expresión: Para definir los perfiles de transcripción RCCnc, los valores Z para cada exón I_H se sometieron a agrupamiento por k -medias usando el software Cluster 3.0 (de Hoon et al., 2004). Se generaron diez grupos, con 100 ejecuciones iterativas y distancia euclídea como medida de distancia. Los datos agrupados se visualizaron con java Treeview 3.0 (<https://sourceforge.net/projects/jtreeview/>).

Pruebas de enriquecimiento de tejido con respecto al perfil transcripcional RCCnc: Para identificar predominio de un perfil de transcripción RCCnc en un tejido sano particular (conjunto de datos GTEx), se realizó un análisis de enriquecimiento de muestras de tejido de acuerdo con el clúster Z , utilizando una prueba exacta de Fisher de dos lados, con la función de `Fisher.test (c, alternative = "two.sided")`. Para ello, se construyó una tabla de contingencia de 2×2 para cada clúster Z , con el número de muestras por tejido que pertenecen y las que no pertenecen al clúster. H_0 indica que no hay diferencia en la distribución de probabilidad entre los grupos Z y el tejido. El ajuste del valor p con el método Benjamini-Hochberg se realizó para corregir por pruebas múltiples con la función `R p.adjust (p, method = "bh", n = length (p))`. Se consideró una tasa de descubrimiento falso (FDR) < 0.01 como un enriquecimiento significativo.

Análisis de expresión diferencial: Los valores de RPKM para los 10 exones I_H y sus exones C_H correspondientes de estudios seleccionados fueron utilizados para analizar su expresión diferencial mediante las funciones `limFit ()` y `eBayes ()` del paquete *limma* R v3.34. Debido a que en *recount2* el tamaño de las librerías (datos crudos) fue estandarizado a 40×10^6 reads (Collado-Torres et al., 2017), se utilizó el enfoque `limmaTrend`, en lugar de `voom` (Law et al., 2014). En caso de que algún estudio contara con replicas técnicas, se ajustó la correlación inducida mediante la función de `duplicateCorrelation()` de *limma*. Los valores resultantes de p ajustados por Bonferroni inferiores a 0.05 se consideraron como estadísticamente significativos. Usamos Bonferroni en lugar de FDR, dado que probamos 19 regiones (9 C_H 's y 10 I_H 's) en lugar de la cantidad habitual de miles de genes.

Resultados

Mapeo de la transcripción de RCCnc en Linfocitos B humanos aislados: *recount2* (Collado-Torres et al., 2017) es una herramienta informática que comprende un catálogo muy heterogéneo de experimentos de RNA-seq humanos provenientes de 2,036 estudios independientes, sumando un total de 70,603 muestras, muchas de las cuales representan replicas biológicas, tanto de casos (generalmente enfermedad, líneas celulares, etc); como de controles (tejido proveniente de sujetos sanos). Aunque cada estudio está compuesto por un pequeño número de muestras (34 muestras en promedio), los estudios TCGA (Cancer Genome Atlas Research et al., 2013; Hoadley et al., 2018) y GTEx (SRP012682) (Consortium, 2017; Consortium, 2013) son los proyectos con el mayor número de muestras (11,284 y 9,661 respectivamente), y representan más del 29% de nuestro conjunto de datos (**Figura 6**). El número de muestras y replicas biológicas por estudio es importante para conferir poder estadístico al momento de hacer análisis de expresión diferencial.

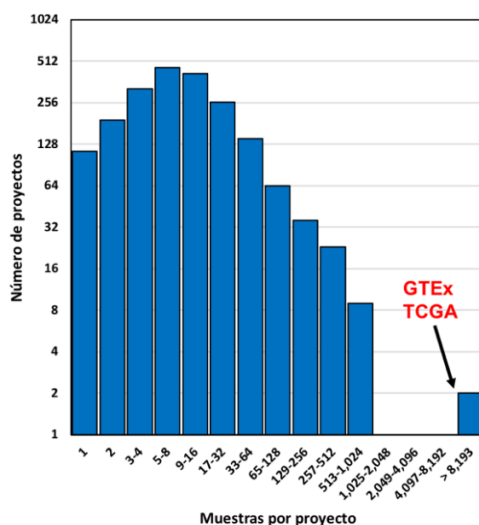


Figura 6. Distribución del número de muestras de RNA-seq analizadas por proyecto. Todo el conjunto de datos *recount2* se clasificó según el número de muestras por proyecto (eje X) y el número de proyectos (eje Y). El número de proyectos analizados con *recount2* fue 2,036 con un total de 70,603 muestras. (Mediana = 8 muestras por proyecto; Media =

32 muestras por proyecto. TCGA y GTEx son dos proyectos que tienen 11,284 y 9,661 muestras, respectivamente, que comprenden el 30% de todas las muestras.

Aunque la transcripción RCCnc humana ha sido evaluada por RT-PCR (Chowdhury et al., 2008; Fear et al., 2004; Litinskiy et al., 2002), los límites precisos de la transcripción, incluidos los sitios alternativos de inicio de la transcripción y las variantes de empalme (*splicing*), siguen sin definirse. Con la finalidad de definir los límites transcripcionales de los exones I_H (RCCnc) en humanos, y para poder medir su expresión mediante la posterior cuantificación de lecturas seleccionamos muestras de RNA-seq derivadas de linfocitos B humanos aislados mediante citometría de flujo de donadores sanos provenientes de distintos proyectos detallados en la **Tabla S1** (Beguelin et al., 2013; Henn et al., 2013; Hoek et al., 2015; Koues et al., 2015; Kushwaha et al., 2016; Linsley et al., 2014; Pascual et al., 2014). Encontramos que en las células B normales, los límites de los exones I_H (RCCnc) están menos definidos que los transcritos codificantes, y como se esperaba, se extienden a las regiones de *switch* (**Figura 7A-C**) (Chowdhury et al., 2008).

Si bien los RCCnc se asocian mecánicamente al cambio de clase, desconocemos si se expresan en otros tipos celulares diferentes a los linfocitos B. Los proyectos SRP045500 (Linsley et al., 2014) y SRP051688 (Hoek et al., 2015) describen el transcriptoma de células inmunes aisladas de sangre periférica (PB) en adultos sanos, incluidos neutrófilos, monocitos, células dendríticas mieloides, linfocitos B, linfocitos T y células NK. Estos proyectos permitieron observar y comprobar que, entre las células diferenciadas derivadas del linaje hematopoyético, la transcripción de RCCnc se restringe a las células B (**Figura 7A**, **Figura 9**). Analizamos la transcripción de I_H y C_H en células B CD19 + aisladas de sangre periférica y observamos que el patrón de transcripción de RCCnc es similar al de sangre completa (**Figura 8**), donde se observa una expresión relativa de I_μ y C_μ alta, seguida por la expresión de I_{γ3} e I_{α1} y finalmente una expresión menor de I_ε (Henn et al., 2013; Linsley et al., 2014).

Debido a que la RCC es un proceso que se da en linfocitos B activados, particularmente de centro germinal, analizamos la transcripción de RCCnc en las

células B de amígdalas y en las células B del centro germinal del proyecto SRP021509 (Beguelin et al., 2013). La transcripción de RCCnc para la mayoría de I_H , pero no I_E , fue relativamente alta tanto en células B vírgenes como en células B de centro germinal (**Figura 8D**). Estos hallazgos indican que, aunque la transcripción de RCCnc es única en células B, no es exclusiva de las células B activadas, lo que concuerda con hallazgos anteriores que demuestran la transcripción constitutiva de RCCnc (Chowdhury et al., 2008; Fear et al., 2004).

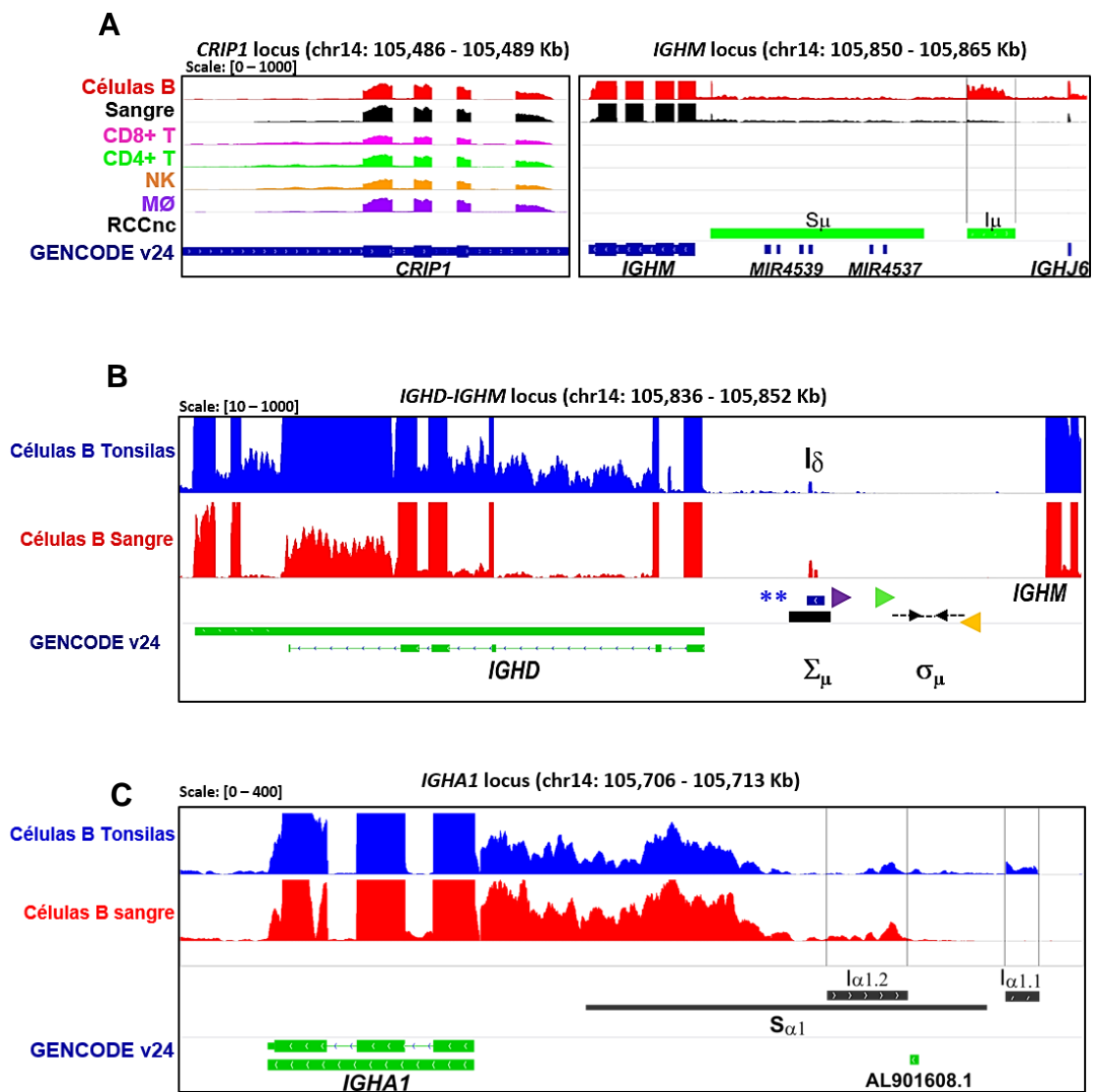


Figura 7. Definición de los límites de transcripción RCCnc. Los proyectos que utilizan células B aisladas fueron seleccionados para definir los límites de la transcripción RCCnc en humano. (A) Gráfico de cobertura del proyecto SRP045500 que muestra el locus CRIP1 (panel izquierdo) y el locus IGHM (panel derecho) en B (pista roja) y en células que no son B. El gen CRIP1 se transcribe en células B y no B, mientras que IGHM e $I\mu$ (dentro de las líneas verticales negras) se transcriben solo en células B y en sangre periférica (pista negra). La región $S\mu$ pronosticada se muestra en la parte inferior (azul) y los genes anotados (GENCODE v24) se muestran en verde(B). Una vista del locus del intrón IGHM - IGHD que muestra los gráficos de cobertura de las amígdalas ingenuas normales purificadas (SRP021509, pista azul) y las células B CD19 + de sangre periférica (SRP060715, pista roja). La flecha negra vertical muestra $I\delta$. La anotación IGHD GENCODE v24 se muestra en verde. El exón $I\delta$ se superpone con la región (Akahori et al., 1988; Arpin et al., 1998; Kluin et al., 1995) y es centromérico con respecto a los sitios mapeados para las uniones RCC (Flechas rayas negras) (Kluin et al., 1995) Primer antisentido (18156) utilizado por Kluin y Arpin, et (P4) se muestran en púrpura y verde, respectivamente. El cebador $S\delta$ Sense utilizado por (Chen et al., 2009) se muestra en naranja. Los asteriscos azules indican los sitios Hind III (Akahori et al., 1988; Yasui et al., 1989; Kluin et al., 1995). (C) Panorama transcripcional del locus IGHA1 que muestra los gráficos de cobertura de IGHA1, la anotación GENCODE v24 (pista verde), las amígdalas (SRR834982, azul) y las células B de sangre periférica (SRR2097501, rojo). Se muestran las dos transcripciones. Se muestra la región $S\alpha1$ (barra negra). Para A-B, la escala de cobertura se muestra en la esquina superior izquierda entre paréntesis.

Se ha descrito que en líneas celulares de mieloma hay cambio de clase mediante recombinación a IgD por un mecanismo no canónico (Akahori et al., 1988; Yasui et al., 1989), por lo cual exploramos la transcripción en la región intergénica entre IgM e IgD. Se identificó una región transcripcionalmente activa de 308 pares de bases (pb) dentro del intrón *IGHM-IGHD* (referida en lo sucesivo como $I\delta$) y se incluyó para el análisis subsecuente (**Figura 9C**). Esta región es homóloga a $I\mu$, y se superpone con una región repetida descrita previamente llamada $S\mu$, implicada en RCC de μ - δ en mielomas IgD⁺ (Akahori et al., 1988; Yasui et al., 1989). Para el gen *IGHA1*, identificamos dos posibles exones I_H ($I\alpha1.1$ y $I\alpha1.2$) que se seleccionaron para su posterior análisis (**Figura 9D**). $I\gamma2$ se superpone a un lincRNA previamente anotado, ENST00000497397.1 (AL928742.1) en ENSEMBL (Zerbino et al., 2018).

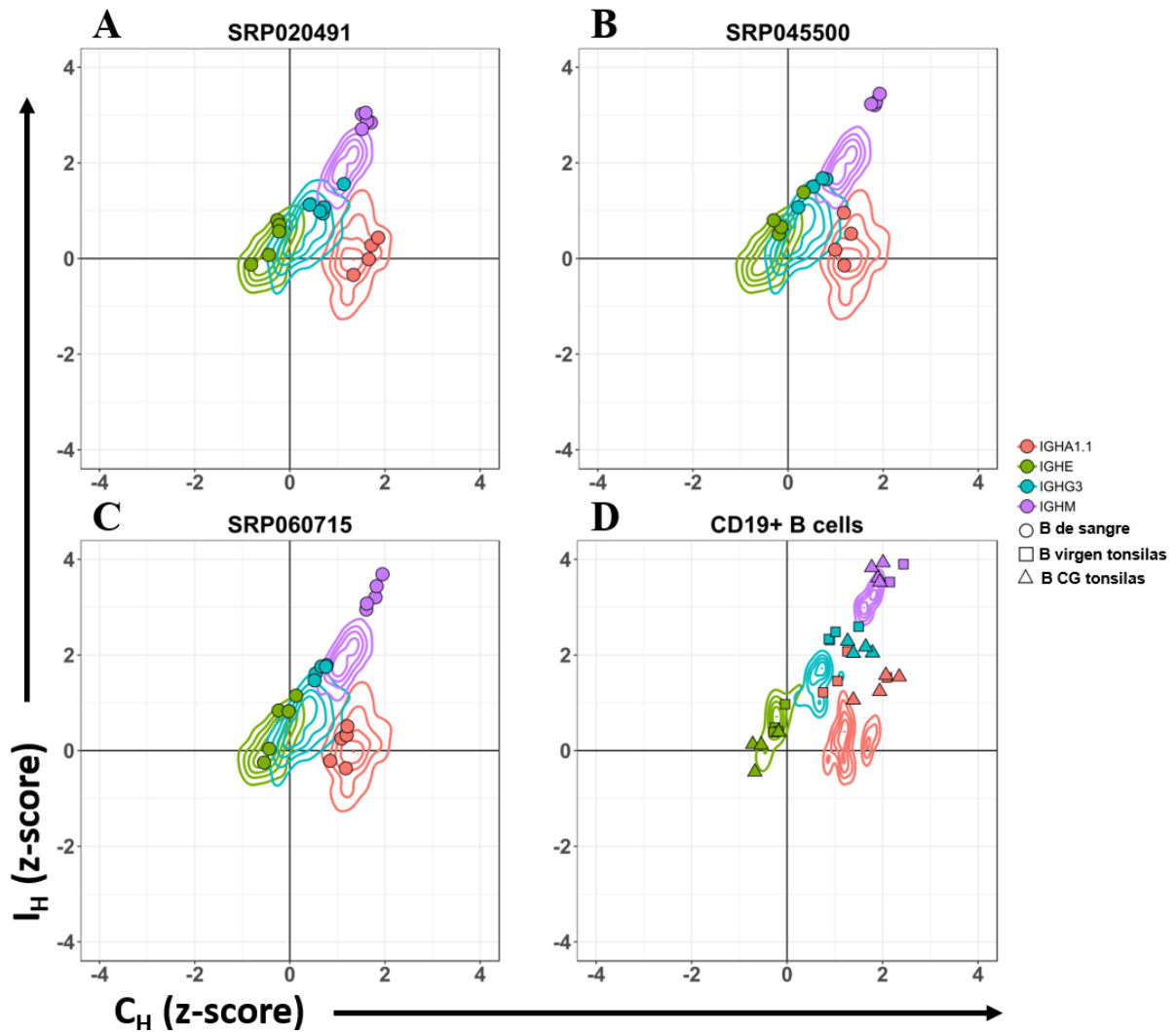


Figura 8. El patrón de transcripción RCCnc en sangre completa es similar a las células B CD19 + purificadas de sangre periférica. Graficas de contorno 2D correlación de la transcripción relativa de RCCnc y C_H (puntuación Z) para IgA (naranja), IgE (verde), IgG3 (azul) e IgM (morado) en el ARN de sangre total del conjunto de datos GTEx ($n = 456$ muestras) en (A-C). Los círculos de colores rellenos corresponden a las puntuaciones Z para IgA (naranja), IgE (verde), IgG3 (azul) e IgM (morado) en las células B CD19 + purificadas de sangre periférica de tres proyectos independientes de RNA-seq (Henn et al., 2013; Linsley et al., 2014), superpuestos a los datos de GTEx (archivo complementario 2: tabla S1). (D) Grafica de contorno 2-D de los valores Z en las células B periféricas CD19 + purificadas utilizadas en A-C ($n = 14$), superpuestos con las puntuaciones Z correspondientes de células B vírgenes de tonsilas (cuadrados. $N = 4$) y células B de centro

germinal obtenidas d amígdalas (triángulos. n = 4) del proyecto SRP021509 (Beguelin et al., 2013

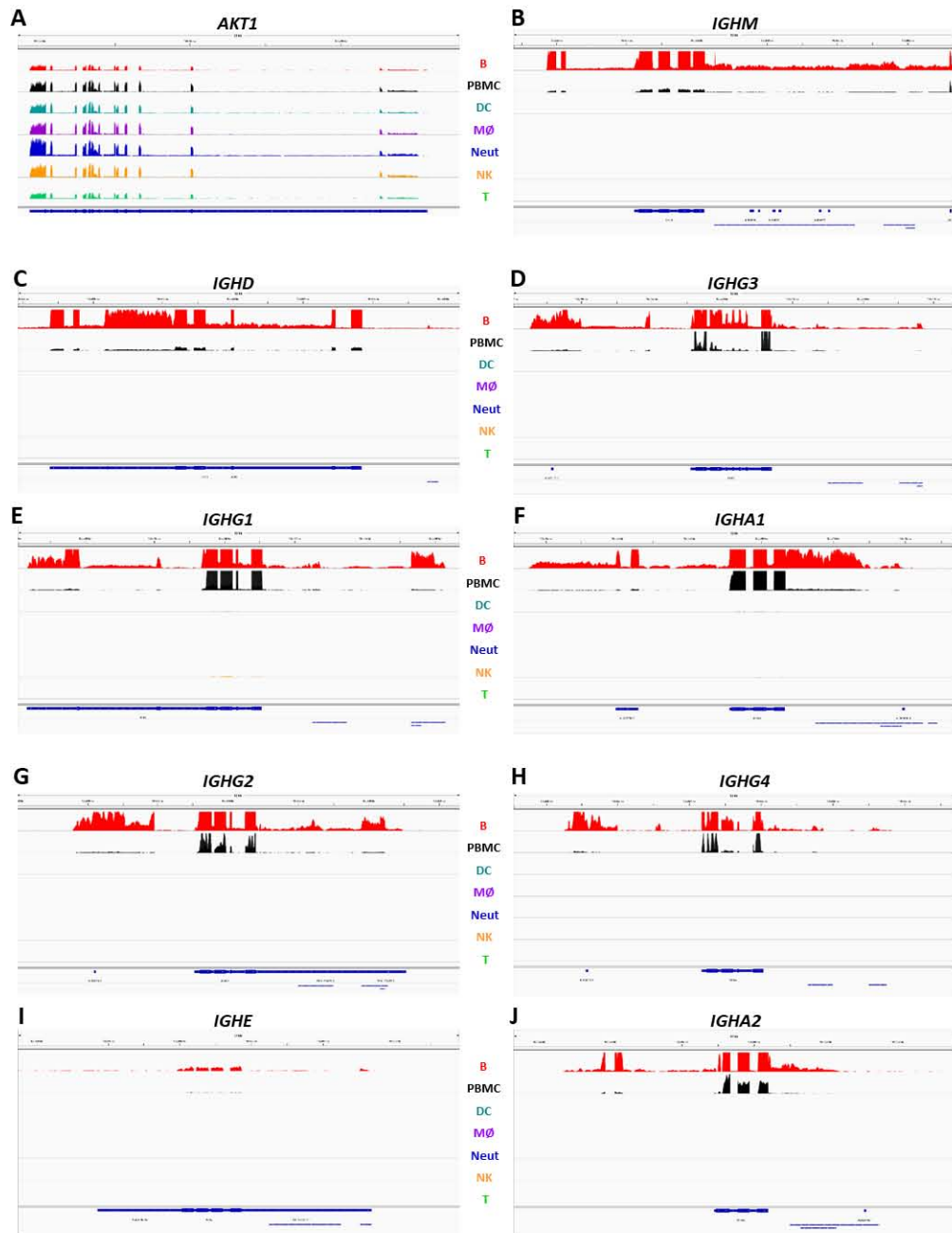


Figura 9. La transcripción RCCnc es específica de las células B. Gráficos de cobertura del locus AKT1 (A) e IGH (B-J) que muestran la actividad transcripcional en células diferenciadas derivadas de células hematopoyéticas aisladas del proyecto SRP051688 (Hoek et al., 2015). Células B (B, pista roja), PBMC totales (PBMC, pista negra), células dendríticas mieloides (DC, pista verde turquesa), monocitos (MØ, pista púrpura), neutrófilos

(Neut, pista azul), células asesinas naturales (NK), pista naranja) y células T (T, pista verde).

Tabla 1. Mapeo de exones I (I_H) y regiones de <i>switch</i> (S_H) en el cromosoma 14 humano (GRCh38.p10)			
Gen	Inicio	Fin	Longitud (pb)
<i>IGHM</i>			
I _μ	105,861,311	105,862,213	903
S _μ	105,856,501	105,860,500	4,000
<i>IGHD</i>			
I _δ	105,847,549	105,847,857	309
<i>IGHG3</i>			
I _{γ3}	105,775,023	105,775,702	680
S _{γ3}	105,773,001	105,774,000	1,000
<i>IGHG1</i>			
I _{γ1}	105,747,322	105,748,299	978
S _{γ1}	105,744,501	105,745,500	1,000
<i>IGHA1</i>			
I _{α1.2}	105,711,304	105,711,907	604
I _{α1.1}	105,712,639	105,712,892	254
S _{α1}	105,709,501	105,712,500	3,000
<i>IGHG2</i>			
I _{γ2}	105,647,803	105,648,549	747
S _{γ2}	105,627,501	105,628,500	1,000
<i>IGHG4</i>			
I _{γ4}	105,629,000	105,629,500	501
S _{γ4}	105,623,001	105,624,000	1,000
<i>IGHE</i>			
I _ε	105,605,036	105,605,357	322
S _ε	105,602,501	105,604,500	2,000
<i>IGHA2</i>			
I _{α2}	105,589,300	105,590,300	1,001
S _{α2}	105,589,001	105,591,500	2,500

Las coordenadas genómicas identificadas y analizadas en mayor detalle de cada exón I_H se encuentran en la **Tabla 1**. El recorrido a través del locus IgH utilizando el Navegador de Genomas ENSEMBL (Zerbino et al., 2018) permitió confirmar que los exones I_H predichos, incluyendo I_δ y regiones donde se da la unión de RNAPolIII, H3K36 y H3K4. Además, se observó enriquecimiento de la trimetilación y acetilación de lisinas (H3K36me3, H3K3me3 y H3K27ac) marcas de transcripción activa (cromatina abierta), en células B de sangre periférica y células B transformadas con EBV generadas por los proyectos Roadmap Epigenomics y ENCODE (Consortium, 2012; Roadmap Epigenomics et al., 2015) (**Material Suplementario: Figura S1**). En general, estos resultados concuerdan con el modelo actual de transcripción RCCnc (Chowdhury et al., 2008); sin embargo, la identificación de novedosos elementos transcripcionales complica el estudio y entendimiento de la regulación transcripcional del locus IgH.

Hallazgos principales

- *La transcripción de RCCnc es exclusiva de células B, aunque no es exclusiva de las células B activadas.*
 - *Identificación de un nuevo transcrito no codificante Id en la región intergénica IGHM-IGHD*
-

Caracterización de la transcripción de RCCnc *in vitro* y en respuesta a la vacunación: La transcripción RCCnc aumenta en respuesta a la activación de células B (Chowdhury et al., 2008). Para confirmar que los exones I_H identificados *in silico* son transcritos y se pueden medir, validamos la transcripción de I_μ, I_{γ3} y I_{γ1} en reposo y en respuesta a la activación. Linfocitos B de donadores sanos fueron estimulados *in vitro* durante 3 y 6 días con activación tipo T-dependiente (ligando de CD40, IL-21) y activación T-independiente (PWM, SAC) para ver si el tipo de activación hacia una diferencia en la transcripción de los exones (**Figura 10**).

Detectamos un incremento de la transcripción de $I\mu$, $I\gamma_3$ e $I\gamma_1$ en células B activadas tanto por activación tipo T-dependiente como por activación T-independiente. Sin embargo, la expresión fue significativamente más alta para todos los transcritos con la activación tipo T-dependiente. En ambos tipos de activación, la transcripción *RCCnc* fue superior a los 6 días después de la activación (**Figura 10D-G**). La transcripción más alta fue para $I\mu$ (3 veces mayor que $I\gamma_1$ e $I\gamma_3$). La transcripción de AID mostró el mismo patrón que la transcripción *RCCnc*, siendo mayor a los 6 días post-activación con estímulo T-dependiente (**Figura 10D**). Observamos un decremento de la transcripción en la activación tipo T independiente para $I\mu$, $I\gamma_1$ e $I\gamma_3$ al día 3 el cual podría deberse un efecto de la metodología donde al haber una activación de linfocitos B aumenta el número de células que no expresan necesariamente estos transcritos, aunque eso no explicaría porque para la activación tipo T-dependiente no se observó ese fenómeno.

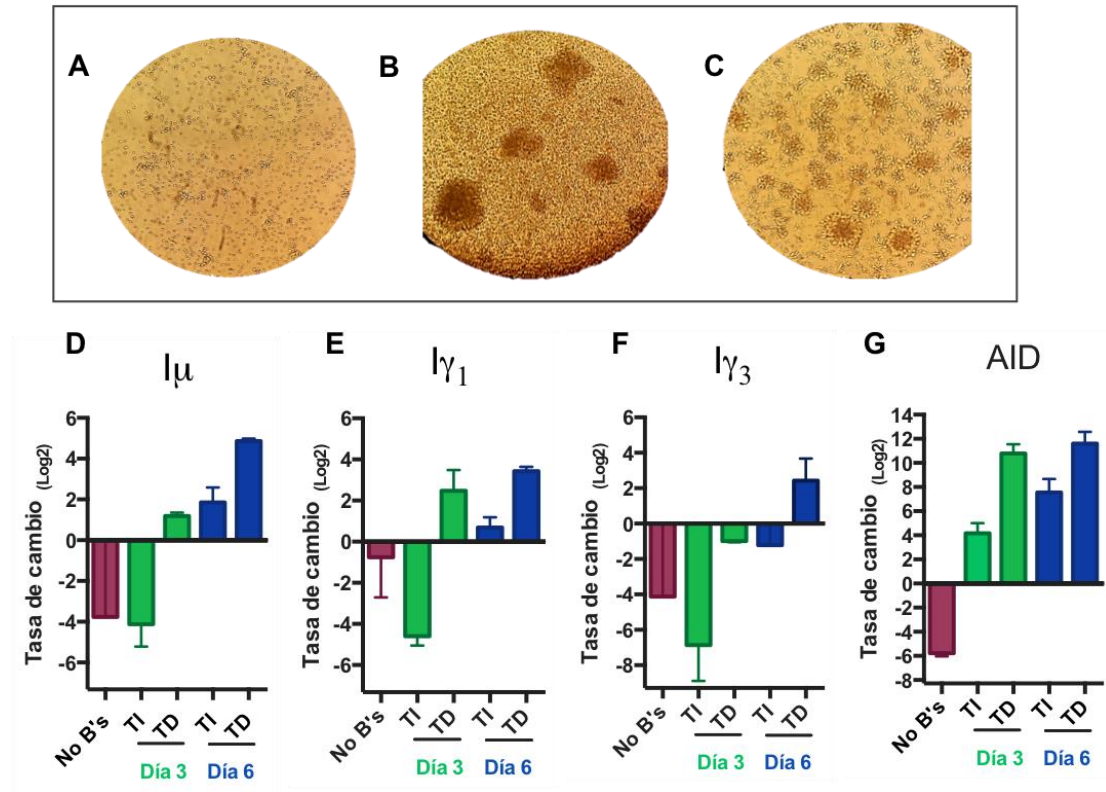


Figura 10. Activación de células B y cuantificación de transcritos RCCnc *in vitro*. (A-C) Células B enriquecidas cultivadas durante 3 días. Imagen de microscopio de cultivo de células B a los 3 días de: (A)cultivo control sin activar, (B) activación independiente de T (TI) (PWM, SAC) y (C) activación dependiente de T (TD) (IL-21, α -CD40). (D-G) Cuantificación de transcritos $I\mu$, $I\gamma_3$, $I\gamma_1$ y AID por medio de qRT-PCR. Las gráficas de barras representan el cambio de expresión de los transcritos en células activadas tipo TI y TD a los 3 y 6 días con respecto a células B no activadas (células no B como control). $I\mu$ (D), $I\gamma_1$ (E), $I\gamma_3$ (F) y AID(G).

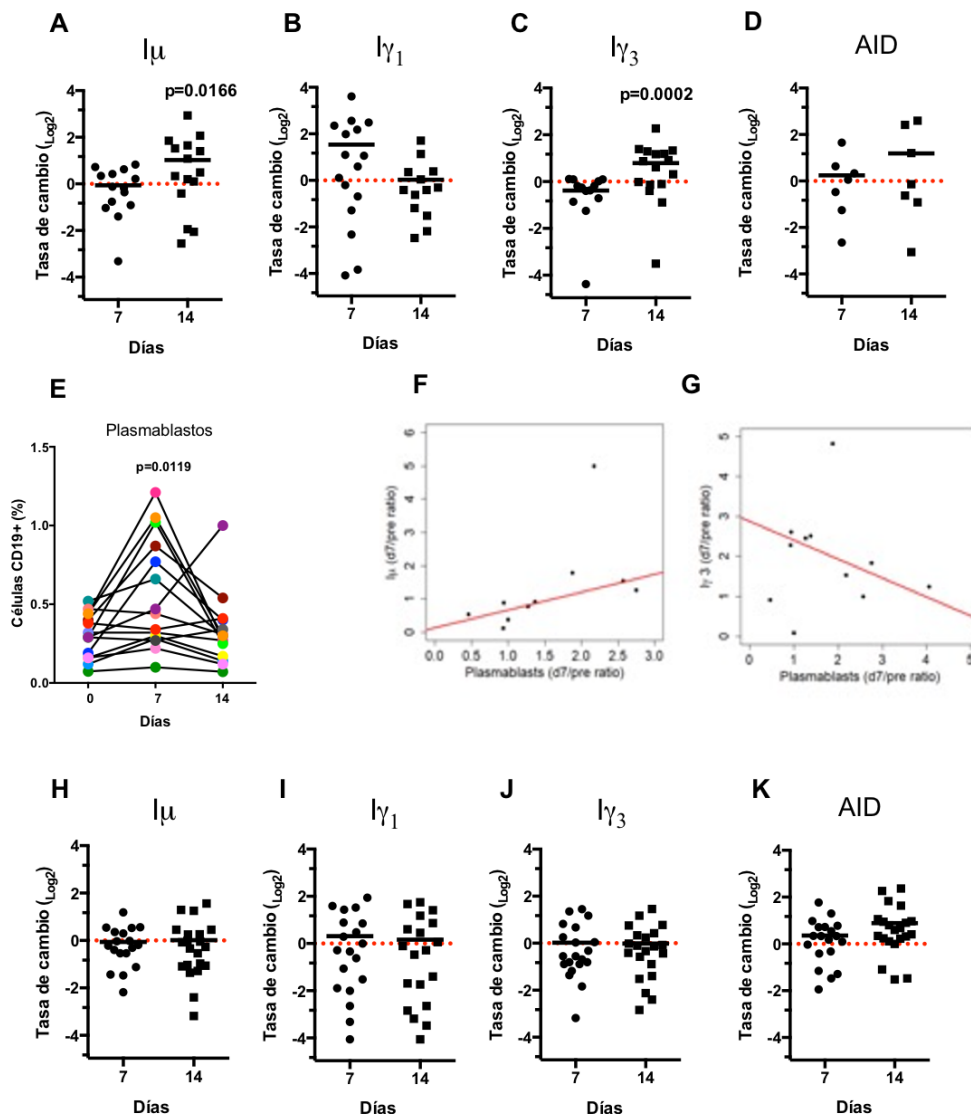


Figura 11. Expresión de transcritos RCCnc en individuos vacunados (A-D) qPCR para $I\mu$, $I\gamma_3$, $I\gamma_1$ y AID a partir del ARN total obtenido de las PBMC de donantes vacunados contra la hepatitis B y / o tétanos-difteria, tomadas en la preinmunización (día 0) y en los días 7 y

14 post-inmunización. (Prueba de Wilcoxon. $P < 0.05$). **(E)** Movilización de plasmablastos (CD3-CD19 + CD20-CD27 + CD38 +) en sangre periférica expresada como porcentaje de células B CD19 + (prueba de Wilcoxon, $p = 0,005$). **(F)** Correlación positiva entre la relación de plasmablastos del día 0/7 y la relación de $I\mu$ del día 0/7 (método de regresión LTS. R^2 ajustado: 0,53, valor de p : 0,015). **(G)** Correlación negativa entre la proporción de plasmablastos del día 0/7 y la proporción de $I\gamma 3$ del día 0/14 (método de regresión LTS. R^2 ajustado = 0,63, valor de p : 0,011). **(H-K)** Transcripción de RCCnc evaluados por la qPCR 7 y 14 días después de la vacunación contra la influenza inactivada trivalente (prueba de Wilcoxon. $P > 0.05$). **(H)** $I\mu$, **(I)** $I\gamma 3$, **(J)** $I\gamma 1$ y **(K)** AID. La línea roja punteada indica transcripción base Fold-change = 1.0.

La inmunización promueve la movilización de plasmablastos antígeno-específicos a sangre periférica alrededor del día 7 post-vacunación (Odendahl et al., 2005; Wrammert et al., 2008), y se ha propuesto que los plasmablastos derivan de centros germinales (Frolich et al., 2010). Por lo tanto, hipotetizamos que podía existir una correlación entre el nivel de transcripción RCCnc en sangre periférica y la proporción de plasmablastos circulantes. Evaluamos la transcripción de RCCnc ($I\mu$, $I\gamma 3$ $I\gamma 1$) y AID, así como la proporción de plasmablastos en sangre periférica de individuos sanos antes (día 0), a los 7 y a los 14 días después de la vacunación solo contra hepatitis B (HB)(la que contiene el antígeno de superficie recombinante) o en combinación con la vacuna contra el tétanos / difteria (TT / Dp)(vacuna que contiene las toxinas tetánica y diftérica modificadas) (**Figura 11A-G**). La transcripción de $I\mu$ e $I\gamma 3$ aumentó modestamente a los 14 días después de la vacunación con ambas vacunas (HB y TT/Dp) con respecto a los niveles pre-inmunes (**Figura 11A-C**) y fue significativamente mayor que al día 7 ($p < 0.05$). La transcripción de $I\gamma 1$ se incrementó al día 7 post-vacunación. Los niveles de plasmablastos alcanzaron su punto máximo al día 7 después de la vacunación, y volvieron a los niveles previos a la vacunación 14 días después (**Figura 11E**), conforme a lo descrito previamente (Odendahl et al., 2005; Wrammert et al., 2008). Aunque hubo una correlación positiva entre el aumento de plasmablastos y la transcripción de $I\mu$ al día 7 (**Figura 11F**. Método LTS, R^2 ajustado = 0.53; F-estadístico = 10.17 en 1 y 7 DF, $p = 0.015$), no hubo correlación con el cambio de expresión de $I\gamma 1$. Adicionalmente, $I\gamma 3$ correlacionó

negativamente con el aumento de plasmablastos (**Figura 11G**, método LTS, R^2 ajustado = 0.63, estadística $F = 13.04$ en 1 y 6 DF, $p = 0.011$).

A diferencia de los virus de Hepatitis B o las bacterias del tétanos y difteria, el virus de influenza es un virus que muta constantemente, por lo cual las vacunas contra influenza se renuevan sistemáticamente (“deriva antigénica”). Por lo tanto, también evaluamos la transcripción de RCCnc en individuos vacunados contra influenza. Sin embargo, en este caso no detectamos cambios en la transcripción de $I\mu$, $I\gamma1$ e $I\gamma3$ al día 7 ni al día 14 post vacunación (**Figura 11H-K**). En general, estos resultados sugieren que el aumento en la transcripción RCCnc en sangre periférica depende del tipo de vacuna y que la movilización de plasmablastos a sangre periférica contribuye poco o nada a la transcripción RCCnc en sangre periférica.

Hallazgos principales

- *Mayor transcripción de exones I_H in vitro ante activadores de respuesta tipo centro germinal.*
 - *La vacunación contra hepatitis B se asocia con movilización de plasmablastos e incremento de $I\gamma3$, mientras que la vacunación contra influenza no induce cambios significativos en los transcritos, a pesar de que si induce movilización de plasmablastos.*
-

La transcripción RCCnc es prevalente en el set de recount2, y predomina la transcripción de $I\mu$. Con la finalidad de explorar en forma panorámica la expresión de I_H 's en el set de datos que comprende recount2, las cuentas normalizadas (RPKM) obtenidas con *recount2* se utilizaron para evaluar la transcripción RCCnc en los conjuntos de datos SRA, TCGA y GTEx. En una fracción sustancial de todo el conjunto de datos, no encontramos transcripción RCCnc (RPKM = 0) en ninguno

de los 10 exones de I_H ($n = 26,512$ muestras, 37%) (**Tabla 2**). Se encontró una mayor transcripción (\log_2 RPKM), así como una transcripción más generalizada (proporción de muestras) para I_μ para todos los conjuntos de datos (SRA, TCGA y GTEx) (**Figura 12**). Esto respalda el modelo actual de transcripción RCCnc en el que I_μ es constitutivamente activo, y los otros I_H están regulados de acuerdo con señales microambientales particulares. El promedio de I_H RPKM \log_2 -transformada (2.65) para cada muestra de RPKM distinta de cero ($n = 44,091$; 62,4%) se utilizó para definir un límite de expresión. La transcripción RCCnc "alta" se definió como una RPKM \log_2 promedio de 2.65 o superior. El 29.8% de las muestras de RNA-seq ($n = 21,017$) estuvo por encima de este límite de expresión. La transcripción "baja" de RCCnc se definió como una RPKM \log_2 promedio $< 2,65$ (**Figura 13A**) que representó el 32,7% restante de las muestras ($n = 23,074$). Los niveles de transcripción de la RCCnc variaron de acuerdo con el I_H .

Métricas generales de corridas de secuenciación utilizadas en el análisis de expresión de RCCnc

	\log_2 RPKM	Muestras	(%)
Total	NA	70,603	100.0
Alta expresión	> 2.65	21,017	29.8
Baja expresión	$< 2.6, > 0$	23,074	32.1
No expresión	0	26,512	37.1
Utilizado para agrupamiento	> 0	44,091	62.9

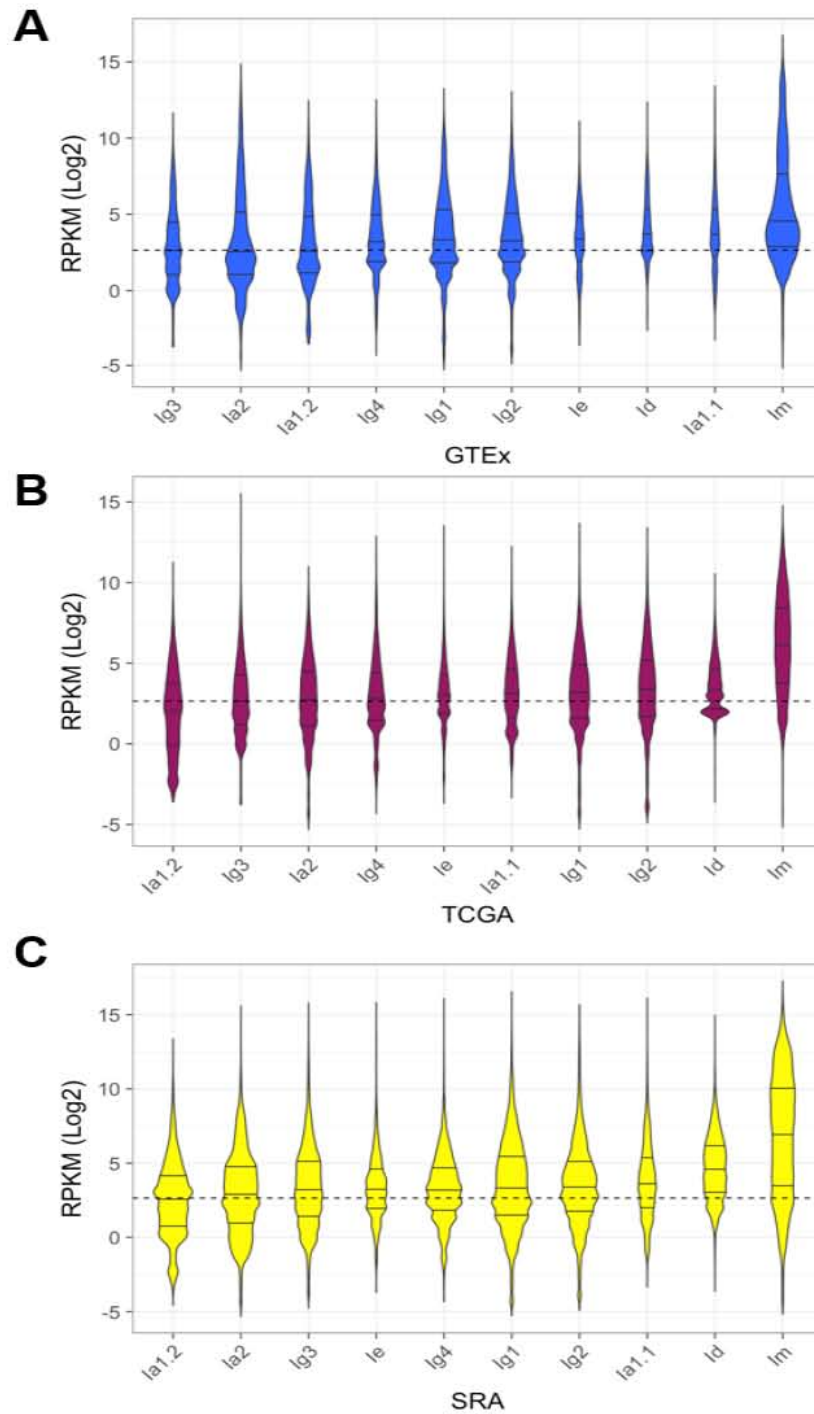


Figura 12. Transcripción CSRnc por I_H y de acuerdo al conjunto de datos del proyecto. Diagramas de violín de la distribución RPKM \log_2 por transcripción RCCnc (I_H) para (A) GTEx, (B) TCGA y (C) SRA conjuntos de datos. El área de cada violín está escalado respecto al número de muestras.

El set de recount2 se agrupa en perfiles transcripcionales bien definidos. El microambiente de las citocinas y otras señales regulan la RCC. Para identificar los perfiles transcripcionales de la RCCnc, que pueden resultar de diferentes microentornos y, así identificar conjuntos de muestras con expresión de ciertos I_H en particular, caracterizamos la transcripción RCCnc de acuerdo con el patrón de transcripción relativo de los distintos I_H . Con un universo de datos tan grande (> 70,000 muestras) y un amplio rango en el RPKM de las muestras para cada uno de los diferentes I_H , convenimos utilizar únicamente las muestras que tuvieran un RPKM diferente de 0 para poder agruparlas por perfiles de expresión en 10 conjuntos predefinidos (Z0-Z9), mediante el agrupamiento de *k*-medias (**Figura 13B-C**). Utilizando al grupo Z0 como referencia, cada uno de los I_H mostró un perfil de expresión distintivo de los grupos restantes de Z (\log_2 análisis de regresión RPKM. F-estadística: valor p: <2.2e-16). Los grupos Z0, Z2 y Z8 se caracterizaron por una baja transcripción de RCCnc (media \log_2 RPKM <2.65 o puntuación Z <0) en todas las clases de I_H . Los grupos Z4, Z6 y Z7 se caracterizaron por una expresión "alta" de I_μ solamente (RPKM \log_2 promedio > 2,65). El grupo Z3 mostró una expresión "alta" en I_μ e $I_{\alpha 2}$. El grupo Z1 mostró alta expresión en I_μ , $I_{\gamma 1}$ e $I_{\gamma 4}$. Finalmente, los grupos Z5 y Z9 mostraron una expresión "alta" en todos los I_H (**Figura 13C**)

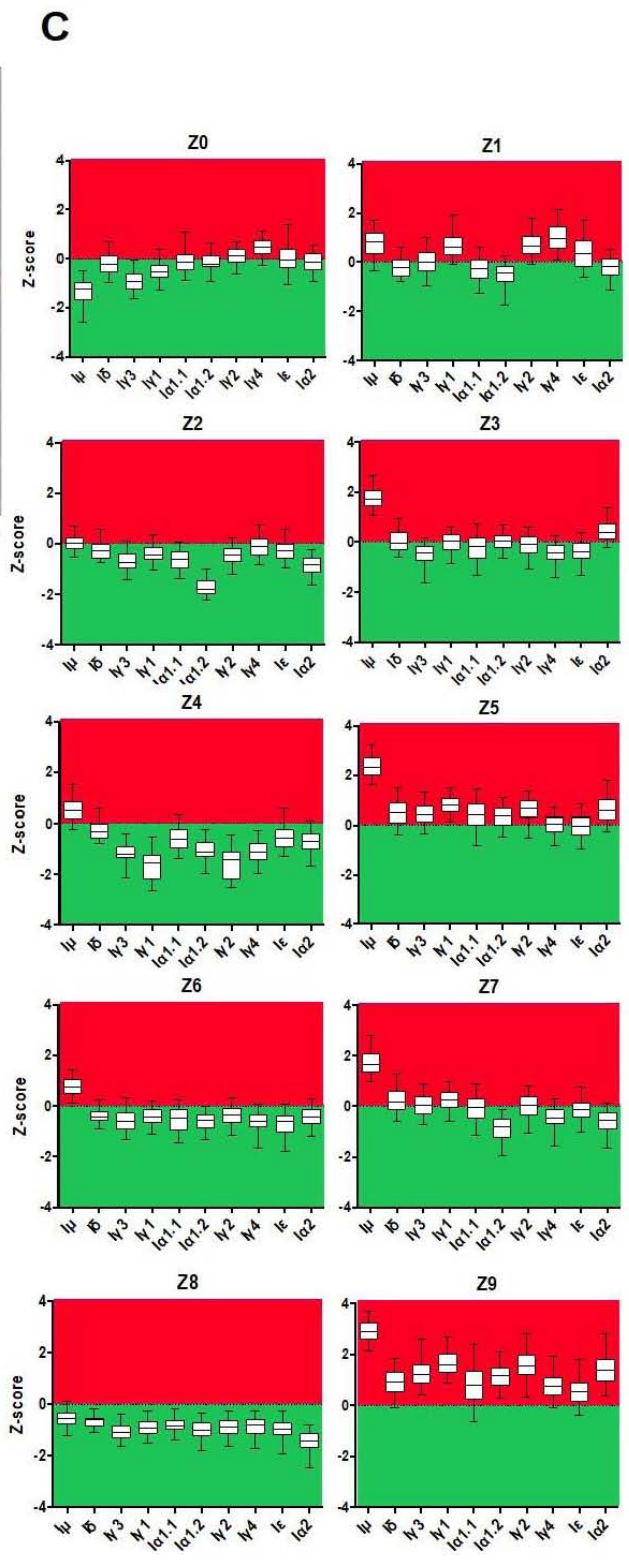
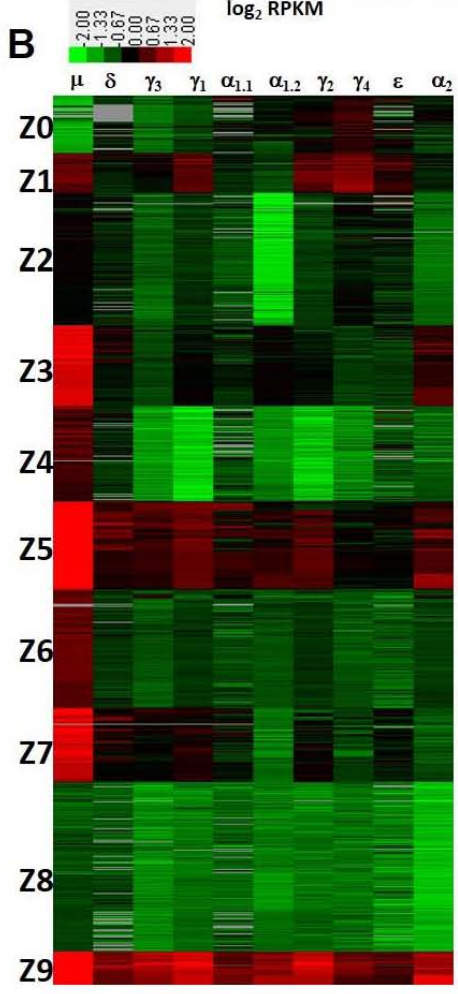
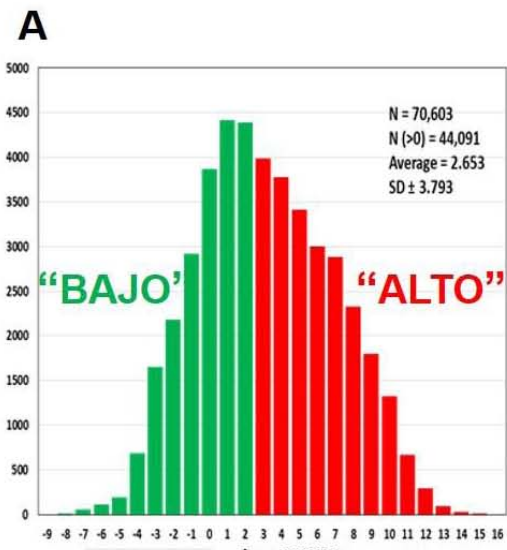


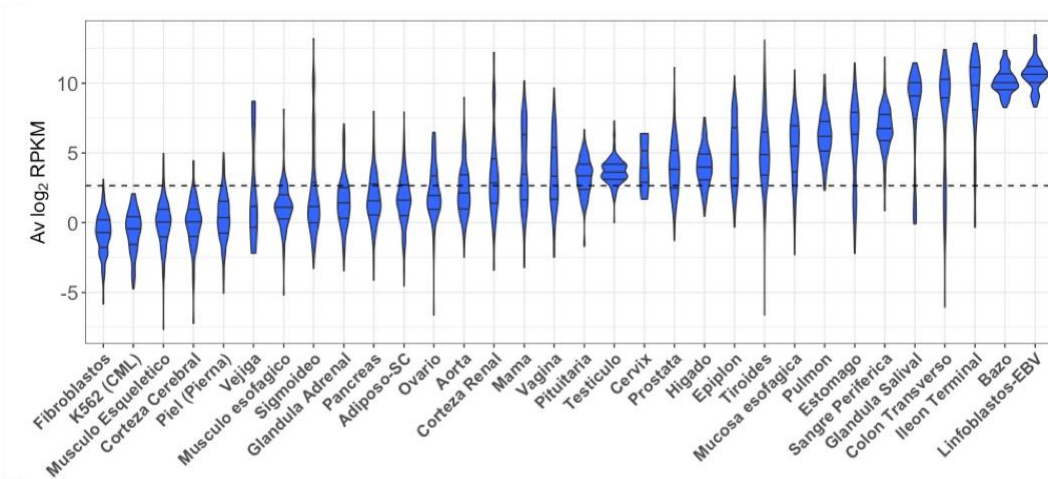
Figura 13. Análisis cuantitativo de la transcripción RCCnc utilizando *recount2*. (A) Distribución RPKM media transformada en \log_2 de los 10 exones I_H por muestra. Hubo 44.091 muestras con RPKM no cero. "Altamente" transcrito (\log_2 RPKM > 2.65) (mostrado en rojo) y "bajo" (\log_2 RPKM) <2.65 (verde). (B) RCCnc perfil de transcripción por agrupación de valores Z (Z-score). Los valores de transcripción \log_2 RPKM RCCnc de los conjuntos de datos GTEx + TCGA + SRA se transformaron en valores Z estandarizados y se sometieron a agrupación de k-medias con un número predefinido de 10 grupos (panel izquierdo). Los datos agrupados se representan en un mapa de calor donde los I_H son columnas y los grupos Z están en filas. Los valores Z negativos (es decir, <2.65 \log_2 RPKM) se muestran en verde, los valores Z positivos (es decir, > 2.65 \log_2 RPKM) se muestran en rojo. Los valores de ausencia (RPKM = 0) se muestran en gris. Los valores de Z cercanos a 0 se muestran en negro. (C) Patrón de expresión de cada grupo representado en diagramas de caja. La distribución del valor Z (eje y) por I_H (eje x) para cada grupo de los diez grupos generados. La línea de puntos indica un valor Z = 0, que corresponde a la expresión media por muestra (\log_2 RPKM = 2.65). La transcripción superior a la media se marca en rojo, mientras que la media se marca en verde.

La transcripción RCCnc está ampliamente distribuida en tejidos sanos y su perfil depende del tejido. El microambiente en el que se desarrolla la respuesta humoral influencia la elección de cambio de isotipo. Para evaluar los patrones transcripcionales de RCCnc, presentes en tejidos sanos de humanos adultos, utilizamos las muestras del proyecto GTEx como referencia. El promedio de RPKM por muestras de tejido diferentes de cero se clasificó según su media de RPKM \log_2 . Encontramos que la transcripción media más alta estaba presente en tejidos linfoides tales como bazo, linfocitos B transformados por EBV y sangre completa, como era esperado ya que son tejidos donde se encuentran células B de manera abundante (Chaudhuri and Alt, 2004; Dudley et al., 2005; Kenter, 2005).

Sin embargo, también en órganos con tejido linfoide asociado a la mucosa (MALT) como íleon terminal, colon transversal, estómago, pulmón y mucosa esofágica se observó una alta transcripción media (**Figura 14A**). Se observó una notable diferencia en el promedio RPKM \log_2 entre colon transversal (media 7.5 ± 3.9) y el colon sigmoide (media 2.1 ± 3.2). Interesantemente, la expresión de las glándulas salivales se encontró entre los tejidos con mayor transcripción, y tejidos no mucosos como la glándula tiroidea y la glándula pituitaria, mostraron una transcripción de I_H media alta. Otra diferencia notable en RPKM \log_2 se observó entre tejido adiposo visceral (omento, media 4.9 ± 2.3) y tejido adiposo subcutáneo

(media $1.9 \pm 1,5$). Por el contrario, muestras de tejidos tales como cerebro, músculo esquelético y cardíaco, piel, así como en la línea celular de leucemia mieloide crónica K562 y líneas celulares de fibroblastos dérmicos transformados, mostraron los niveles más bajos de transcripción de RCCnc (RPKM \log_2 promedio < 2.65) (Figura 13 A).

A



B

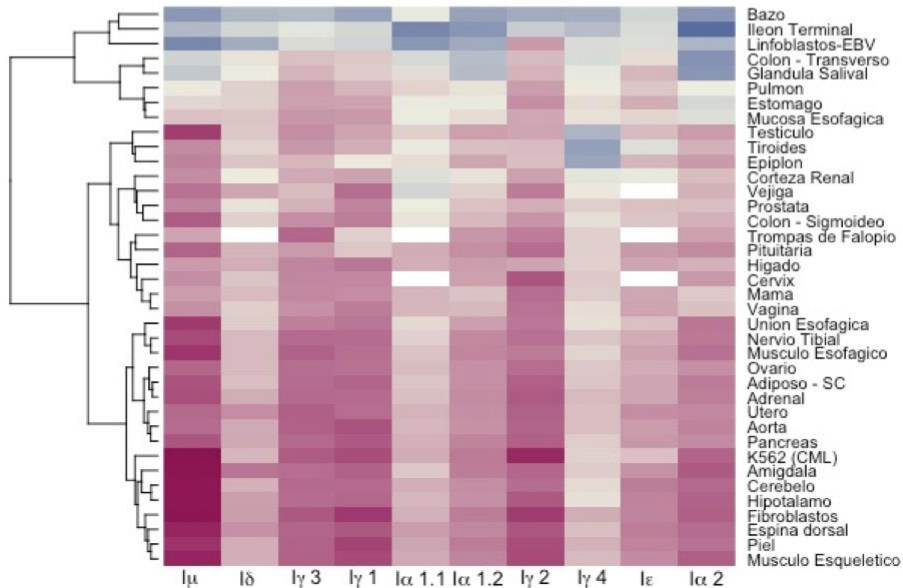


Figura 14. Transcripción de RCCnc en tejidos adultos sanos. (A) El conjunto de datos GTEx se dividió según el tipo de tejido. Diagrama de violín de la distribución log2RPKM promedio de los tejidos representativos del proyecto GTEx, ordenados de izquierda a derecha según cada log2RPKM mediano de tejido. El área del violín se escaló según el recuento de la muestra y se muestran la mediana y los cuartiles. Una línea negra punteada marca el promedio de log2RPKM (2.65). Para simplificar, solo se incluyó la corteza cerebral como una muestra representativa del sistema nervioso central. (B) Análisis de regresión de la transcripción RCCnc en función del tipo de tejido y el exón I. La transcripción RCCnc de sangre total se utilizó como tejido de referencia (filas) para el análisis de regresión de acuerdo con I-exon (columnas). La estimación para cada comparación se expresa como un mapa de calor. Una transcripción RCCnc mayor que en la sangre se representa en tonos azules, mientras que la transcripción inferior se muestra en tonos rosados. Los valores de estimación cero se muestran en marfil. Los valores faltantes (NA) se muestran en blanco. La distancia euclidiana se utilizó para agrupamiento jerárquico por fila. Por simplicidad, solo se muestran los tejidos representativos.

Considerando que los tejidos se encuentran irrigados por sangre y para distinguir entre la transcripción presente en sangre y el tejido específico, realizamos un análisis de regresión lineal de cada I_H por tipo de tejido, utilizando sangre periférica como tejido de referencia. En sangre periférica, el patrón de transcripción de cada I_H fue significativamente diferente ($p = < 2.2e-16$). La transcripción de I_μ fue la más elevada en todos los tejidos y predominó en bazo, íleon terminal, glándula salival y colon transversal. Aunque no se detectó ninguna diferencia en la transcripción de I_μ entre sangre y pulmón ($p = 0.23$), para el resto de los tejidos se detectó significativamente menos transcripción I_μ que en sangre ($p < 0.05$). Es de destacar que la transcripción I_μ en testículo fue particularmente baja, a pesar de una transcripción promedio alta. La transcripción I_δ es similar a la transcripción I_μ , pero solo en bazo e íleon terminal fue significativamente más alta. Por su parte, $I_\gamma3$ e $I_\gamma1$, $I_\gamma2$ compartieron un perfil transcripcional similar, con alta expresión en bazo e íleon terminal con respecto a sangre ($p < 2e-16$) (**Figura 14B**).

Para la mayoría de los tejidos, la transcripción de $I_{\alpha1.1}$ e $I_{\alpha1.2}$ se encontró altamente correlacionada, sin embargo, la transcripción de $I_{\alpha1.1}$ fue mayor en íleon terminal y colon transversal que en sangre ($p < 0.001$), mientras que la transcripción de $I_{\alpha1.2}$ fue mayor en bazo, glándula salival y riñón ($p < 2e-16$). En función de lo

anterior, parece que la RCC a *IGHA1* puede implicar el inicio y / o *splicing* de la transcripción alternativa dependiente del tejido (**Figura 14B**).

El perfil transcripcional de $I\gamma 4$ fue el más particular, ya que se observó alta expresión en tejidos glandulares como tiroides y testículo, y epiplón (omento) con respecto a sangre ($p < 3.9e-05$). En cuanto a órganos linfoides, la transcripción de $I\gamma 4$ fue mayor en bazo, íleon terminal, colon transverso, glándula salival, estómago y mucosa esofágica que en sangre ($p < 1.1e-05$). No se observaron diferencias entre sangre y pulmón ($p > 0.05$) (**Figura 14B**).

Para identificar si el perfil de expresión de RCC por tejido es estereotipado o existen diferencias entre individuos sanos, de los 10 perfiles de expresión identificados mediante agrupamiento de k-medias (cluster Z0-9), probamos como estaban distribuidos los perfiles de expresión de RCCnc, de acuerdo a un tejido en particular. Consistentemente con el análisis de regresión de RCCnc según la expresión relativa de I_H (*cluster Z*), reveló que el íleon terminal, bazo, colon transverso, sangre total y glándula salival comparten un patrón de expresión similar y que se encuentran altamente enriquecidos en los grupos Z5 y Z9 (alta transcripción en todos los I_H) (FDR < 0.001) (**Figura 15**). Algunos tejidos con MALT como el estómago, la mucosa esofágica y el pulmón compartieron un patrón de transcripción I_H similar y predominan en el grupo Z5 y Z3 (alta transcripción I_μ e $I\alpha 2$) (FDR < 0.001), mientras que otros como mama, vagina e hígado están enriquecidos en los grupos Z3 y Z6 (solo I_μ) (FDR < 0.001) (**Figura 15**). De particular interés fue que el tejido adiposo visceral, la tiroides y la pituitaria que predominan en el cluster Z1 (alta expresión de $I\gamma 1$ e $I\gamma 4$). Estos resultados indican diferencias en el perfil de expresión en ciertos tejidos de individuos sanos, lo cual sugiere que la RCCnc puede estar determinada por variación genética regulatoria, exposición ambiental o ambas.

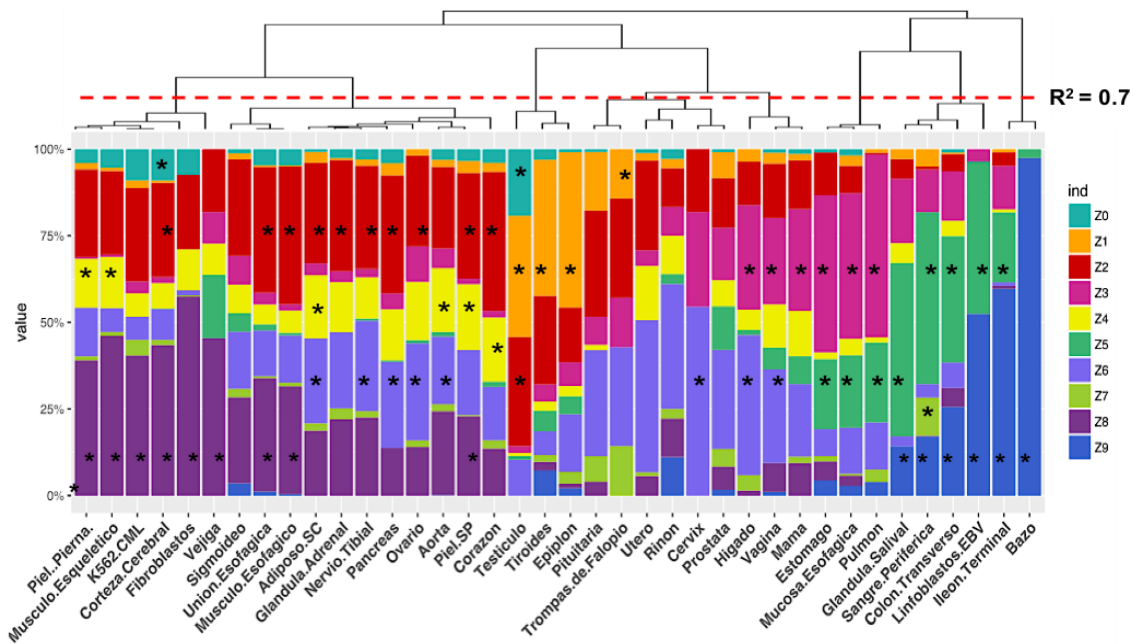


Figura 15. Variación del perfil de transcripción RCCnc en tejido sano. El conjunto de datos GTEx se clasificó según el tejido (eje x) y la proporción relativa de muestras (eje y) que pertenecen a cada grupo Z (en colores). Las proporciones se agruparon jerárquicamente utilizando la correlación de Pearson como métrica de distancia con Cluster3.0 (de Hoon et al., 2004). La línea roja punteada marca una correlación $R^2 = 0.7$. El asterisco negro indica un enriquecimiento significativo de un tejido dado en un grupo Z particular (prueba de Fisher, ajuste de Benjamini-Hochberg para corrección múltiple. FDR < 0.01).

La distribución anatómica observada y la abundancia de la transcripción de RCCnc sugieren que la cantidad de transcripción de RCCnc puede ser dependiente de la abundancia de células B presentes en un tejido dado. Usamos el RPKM \log_2 de la transcripción C_H codificante como un proxy de la cantidad de células B en el tejido. En general, la transcripción C_H codificante fue de 10 a 100 veces mayor que la transcripción RCCnc. Para correlacionar la transcripción de RCCnc con la transcripción C_H correspondiente, se compararon sus valores Z. Como se esperaba para cada clase, el valor Z de RCCnc y C_H correlacionaron significativamente ($p < 1e-16$), lo que sugiere que a mayor número de células B o células plasmáticas en un tejido dado, mayor transcripción RCCnc. Sin embargo, notamos que, para la mayoría de las clases, una fracción de las muestras se desvía de la correlación ortogonal esperada, lo que indica una mayor transcripción RCCnc en relación con la transcripción C_H . Esto es particularmente notable para I_{μ} , pero también, $I_{\gamma 1}$, $I_{\gamma 4}$ y

$I_{\alpha 2}$ (Figura 16). Estos resultados sugieren que esta expresión relativa más alta de I_H con respecto a C_H indica RCC activa.

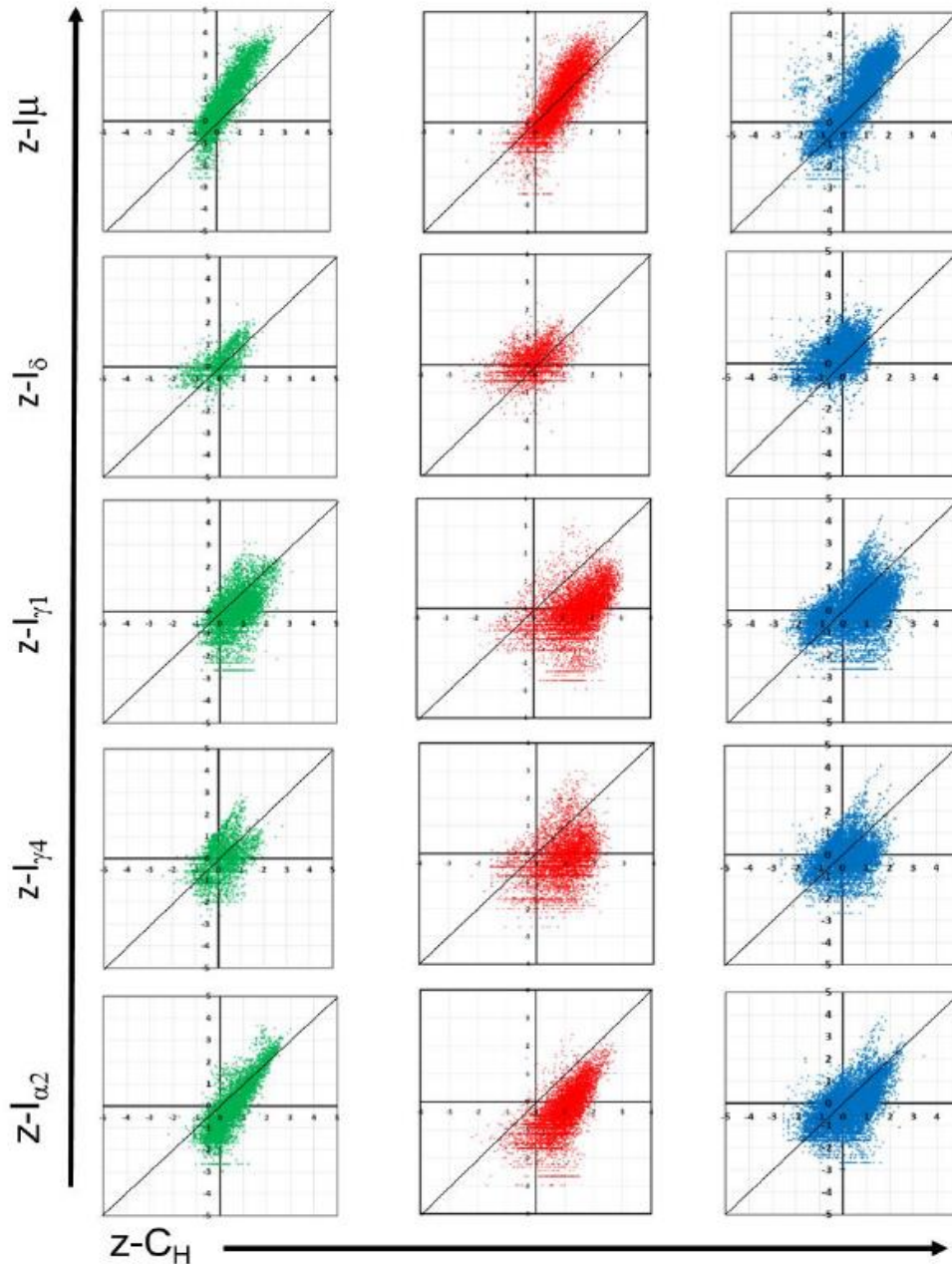


Figura 16. Correlación entre la transcripción RCCnc I_H y la transcripción C_H codificante. En cuanto a las transcripciones RCCnc I_H , la codificación C_H log2RPKM correspondiente se transformó en puntajes Z para hacer comparaciones de sus respectivas expresiones relativas en dos dimensiones (valor Z C_H , eje x, valor Z I_H , eje y). Se muestran

gráficos representativos para I / C μ , I / C δ , I / C γ 1, I / C γ 4 y I / C α 2 para GTEX (verde), TCGA (rojo) y SRA (azul). La diagonal negra se muestra para enfatizar las desviaciones del plano ortogonal. Las muestras por encima de la diagonal indican una transcripción RCCnc relativa más alta que su correspondiente homólogo de codificación C_H.

I μ se expresa en precursores linfoides tempranos e I γ 4 en tejidos fetales.

La RCC es un mecanismo que se da en linfocitos B maduros activados por el antígeno. Sin embargo, el locus de inmunoglobulina es transcripcionalmente muy activo desde el compromiso al linaje linfocito B. Por lo tanto, investigamos la transcripción de RCCnc a lo largo del desarrollo linfocito temprano usando datos del estudio SRP058719, que aborda la transcripción en la diferenciación linfocito temprana antes y después del compromiso del linaje de células B y T usando RNA-seq de distintos precursores linfocitos aislados mediante citometría de flujo (Casero et al., 2015). Observamos que los precursores del linaje B y T expresaron I μ . Células madre hematopoyéticas enriquecidas (CD34⁺ CD38⁻ lin⁻ HSC), progenitores linfocitos multipotentes (LMPP's, CD34⁺ CD38⁺ CD10⁻ CD45RA⁺ lin⁻), progenitores linfocitos comunes o CLP's (CD34⁺ CD38⁺ CD10⁺ CD45RA⁺ lin⁻), precursores tímicos CD34⁺ CD7⁻ CD1a⁻ CD4⁻ CD8⁻ (Thy1) y precursores B completamente comprometidos (CD34⁺ CD38⁺ CD19⁺) expresaron I μ , indicando que además de la expresión de los transcritos provenientes de las células B que recirculan la sangre, los precursores hematopoyéticos, incluidos los precursores de las células B, contribuyen a la expresión de RCCnc en la médula ósea (**Figura 17A**).

Adicionalmente, la expresión de RCCnc se analizó en tejidos fetales mediante el uso de los datos del proyecto SRP055513, que reportó un extenso análisis con RNA-seq de 20 tejidos fetales durante las semanas gestacionales 9-22 (Roost et al., 2015). A pesar de que no se detectó transcripción RCCnc para la mayoría de los isotipos, ni transcripción de IGH, sorprendentemente, se detectó una transcripción robusta de I γ 4, aunque no se detectó expresión de IGHG4 en todos los tejidos fetales analizados, independientemente de la semana de gestación, (**Figura 17B**). La mayor expresión de I γ 4 promedio se encontró en el bazo, seguida de pulmón e hígado, siendo que este último tiene una función linfohematopoyética

respectivamente. Se observa una expresión de $I_{\gamma 4}$ superior a la media en todos los tejidos analizados. Código de color de la puntuación Z como en (A).

Hallazgos principales

- *Hay diferencias importantes en la expresión de I_H en diferentes órganos siendo mayor en tejidos linfoides secundarios*
 - *Hay correlación entre la expresión de I_H con su respectivo C_H . Sin embargo, en diferentes proyectos se observan muestras donde la expresión relativa de I_H es mayor que C_H*
 - *Destaca un perfil de expresión caracterizado por la expresión de $I_{\gamma 1}$ y $I_{\gamma 4}$ en testículo, tiroides, pituitaria y tejido adiposo visceral (epiplón)*
 - *Hay variabilidad en la expresión de I_H 's en individuos sanos en ciertos tejidos, lo que puede indicar diferencias genéticas o ambientales o una mezcla de ambas.*
 - *La expresión de I_{μ} se expresa desde los precursores linfoides T y B*
 - *$I_{\gamma 4}$ tiene expresión en órganos fetales linfoides y no linfoides cuya función biológica desconocemos.*
-

Transcripción de RCCnc en cáncer. Los datos del proyecto TCGA se usaron como referencia para estudiar la transcripción de RCCnc en una amplia variedad de neoplasias en humanos. Así como en el caso de GTEx, analizamos la transcripción de RCCnc (promedio \log_2 RPKM de I_H) en 33 tipos de cáncer (**Figura 18**). La transcripción de I_H en el linfoma difuso de células B grandes (DLBCL) y leucemia mieloide aguda (AML) fueron las más altas (**Figura 18**). En general, la transcripción de RCCnc en tejido neoplásico es similar a su contraparte no neoplásica, presentando alta transcripción en carcinoma de pulmón, estómago y testículo. Por

el contrario, la transcripción de RCCnc fue baja en tumores gliales y melanoma (**Figura 18**).

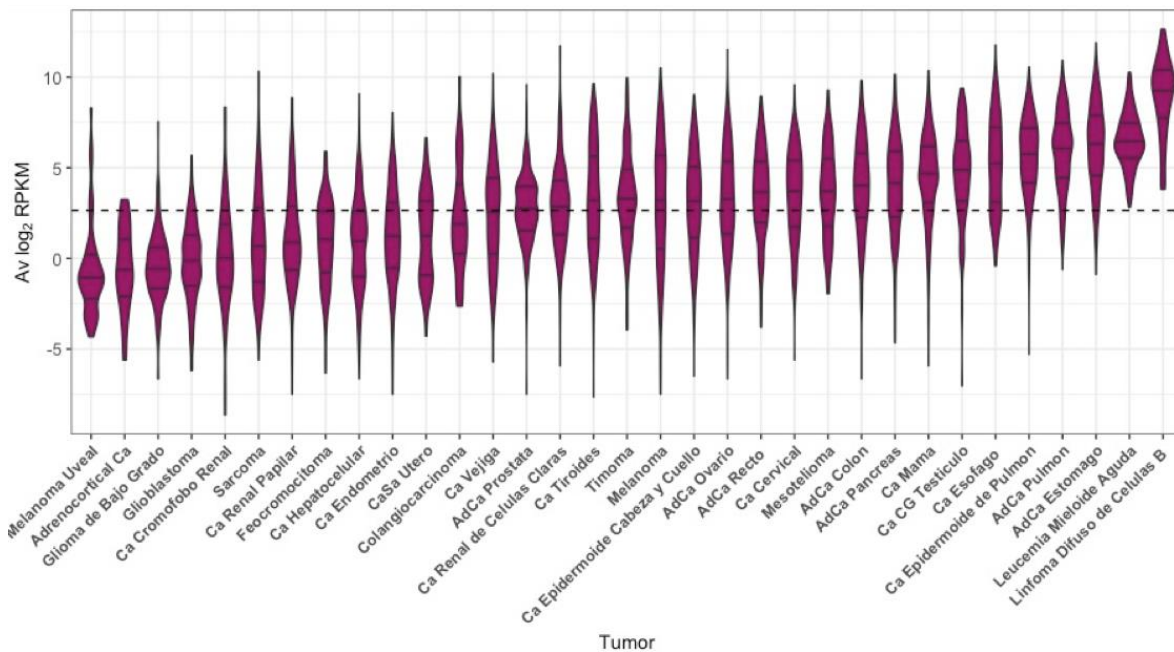


Figura 18. Transcripción de RCCnc en cáncer. Se utilizaron datos de RNA-seq del proyecto TCGA (Cancer Genome Atlas Research et al., 2013) para analizar la transcripción de RCCnc en 33 tipos de cáncer. A) Diagramas de violín de RPKM \log_2 promedio, ordenados de izquierda a derecha de acuerdo a la mediana. El área del violín no corresponde al número de muestras de cada tumor. Una línea negra punteada marca el RPKM \log_2 promedio = 2.65.

La comparación directa entre la transcripción de RCCnc en tejido sano (GTEx) y su contraparte de tejido neoplásico, permitió la identificación de tres patrones distintos (**Figura 19**): 1) Tumores donde la RCCnc promedio es menor que en tejido sano, como en DLBCL, cáncer de próstata, tiroides, hígado y colon (**Figura 19 A-E**). 2) Tumores donde la transcripción promedio de RCCnc fue más alta que su contraparte del tejido sano, como en mama, recto, células germinales testiculares, carcinoma de páncreas, cistadenoma ovárico y melanoma cutáneo (**Figura 19 G-L**). 3)

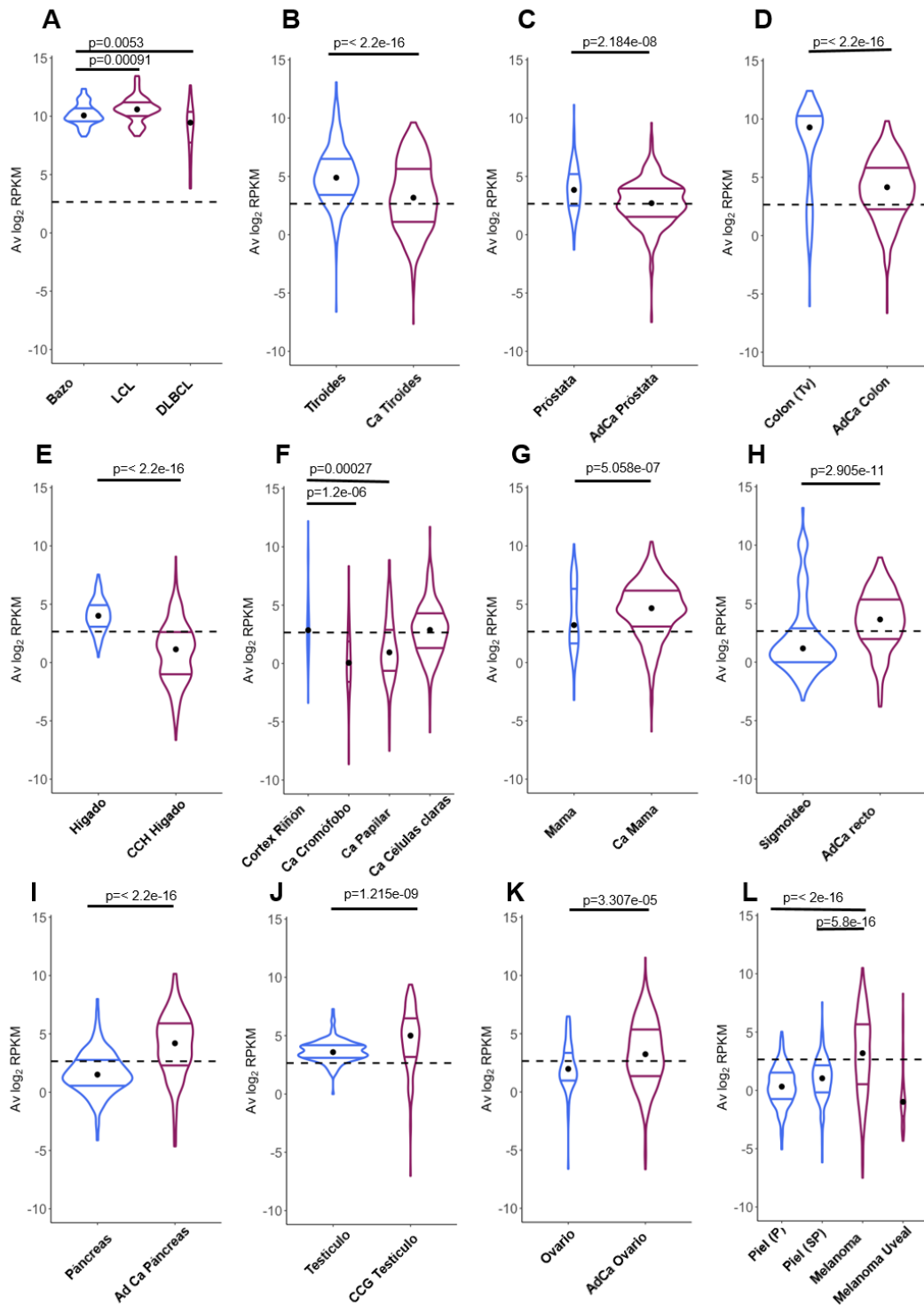


Figura 19. Comparación de la transcripción de RCCnc en tejido sano y su contraparte tumoral. Diagramas de violín de la distribución promedio de log₂RPKM en tejido sano (azul) en comparación con su contraparte de tejido canceroso (púrpura). El área del violín no se

escala para el recuento de la muestra. Mediana (punto negro) y cuartiles se muestran para cada violín. La línea negra discontinua marca el promedio de log₂RPKM (2.65). Se observó una reducción de la transcripción de RCCnc en tumores en (A-E). Se observó un aumento de la transcripción de RCCnc en tumores frente a tejido sano en (G-K). Algunos tipos de cáncer de riñón y melanomas mostraron patrones opuestos con respecto a la contraparte de tejido sano (F y L). La prueba estadística realizada fue la prueba de suma de rangos de Wilcoxon con corrección de continuidad para comparaciones de dos muestras, y la prueba de Kruskal-Wallis con corrección de prueba de Dunn *post hoc* para comparaciones múltiples.

Tumores sin diferencia en la transcripción de RCCnc entre tejido neoplásico y su homólogo sano, como en estómago y esófago (**Figura 20**). Interesantemente, utilizando la corteza renal sana como referencia para los tumores renales, se observó que la transcripción RCCnc varió de acuerdo con el tipo de cáncer renal: la transcripción RCCnc fue significativamente menor en carcinoma renal cromóforo y carcinoma papilar, pero no en el carcinoma de células claras. (**Figura 19F**). En la glándula suprarrenal, se observó un patrón similar, con una menor transcripción de RCCnc en el carcinoma adrenocortical, pero no en el feocromocitoma (**Figura 20**). La correlación entre RCCnc y la transcripción C_H (**Figura 15**) sugiere que, como para los tejidos sanos, la transcripción RCCnc deriva de infiltrados linfoides terciarios como resultado de la inflamación asociada al tumor y los correspondientes tejidos linfoides asociados a la mucosa (Colbeck et al., 2017). Para abordar esta cuestión, analizamos la transcripción de RCCnc en una amplia variedad de líneas celulares derivadas de tumores provenientes de proyectos en SRA. No se detectó transcripción RCCnc en líneas celulares derivadas representativas de tumor (células A549 de cáncer de pulmón, MCF-7 de cáncer de mama y HCT116 de cáncer de colon) (**Material Suplementario: Tabla S2**), indicando que la transcripción RCCnc en cáncer se deriva de la infiltración de células B y no de las células neoplásicas *per se*.

Para poder comprobar la robustez de nuestros resultados también utilizamos el conjunto de datos SRA. Usando ontologías de enfermedad a través de MetaSRA (Bernstein et al., 2017; Kibbe et al., 2015), identificamos un enriquecimiento significativo de la expresión de RCCnc en distintos tipos de cáncer como AML,

mama y pulmón en muestras de SRA , lo cual confirma nuestras observaciones derivadas del análisis de datos TCGA.

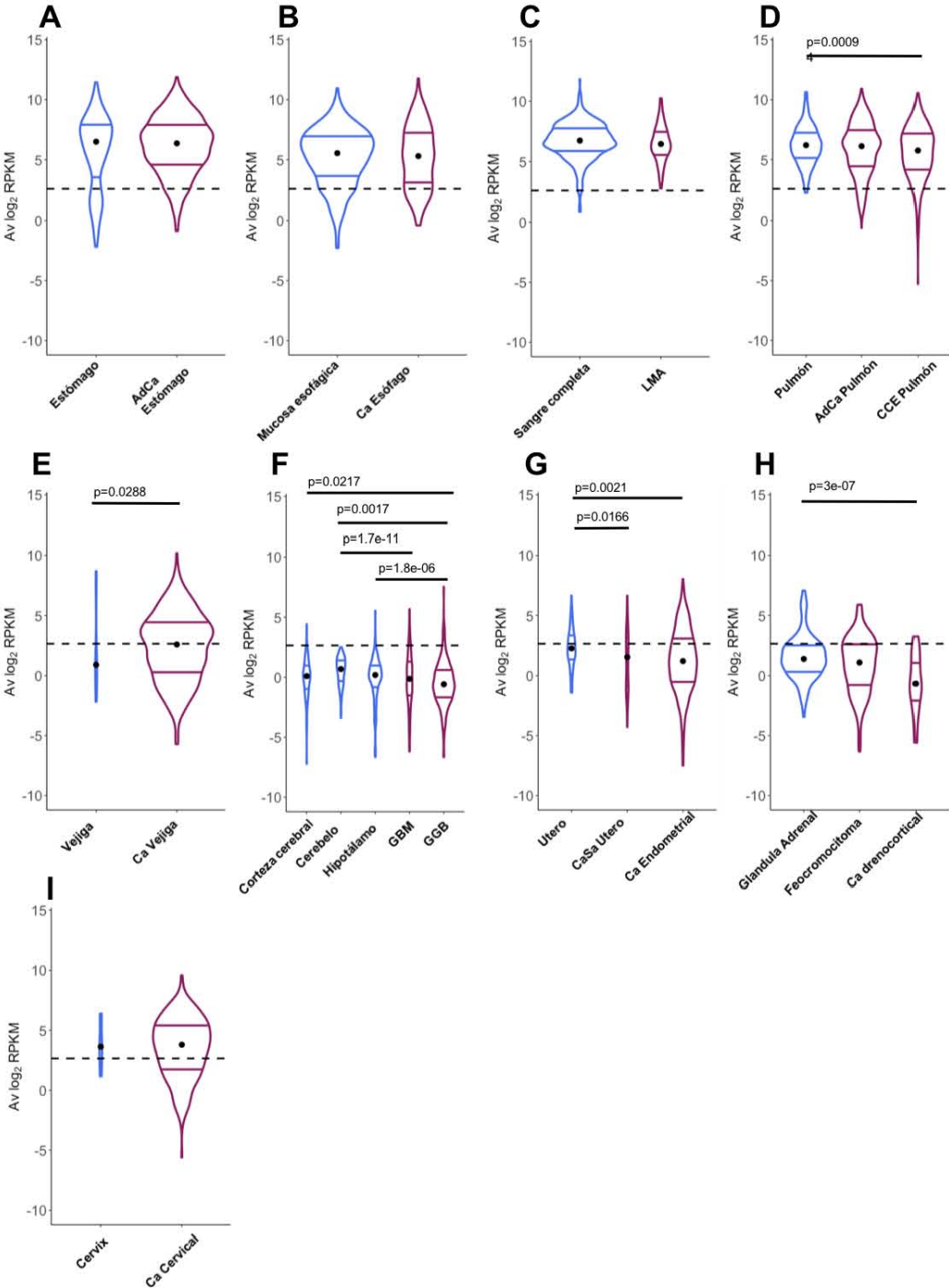


Figura 20. Comparación de la transcripción de CSRnc en tejido sano y su contraparte tumoral. Diagramas de violín de la distribución promedio de log2RPKM en tejido sano (azul) en comparación con su contraparte de tejido canceroso (púrpura). El área del violín corresponde al conteo de la muestra. Mediana (punto negro) y cuartiles se muestran para cada violín. La línea negra discontinua marca el promedio de log2RPKM (2.65). No se observaron diferencias en la transcripción de CSRnc en los tumores en comparación con su contraparte saludable en la A-D. No se detectaron diferencias en el adenocarcinoma sano de pulmón y pulmón, sin embargo, la transcripción CSRnc fue menor en el carcinoma de células escamosas de pulmón (E). También se observó una menor transcripción de CSRnc en los tumores del sistema nervioso central (F), carcinosarcoma uterino y carcinoma endometrial (G) y carcinoma adrenocortical, pero no feocromocitoma (H). Se detectó un aumento de la transcripción de CSRnc en el cáncer de vejiga (I). La prueba estadística realizada fue la prueba de la suma de rangos de Wilcoxon con corrección de continuidad para comparaciones de dos muestras, y la prueba de Kruskal-Wallis con corrección de la prueba de Dunn post hoc para comparaciones múltiples.

Hallazgos principales

- *La expresión de I_H's en tejido tumoral varía de acuerdo al origen del tejido y probablemente refleja el grado de infiltración de linfocitos B. Además en líneas celulares cancerígenas no se observó expresión de I_H's, reforzando la idea que la expresión proviene de los linfocitos infiltrantes y no del tumor mismo.*
- *En algunos tejidos tumorales no observamos diferencia significativa en la expresión global de I_H's con su contraparte sana (cervix, estómago, esófago). Por otro lado, para tiroides, próstata, hígado y colon la expresión de I_H's fue mayor, mientras que en cáncer de mama, testículo, páncreas y ovario la expresión de I_H's fue menor. Esto nos puede sugerir que el microambiente*

Una firma transcripcional de I_H se asocia a la enfermedad de Crohn pediátrica con ulceración profunda: El SRA es un reservorio de muestras muy diversas, por lo que es fuente de información sobre la presencia y expresión de I_H en condiciones

patológicas de distinto origen (infeccioso, no-infeccioso o autoinmune). A diferencia de GTEX y TCGA, algunos proyectos contenidos dentro del reservorio de SRA permitieron realizar el análisis de expresión diferencial de los distintos I_H's, ya que tenían tanto muestras control como muestras patológicas.

Tabla 3. Análisis de expresión diferencial de C _H e I _H													
Proyecto	Descripción	Comparación (Exp vs control)	Muestra	valor p*									
				I _μ	I _δ	I _{γ3}	I _{γ1}	I _{α1.1}	I _{α1.2}	I _{γ2}	I _{γ4}	I _ε	I _{α2}
SRP032775	Malaria	Post-infección vs Pre-infección	Sangre	0.92	0.14	0.15	0.05	1.00	1.00	0.16	1.00	1.00	1.00
SRP033696	Influenza	Infección H7N9 vs control sano	Sangre	1.00	1.00	0.49	0.02	1.00	0.00	0.47	1.00	1.00	1.00
SRP059039	Diarrea pediátrica	<i>Shigella</i> vs Control	Sangre	0.03	0.74	0.13	0.01	1.00	0.89	0.51	0.03	1.00	1.00
SRP059039	Diarrea pediátrica	<i>Salmonella</i> vs Control	Sangre	1.00	1.00	0.03	0.08	1.00	1.00	0.05	0.25	1.00	1.00
SRP059039	Diarrea pediátrica	Rotavirus vs Control	Sangre	1.00	1.00	1.00	0.43	1.00	1.00	0.05	0.36	1.00	1.00
SRP059172	Brucelosis/Leishmaniasis	<i>Brucella</i> vs Control	Sangre	0.42	0.00	1.00	1.00	1.00	0.72	0.61	1.00	0.35	1.00
SRP059172	Brucelosis/Leishmaniasis	<i>Leishmania</i> vs Control	Sangre	1.00	1.00	0.20	0.47	1.00	1.00	0.01	1.00	0.48	1.00
SRP042228	Crohn pediátrico	Crohn vs No-IBD	Biopsia ileal	0.00	0.03	1.00	1.00	0.01	0.54	0.91	1.00	1.00	0.02
SRP042228	Crohn pediátrico	Crohn vs No-IBD	Biopsia ileal	0.14	0.28	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SRP042228	Crohn pediátrico	Colitis Ulcerativa vs No-IBD	Biopsia ileal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86
SRP051688	Influenza	Post-inmunizado (día 7) vs pre-inmune	Células B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SRP020491	Influenza	Post-inmunizado (día 7) vs pre-inmune	PBMC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SRP062966	Lupus eritematoso generalizado	Lupus vs control sano	Sangre	0.93	0.82	1.00	1.00	1.00	1.00	0.39	1.00	0.48	0.77

*Valor p ajustado por Bonferroni. En verde, reducción significativa de expresión con respecto a control. En rojo, aumento significativo de expresión con respecto a control.

El análisis del conjunto de datos de SRA nos permitió identificar dentro de clúters de alta expresión (\log_2 RPKM > 2.65) de RCCnc a proyectos con muestras de enfermedades infecciosas como diarrea, brucelosis y malaria (**Prueba de Fisher**). Sin embargo, estos proyectos derivaban de muestras de sangre periférica. Para distinguir si el enriquecimiento implica el aumento de la expresión debido a la condición patológica o por el simple hecho de provenir de sangre periférica (**Figura 14B**), se realizó un análisis de expresión diferencial que comparó los grupos experimentales proporcionados contenidos en cada proyecto. Detectamos una reducción significativa de la transcripción RCCnc pero no en la transcripción de C_H correspondiente en enterobacterias, pero no en la enfermedad diarreica por rotavirus (SRP059039), malaria (SRP032775) (Tran et al., 2016) brucelosis y leishmaniasis (SRP059172) (**Tabla 3**). Estos resultados sugieren que en este tipo de infecciones se altera la capacidad de cambiar de clase.

Realizamos el análisis de expresión diferencial para influenza, donde utilizamos un proyecto de SRA con muestras de individuos vacunados contra la influenza (SRP020491) (Henn et al., 2013). Lo que observamos con el resultado fue consistente con nuestros propios datos experimentales, ya que no hubo cambios significativos en la transcripción de RCCnc 7 días después de la vacunación, ni tampoco encontramos cambios en los transcritos codificantes de RCC (Material suplementario, Tabla S3), independientemente de la onda de migración de plasmablastos a sangre periférica al día 7, previamente descrita (Wrammet, 2008). Esto indica lo que los transcritos de RCCnc observados en sangre periférica no se derivan de plasmablastos recientemente cambiados de clase (**Figuras 11H-K**). Sin embargo, lo observado en respuesta a la vacunación contra la influenza contrastó con el resultado del análisis de datos de un proyecto con muestras de una infección natural con H7N9 (Mei et al., 2014) (SRP033696), en el que observamos un aumento en la transcripción de RCCnc ($I\gamma 1$ e $I\alpha 1$) durante la infección aguda (**Tabla 3**). Además, se observó un aumento en la transcripción de transcritos codificantes IgM, IgA1, IgG2 e IgG4 (Tabla S3).

Dentro del conjunto de datos de SRA también, encontramos que había una alta expresión de I_H 's ($\log_2 < 2.65$) en proyectos sobre enfermedades autoinmunes, como lupus eritematoso sistémico y enfermedades autoinflamatorias como la enfermedad inflamatoria intestinal. No se observó expresión diferencial significativa en transcriptomas de sangre periférica en pacientes con lupus eritematoso sistémico en comparación con sujetos sanos (SRP062966) (Hung et al., 2015) (**Tabla 3**). Sin embargo, un estudio comparativo entre el transcriptoma ileal en la enfermedad de Crohn pediátrica (EC) con y sin ulceración profunda y con o sin afectación ileal (SRP042228) (Haberman et al., 2014), reveló un aumento de la expresión para I_μ , I_δ y $I\alpha 2$ en pacientes con EC con ulcera profunda, independientemente de la presencia de inflamación macroscópica o microscópica (**Tabla 3**) (**Figura 21**). Se ha descrito que en pacientes con EC que presentan ulcera profunda hay un aumento en la expresión de genes asociados a la señalización de citocinas proinflamatorias y relacionados con la activación de linfocitos en células del sistema inmune incluidos

los linfocitos B (Haberman et al., 2014), por lo que el aumento de los I_H podría estar relacionado con el contexto celular y de expresión específico para esta patología donde se presenta un estado inflamatorio severo.

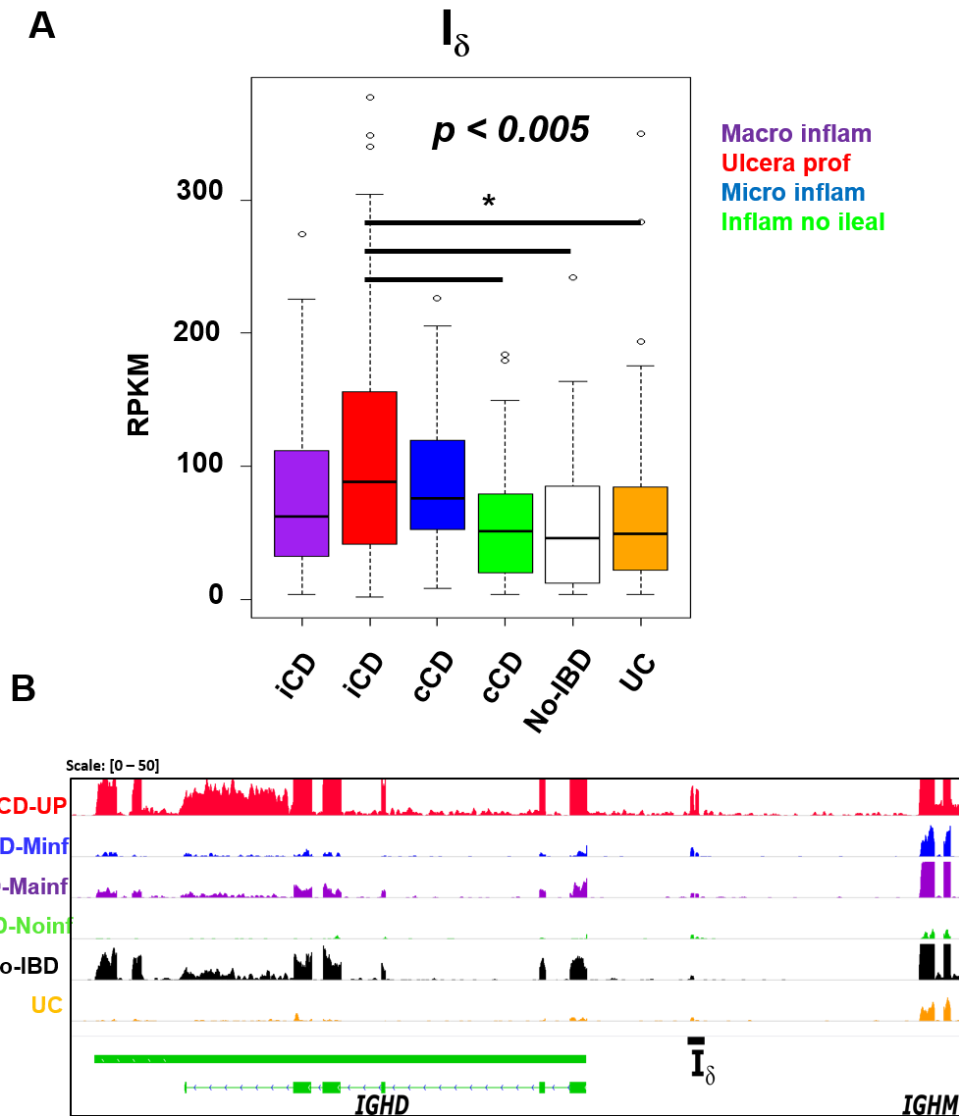


Figura 21. Aumento de la transcripción de I_δ en la mucosa ileal de la enfermedad de Crohn. Los pacientes con enfermedad de Crohn pediátrica del proyecto SRP042228 se clasificaron en dos grupos según la participación de la mucosa ileal (I) o la no participación (CCD) (Haberman et al., 2014). Los pacientes se clasificaron según el grado de inflamación de la mucosa ileal. La mucosa ileal de la enfermedad intestinal no inflamatoria (nIBD) y la colitis ulcerosa (CU) se utilizaron como controles (Haberman et al., 2014). (A) Diagrama de caja que muestra un aumento de la transcripción de I_δ (RPKM, eje y) en la enfermedad de Crohn ulcerada profunda con afectación ileal (un factor ANOVA, Tukey comparaciones múltiples de medias. $P < 0,005$). (B) Paisaje transcripcional del locus IGHD-IGHM en el

cromosoma 14 humano, que muestra un aumento de la transcripción I en muestras representativas de CD con afectación ileal ulcerada profunda (iCD, pista roja), CD con inflamación microscópica (cCD, pista azul), iCD con macroscópico Inflamación (pista púrpura), enfermedad intestinal no inflamatoria (EII, pista negra) y colitis ulcerosa (CU, pista naranja).

Hallazgos principales

- *En enfermedades infecciosas como diarrea, brucelosis y malaria observamos una disminución en la expresión de transcritos RCCnc*
 - *En un proyecto de vacunación contra influenza observamos en el análisis de expresión diferencial, que al igual que en nuestro experimento no hay diferencias post vacunación en los RCCnc ni tampoco en los transcritos codificantes. Sin embargo, se observa un incremento en Ig1, Ia1 al igual que en RCC codificantes (IgM, IgA1, IgG2 e IgG4) durante la infección aguda con H7N9.*
 - *En enfermedad de Crohn identificamos un incremento en la expresión de $I\mu$, $I\delta$ e $I\alpha 2$ en pacientes que presentaban ulceración profunda lo que refleja el microambiente con inflamación aguda.*
-

Discusión

En este trabajo realizamos un análisis integral y sistemático de la transcripción RCCnc en humanos utilizando un enfoque experimental y de minería de datos públicos de un conjunto grande y diverso de datos de RNA-seq con la herramienta *recount2* (Collado-Torres et al., 2017). Aunque se han cuantificado los transcritos RCCnc previamente, los límites de los exones I_H para cada clase no han sido definidos y no están anotados en la versión actual del genoma humano (GRCh38.p12).

En las primeras observaciones de este estudio, pudimos constatar que la transcripción RCCnc se restringe a linfocitos B en el adulto (**Figura 7A**), mientras que $I\gamma_4$ se expresa en diversos tejidos linfoides y no linfoides en la etapa fetal, lo cual requiere de un análisis más profundo para poder entender el papel de $I\gamma_4$ durante el desarrollo fetal (**Figura 18B**). Además, observamos que la transcripción RCCnc, así como la transcripción codificante está presente en células progenitoras de linfocitos B principalmente $C\mu$ e $I\mu$, aunque también se observa el incremento de la presencia de $I\delta$ a medida que se da la diferenciación a linfocitos B (**Figura 18A**). En un análisis general sobre la expresión de los transcritos RCCnc clasificada de acuerdo a los tres grandes tipos de iniciativas englobadas en proyectos (GTEx, TCGA o SRA), observamos que $I\mu$ tiene los niveles más altos de transcripción y es el más expresado en los tres proyectos. Siguió una fuerte disminución en la transcripción de las transcripciones de I_H restantes. Inesperadamente, la transcripción de $I\delta$ y $I\epsilon$ fue relativamente alta, aunque su transcripción se restringió a una pequeña proporción de muestras. En contraste, los niveles de transcripción de $I\gamma_3$ e $I\gamma_2$ fueron los más bajos en todos los conjuntos de datos y se transcribieron en una proporción relativamente alta de muestras. Estos datos indican que la expresión de $I\mu$ es mayor y más generalizada (es decir, expresada en muchas muestras y proyectos de diversos orígenes), mientras que otros niveles de transcripción de I_H fueron generalmente más bajos y su transcripción es más restringida (**Figura 13**). Se identificó un nuevo elemento transcrito ($I\delta$) ubicado en la región intergénica *IGHM-IGHD*, que se superpone con una región repetida descrita anteriormente $\Sigma\mu$, la cual se encuentra río abajo de una región de *switch* atípica ($\sigma\delta$) involucrada en RCC no clásica a IgD en ratones (Rouaud et al., 2014) y humanos (Arpin et al., 1998; Chen et al., 2009; Kluin et al., 1995). La región $\Sigma\mu$ se describió originalmente como mediador de RCC μ - δ por recombinación homóloga en líneas celulares de mieloma (Akahori et al., 1988; Yasui et al., 1989). Sin embargo, el papel biológico de $\Sigma\mu$ es incierto, porque las líneas celulares transformadas IgD⁺ EBV y las células B tonsilares humanas se someten a μ - δ RCC mediante recombinación no homóloga usando $\sigma\delta$ como aceptor de la

región de *switch* (Arpin et al., 1998; Kluin et al., 1995), la cual también es dependiente de AID (Chen et al., 2009). Aquí, demostramos que $I\delta$ ($\Sigma\mu$) puede transcribirse activamente (**Figura 7B**). No obstante, $I\delta$ como I_H acceptor no encaja en el modelo general de RCC, ya que se encuentra río abajo de la región $\sigma\delta$. Se requiere más investigación para dilucidar si la transcripción $I\delta$ está implicada en RCC no clásica de $\mu-\delta$.

El presente estudio tiene ciertas limitaciones: 1) Dado que la mayoría de los estudios de RNA-seq públicamente disponibles están dirigidos a el análisis de la transcripción de genes codificantes, es posible que por razones técnicas se haya subestimado la expresión absoluta del RCCnc, ya que solo una fracción de estos se encuentra poliadenilada (Quinn and Chang, 2016). 2) la transcripción de RCCnc es obligatoria y precede a la RCC, sin embargo, no es condición *sine qua non* de RCC. Entre los factores más determinantes que influyen en la cantidad de transcripción de RCCnc observada en una muestra de RNA-seq está la cantidad relativa de células B que expresan un I_H particular, como lo sugiere la alta correlación del valor Z entre RCCnc y su correspondiente transcripción codificante (No mostrado). Proponemos que una mayor proporción del valor Z codificante/no codificante indica una mayor proporción de células B sometidas a un evento de RCC particular. 3) Otros factores que influyen en la cantidad de transcripción RCCnc es que los círculos de *switch* son transcripcionalmente activos (Kinoshita et al., 2001), y nuestro análisis actual no puede diferenciar si la transcripción RCCnc deriva del cromosoma o del círculo de *switch*. Los círculos de *switch*, son episomas no replicantes que se forman como resultado de la escisión del fragmento de ADN que se encuentra entre la región de *switch* donadora hasta una parte de la región de *switch* acceptora, el cual se diluyen con la división de células B (Iwasato et al., 1990; Kinoshita et al., 2001; Matsuoka et al., 1990). Por lo tanto, las células B sometidas a división celular a alta velocidad (es decir, células B de CG) deberían diluir la cantidad de templado de los círculos en mayor medida que las células B que no se dividen. Sin embargo, nosotros no llevamos a cabo esta medición.

Una motivación importante para este trabajo fue poder identificar firmas transcripcionales tempranas en la sangre que se correlacionen con la intensidad y la calidad de la respuesta humoral, y en particular con la respuesta de CG. Por lo que analizamos la transcripción de I_{μ} , $I_{\gamma 1}$, $I_{\gamma 3}$ y AID en muestras de individuos vacunados. La vacunación contra hepatitis B y/o Tétanos/Difteria indujo la transcripción de I_{μ} e $I_{\gamma 3}$ en sangre periférica el día 14 después de la vacunación, no coincidiendo con el pico de plasmablastos (día 7 después de la vacunación), lo que sugiere que la transcripción de RCCnc puede no ser el resultado de la movilización de plasmablastos a sangre periférica, como originalmente teníamos hipotetizado (**Figura 12**). Por otra parte, se observó un aumento de la transcripción $I_{\gamma 1}$ e $I_{\alpha 1}$ en la infección natural H7N9, pero no en la vacunación contra influenza, o en la infección por rotavirus (**Tabla 3**). La respuesta diferencial observada en la vacunación contra la influenza y la infección natural por rotavirus, en contraste con la vacuna contra el VHB / tétanos-difteria y la infección H7N9, puede ser el resultado de la exposición frecuente de la población humana a la influenza estacional y/o rotavirus, que reactivaría células Bm IgG^{+} en ausencia de RCC (Guthmiller and Wilson, 2018). Aunque la transcripción de I_H es altamente inducible tras la activación de células B (**Figura 11D-G**), la minería de datos en GTE_x y SRA reveló que la transcripción I_{μ} es más alta que otras I_H , siendo consistente con el modelo actual de RCC, en el que I_{μ} se transcribe constitutivamente bajo el control del *enhancer* intrónico de μ (E_{μ}), y su transcripción es necesaria para la RCC, independientemente de la clase aceptora (Li et al., 1994). E_{μ} participa en la remodelación de la cromatina del locus IGH durante el compromiso del linaje y la recombinación VDJ (Afshar et al., 2006; Bolland et al., 2007), y su eliminación altera el desarrollo de las células B, la transcripción I_{μ} y la RCC (Perlot et al., 2005).

La combinación del valor Z y el análisis por clústeres permitió la identificación de patrones de transcripción RCCnc distintivos relacionados con funciones inmunológicas conocidas, como la coexpresión de I_{μ} e $I_{\alpha 2}$ en órganos ricos en MALT, donde tiene lugar la secreción activa de IgA. De particular interés es el grupo Z1, definido por una firma transcripcional $I_{\gamma 1}$, $I_{\gamma 2}$ e $I_{\gamma 4}$. Este grupo, fue el único con

mayor expresión de $I\gamma 4$, y fue característico de los testículos, la glándula tiroides y el tejido adiposo visceral (omento), pero no del tejido adiposo subcutáneo. Las células B linfoides adiposas viscerales son células maduras e inmunológicamente activas, implicadas en la homeostasis del tejido adiposo que pueden desempeñar importantes funciones proinflamatorias asociadas con el síndrome metabólico y la obesidad (Winer et al., 2014). La expresión de la firma transcripcional $I\gamma 1$ e $I\gamma 4$ en el testículo y la tiroides es inesperada, ya que se consideran sitios “inmunoprivilegiados” carentes de órganos linfoides secundarios / terciarios (Chen et al., 2016) (**Figura 17**). Desconocemos si la RCC a IgG_4 está teniendo lugar en estos órganos, y se requiere más investigación, ya que IgG_4 se considera como una Ig atípica que carece de funciones efectoras mediadas por Fc (Aalberse et al., 2009). Además, la transcripción $I\gamma 4$ en diferentes tejidos fetales sugiere que su transcripción no se limita al tejido linfoides, y puede ser una característica común de la multipotencialidad, ya que no está acoplada a la transcripción codificante, sugiriendo funciones adicionales a la RCC (**Figura 18B**).

Un objetivo principal del proyecto GTEx fue identificar el papel de la variación genética en la expresión génica como un rasgo cuantitativo. El enriquecimiento en tejidos en dos grupos perfil transcripcional RCCnc cualitativamente diferente, como en hígado (Z3 y Z6), testículo (Z1 y Z2), íleon terminal (Z5 y Z9) y sangre completa (Z5, Z7 y Z9) indican heterogeneidad transcripcional de RCCnc entre donadores de tejidos involucrados en el proyecto GTEx (**Figura 17**). Esta heterogeneidad transcripcional podría resultar de diferentes microambientes tisulares (citocinas, microbiota y estímulos ambientales), así como de variación genética, que pueden modificar los patrones de transcripción RCCnc y posiblemente la propia RCC. Recientemente se identificaron cuatro SNPs en el locus *IGH* en humanos, presumiblemente involucrados en la RCC que afectan los niveles de inmunoglobulinas (Jonsson et al., 2017).

El estudio de la transcripción de RCCnc en el cáncer es de particular interés por varias razones: 1) La respuesta antitumoral está mediada en gran medida por la presentación de neoantígenos tumorales a las células T y al equilibrio de linfocitos

T reguladores. Sin embargo, la actividad antitumoral mediada por anticuerpos puede lograrse mediante citotoxicidad dependiente de anticuerpos u otros mecanismos (Colbeck et al., 2017). 2) La presencia de estructuras linfoides ectópicas (o terciarias) (ELE) (Colbeck et al., 2017; Pitzalis et al., 2014) y mayor densidad de células B infiltrantes y células T auxiliares foliculares se correlaciona con una mejor supervivencia en cáncer de pulmón (Germain et al., 2014), mama (Gu-Trantien et al., 2013) y carcinoma colorrectal (Bindea et al., 2013). 3) El microambiente tumoral, que incluye ciertas citocinas, puede modificar los patrones de RCC en las células B infiltrantes y ELE, independientemente de la función efectora del anticuerpo. 4) La actividad de AID contribuye a la mutagénesis fuera del objetivo (*off-target*) y la inestabilidad genómica en las células B neoplásicas (Qian et al., 2014). La transcripción aberrante de AID y RCCnc en células tumorales no linfoides podría contribuir a la kataegis (zonas de hipermutación regional identificadas dentro de algunos tipos de cáncer) mediada por desaminación de citidinas (Taylor et al., 2013).

Hemos encontrado que la expresión promedio de I_H en ciertos tumores analizados en el proyecto TCGA se asemeja a su tejido homólogo no neoplásico, lo cual puede estar relacionado con la cantidad de tejido linfoide asociado, así como tejido linfoide ectópico (Colbeck et al., 2017) (**Figura 19**). Sin embargo, un análisis detallado de los tipos de cáncer individuales reveló que, en aproximadamente un tercio de los tipos de cáncer analizados, incluidos los carcinomas de tiroides, próstata, colon e hígado, la transcripción de I_H promedio es menor que en su contraparte saludable (**Figura 20B-E**). En otro tercio que incluye cáncer de mama, páncreas, ovario y melanoma cutáneo, la transcripción promedio de RCCnc es mayor que en el tejido sano correspondiente (**Figura 20G, I, K y L**). Y finalmente en un tercio no se observaron diferencias en los transcritos RCCnc entre tejido normal y su contraparte tumoral, como cáncer escamoso de pulmón, cáncer cervical y cáncer de estómago (**Figura 21**). La mayor parte de la evidencia que hemos recopilado hasta ahora indica que, el origen de la transcripción de I_H está en las células B infiltrantes y B ELS, en lugar de las células tumorales *per se*. Por lo tanto, las diferencias en la transcripción de I_H en el cáncer pueden ser el resultado de la

edición inmune (Teng et al., 2015), que puede alterar la cantidad y el estado de activación de las células B infiltrantes y B de ELE.

Con base al agrupamiento de firmas de expresión génica, todos los cánceres no hematológicos del proyecto TCGA se clasificaron en seis subtipos inmunológicamente distintos con patrón de aberraciones somáticas distintivas, microambiente tumoral que incluye la cantidad de infiltración de tipo celular y el resultado clínico (Saltz et al., 2018; Thorsson et al., 2018). Actualmente estamos investigando la relación entre estos seis subtipos inmunológicos, el grado de infiltración linfoide y el resultado clínico, con el patrón de transcripción RCCnc, lo cual podría ayudar a comprender el papel elusivo de las células B infiltrantes en la progresión de diferentes tipos de cáncer (Colbeck et al., 2017).

Uno de los descubrimientos que más llama la atención de este estudio, fue el de los patrones transcripcionales de I_H's para la enfermedad de Crohn donde pudimos identificar una firma I δ -I α 2 en la mucosa ileal, donde se ha identificado la presencia de células B en respuesta a la inflamación (Haberman et al., 2014). Sin embargo, el aumento de la transcripción de I δ , particularmente en enfermedad de Crohn asociada a ulceración profunda es un hallazgo intrigante (**Figura 22**), ya que se ha observado que los linfocitos IgD⁺ que han pasado por cambio de clase, se diferencian a células plasmáticas secretoras de IgD que entran en circulación y se esparcen hacia glándulas salivales y lagrimales y mucosas nasales pero no hay diseminación hacia intestino (Brandtzaeg, 2005). Sin embargo, en ratones incapaces de llevar a cabo RCC clásica debido a la pérdida de función de *53BP1*, se observó que tienen una elevación de los títulos séricos de IgD dependiente de la microbiota intestinal y un aumento de las células IgD⁺ B (Choi et al., 2017). Los niveles séricos de IgD son elevados en pacientes con enfermedad de Crohn (Cianci et al., 2016) y otros síndromes autoinflamatorios (Chen et al., 2009; Masters et al., 2009). Una gran proporción de células IgD⁺ mediante RCC μ - δ presentan especificidades autorreactivas y poli-reactivas (Koelsch et al., 2007; Zheng et al., 2004), una característica compartida con "autoanticuerpos naturales", que son reactivas contra componentes catiónicos de la pared bacteriana y pueden

proporcionar inmunidad natural contra la infección (Mannoor et al., 2013). En la mucosa respiratoria humana, las células IgD⁺ B conmutadas median la interacción entre la inmunidad innata, la adaptativa y la inflamación (Chen et al., 2009). Este trabajo representa un punto de partida para dilucidar el papel de la transcripción de I δ , su papel en μ - δ RCC y sus implicaciones en mucosas sanas e inflamadas.

A pesar de la limitación de depender de la información sobre datos públicos cuando a menudo el investigador remitente elige reportar el mínimo de metadatos, el SRA representa una enorme fuente de datos de RNA-seq de estudios muy diversos. En contraste con los criterios metodológicos estandarizados y los protocolos de recopilación de metadatos utilizados por los consorcios GTEx y TCGA, es de esperar una mayor variabilidad metodológica en los datos de SRA, lo que limita las comparaciones entre los estudios.

Conclusiones

Hemos demostrado que la transcripción de RCCnc está ampliamente distribuida en tejidos humanos sanos. La abundancia de la transcripción de RCCnc depende en gran medida de la presencia de tejidos linfoides secundarios como MALT. La transcripción de I_H, principalmente I μ , es activa en células madre hematopoyéticas pluripotentes, linfoides tempranos y precursores de células B y T comprometidas. I γ 4 tiene una expresión única y altamente expresada en varios tejidos fetales, destacando un papel potencial de la transcripción RCCnc más allá de la RCC.

En cáncer, la transcripción de RCCnc depende principalmente del grado de infiltración de células B en tejidos y posiblemente de la presencia de tejido linfóide ectópico (terciario), y en la mayoría de los tipos de cáncer es diferente a su homólogo sano. La transcripción de RCCnc en sangre periférica se modifica con la vacuna contra la hepatitis B / tétanos-difteria, pero no con la vacunación contra la influenza. Nuestro análisis experimental y de minería de datos indica que la transcripción RCCnc en sangre no deriva de los plasmablastos movilizados.

La expresión de I μ , I δ e I α 2 observada en la enfermedad de Crohn sugiere un microambiente inflamatorio particular y puede estar implicado en la RCC no

clásica μ - δ . En general, este trabajo ha generado nuevas hipótesis susceptibles a comprobación directa y es un gran ejemplo de cómo los datos públicos de *recount2* se pueden utilizar para ampliar nuestra comprensión de la transcripción, incluidas las regiones fuera del transcriptoma conocido, así como del uso innovador de “big data” en salud para el fomento de la investigación en medicina traslacional.

Referencias

- Aalberse, R.C., Stapel, S.O., Schuurman, J., and Rispens, T. (2009). Immunoglobulin G4: an odd antibody. *Clin Exp Allergy* 39, 469-477.
- Afshar, R., Pierce, S., Bolland, D.J., Corcoran, A., and Oltz, E.M. (2006). Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *Journal of immunology* 176, 2439-2447.
- Akahori, Y., Handa, H., Imai, K., Abe, M., Kameyama, K., Hibiya, M., Yasui, H., Okamura, K., Naito, M., Matsuoka, H., *et al.* (1988). Sigma region located between C mu and C delta genes of human immunoglobulin heavy chain: possible involvement of tRNA-like structure in RNA splicing. *Nucleic acids research* 16, 9497-9511.
- Allman, D., and Pillai, S. (2008). Peripheral B cell subsets. *Current opinion in immunology* 20, 149-157.
- Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* 152, 417-429.
- Amarasekera, M. (2011). Immunoglobulin E in health and disease. *Asia Pac Allergy* 1, 12-15.
- Arpin, C., de Bouteiller, O., Razanajaona, D., Fugier-Vivier, I., Briere, F., Banchereau, J., Lebecque, S., and Liu, Y.J. (1998). The normal counterpart of IgD myeloma cells in germinal center displays extensively mutated IgVH gene, Cmu-Cdelta switch, and lambda light chain expression. *The Journal of experimental medicine* 187, 1169-1178.
- Baumgarth, N. (2004). B-Cell Immunophenotyping. In *Methods in Cell Biology* (Academic Press), pp. 643-662.
- Beguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K.L., Rosen, M., Shen, H., Yang, S.N., Wang, L., Ezponda, T., *et al.* (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer cell* 23, 677-692.
- Bernstein, M.N., Doan, A., and Dewey, C.N. (2017). MetaSRA: normalized human sample-specific metadata for the Sequence Read Archive. *Bioinformatics* 33, 2914-2923.
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., *et al.* (2013). Spatiotemporal dynamics of

intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782-795.

Black, S., De Gregorio, E., and Rappuoli, R. (2015). Developing vaccines for an aging population. *Science translational medicine* 7, 281ps288.

Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007). Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. *Mol Cell Biol* 27, 5523-5533.

Brandtzaeg, P., Carlsen, H. S. and Farstad, I. N. (2005). The human mucosal B-cell system. In *Mucosal immunology*, J. Mestecky, Lamm, M. E., Strober, W., Bienenstock, J., McGhee, J. R., Mayer, L., ed. (Elsevier).

Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* 45, 1113-1120.

Casero, D., Sandoval, S., Seet, C.S., Scholes, J., Zhu, Y., Ha, V.L., Luong, A., Parekh, C., and Crooks, G.M. (2015). Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nature immunology* 16, 1282-1291.

Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature reviews Immunology* 4, 541-552.

Chen, K., and Cerutti, A. (2011). The function and regulation of immunoglobulin D. *Current opinion in immunology* 23, 345-352.

Chen, K., Xu, W., Wilson, M., He, B., Miller, N.W., Bengten, E., Edholm, E.S., Santini, P.A., Rath, P., Chiu, A., *et al.* (2009). Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils. *Nature immunology* 10, 889-898.

Chen, Q., Deng, T., and Han, D. (2016). Testicular immunoregulation and spermatogenesis. *Seminars in cell & developmental biology* 59, 157-165.

Choi, J.H., Wang, K.W., Zhang, D., Zhan, X., Wang, T., Bu, C.H., Behrendt, C.L., Zeng, M., Wang, Y., Misawa, T., *et al.* (2017). IgD class switching is initiated by microbiota and limited to mucosa-associated lymphoid tissue in mice. *Proceedings of the National Academy of Sciences of the United States of America* 114, E1196-E1204.

Chowdhury, M., Forouhi, O., Dayal, S., McCloskey, N., Gould, H.J., Felsenfeld, G., and Fear, D.J. (2008). Analysis of intergenic transcription and histone modification across the human immunoglobulin heavy-chain locus. *Proceedings of the National Academy of Sciences of the United States of America* 105, 15872-15877.

Cianci, R., Lolli, S., Pagliari, D., Gambassi, G., Frosali, S., Marmo, R., Melioli, G., Orlando, A., Newton, E.E., Serone, E., *et al.* (2016). The involvement of IgH enhancer HS1.2 in the pathogenesis of Crohn's disease: how the immune system can influence a multifactorial disease. *Eur Rev Med Pharmacol Sci* 20, 3618-3627.

- Clem, A.S. (2011). Fundamentals of vaccine immunology. *J Glob Infect Dis* 3, 73-78.
- Colbeck, E.J., Ager, A., Gallimore, A., and Jones, G.W. (2017). Tertiary Lymphoid Structures in Cancer: Drivers of Antitumor Immunity, Immunosuppression, or Bystander Sentinels in Disease? *Frontiers in immunology* 8, 1830.
- Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E., Langmead, B., and Leek, J.T. (2017). Reproducible RNA-seq analysis using recount2. *Nature biotechnology* 35, 319-321.
- Consortium, G. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204-213.
- Consortium, G.T., Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., *et al.* (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204.
- Consortium, T.G. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature genetics* 45, 580-585.
- de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453-1454.
- Defrance, T., Taillardet, M., and Genestier, L. (2011). T cell-independent B cell memory. *Current opinion in immunology* 23, 330-336.
- Dorshkind, K., and Rawlings, D.J. (2018). Chapter 20 - B-Cell Development. In *Hematology (Seventh Edition)*, R. Hoffman, E.J. Benz, L.E. Silberstein, H.E. Heslop, J.I. Weitz, J. Anastasi, M.E. Salama, and S.A. Abutalib, eds. (Elsevier), pp. 210-220.
- Dudley, D.D., Chaudhuri, J., Bassing, C.H., and Alt, F.W. (2005). Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Advances in immunology* 86, 43-112.
- Fear, D.J., McCloskey, N., O'Connor, B., Felsenfeld, G., and Gould, H.J. (2004). Transcription of Ig Germline Genes in Single Human B Cells and the Role of Cytokines in Isotype Determination. *The Journal of Immunology* 173, 4529-4538.
- Frolich, D., Giesecke, C., Mei, H.E., Reiter, K., Daridon, C., Lipsky, P.E., and Dorner, T. (2010). Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells. *Journal of immunology* 185, 3103-3110.
- Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., *et al.* (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature genetics* 50, 956-967.
- Germain, C., Gnjjatic, S., Tamzalit, F., Knockaert, S., Remark, R., Goc, J., Lepelley, A., Becht, E., Katsahian, S., Bizouard, G., *et al.* (2014). Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *American journal of respiratory and critical care medicine* 189, 832-844.

Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., de Wind, A., Ravoet, M., Le Buanec, H., Sibille, C., Manfouo-Foutsop, G., *et al.* (2013). CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *The Journal of clinical investigation* 123, 2873-2892.

Guthmiller, J.J., and Wilson, P.C. (2018). Harnessing immune history to combat influenza viruses. *Current opinion in immunology* 53, 187-195.

Haberman, Y., Tickle, T.L., Dexheimer, P.J., Kim, M.O., Tang, D., Karns, R., Baldassano, R.N., Noe, J.D., Rosh, J., Markowitz, J., *et al.* (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *The Journal of clinical investigation* 124, 3617-3633.

Henn, A.D., Wu, S., Qiu, X., Ruda, M., Stover, M., Yang, H., Liu, Z., Welle, S.L., Holden-Wiltse, J., Wu, H., *et al.* (2013). High-resolution temporal response patterns to influenza vaccine reveal a distinct human plasma cell gene signature. *Scientific reports* 3, 2327.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., *et al.* (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173, 291-304 e296.

Hoek, K.L., Samir, P., Howard, L.M., Niu, X., Prasad, N., Galassie, A., Liu, Q., Allos, T.M., Floyd, K.A., Guo, Y., *et al.* (2015). A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PloS one* 10, e0118528.

Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W., *et al.* (2015). The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* 350, 455-459.

Iwasato, T., Shimizu, A., Honjo, T., and Yamagishi, H. (1990). Circular DNA is excised by immunoglobulin class switch recombination. *Cell* 62, 143-149.

Jonsson, S., Sveinbjornsson, G., de Lapuente Portilla, A.L., Swaminathan, B., Plomp, R., Dekkers, G., Ajore, R., Ali, M., Bentlage, A.E.H., Elmer, E., *et al.* (2017). Identification of sequence variants influencing immunoglobulin levels. *Nature genetics* 49, 1182-1191.

Kenter, A.L. (2005). Class switch recombination: an emerging mechanism. *Curr Top Microbiol Immunol* 290, 171-199.

Kibbe, W.A., Arze, C., Felix, V., Mittraka, E., Bolton, E., Fu, G., Mungall, C.J., Binder, J.X., Malone, J., Vasant, D., *et al.* (2015). Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research* 43, D1071-1078.

Kinoshita, K., Harigai, M., Fagarasan, S., Muramatsu, M., and Honjo, T. (2001). A hallmark of active class switch recombination: transcripts directed by I promoters on looped-out circular DNAs. *Proceedings of the National Academy of Sciences of the United States of America* 98, 12620-12623.

Kluin, P.M., Kayano, H., Zani, V.J., Kluin-Nelemans, H.C., Tucker, P.W., Satterwhite, E., and Dyer, M.J. (1995). IgD class switching: identification of a novel recombination site in neoplastic and normal B cells. *European journal of immunology* 25, 3504-3508.

Koelsch, K., Zheng, N.Y., Zhang, Q., Duty, A., Helms, C., Mathias, M.D., Jared, M., Smith, K., Capra, J.D., and Wilson, P.C. (2007). Mature B cells class switched to IgD are autoreactive in healthy individuals. *The Journal of clinical investigation* 117, 1558-1565.

Koues, O.I., Kowalewski, R.A., Chang, L.W., Pyfrom, S.C., Schmidt, J.A., Luo, H., Sandoval, L.E., Hughes, T.B., Bednarski, J.J., Cashen, A.F., *et al.* (2015). Enhancer sequence variants and transcription-factor deregulation synergize to construct pathogenic regulatory circuits in B-cell lymphoma. *Immunity* 42, 186-198.

Kushwaha, G., Dozmorov, M., Wren, J.D., Qiu, J., Shi, H., and Xu, D. (2016). Hypomethylation coordinates antagonistically with hypermethylation in cancer development: a case study of leukemia. *Human genomics* 10 Suppl 2, 18.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* 15, R29.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS computational biology* 9, e1003118.

Li, S.C., Rothman, P.B., Zhang, J., Chan, C., Hirsh, D., and Alt, F.W. (1994). Expression of I mu-C gamma hybrid germline transcripts subsequent to immunoglobulin heavy chain class switching. *International immunology* 6, 491-497.

Li, W., Wang, S., and Lu, S. (2014). Pilot Study on the Use of DNA Priming Immunization to Enhance *Y. pestis* LcrV-Specific B Cell Responses Elicited by a Recombinant LcrV Protein Vaccine. *Vaccines* 2.

Linsley, P.S., Speake, C., Whalen, E., and Chaussabel, D. (2014). Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PloS one* 9, e109760.

Litinskiy, M.B., Nardelli, B., Hilbert, D.M., He, B., Schaffer, A., Casali, P., and Cerutti, A. (2002). DCs induce CD40-independent immunoglobulin class switching through BLYS and APRIL. *Nature immunology* 3, 822-829.

Lorenz, M., Jung, S., and Radbruch, A. (1995). Switch transcripts in immunoglobulin class switching. *Science* 267, 1825-1828.

Mak, T.W., Saunders, Mary E., Jett, Bradley D. (2014). B Cell Development, Activation and Effector Functions. In *Primer to the Immune Response* (Boston: Academic Cell), pp. 111-142.

Mannoor, K., Xu, Y., and Chen, C. (2013). Natural autoantibodies and associated B cells in immunity and autoimmunity. *Autoimmunity* 46, 138-147.

- Martinez-Barnetche, J. (2016). Linfocitos B. In *Inmunología molecular, celular y traslacional*, M.d.C.J. Lenin Pavon Ph. D, María Eugenia Garcés, ed. (Wolters Kluwer).
- Masters, S.L., Simon, A., Aksentijevich, I., and Kastner, D.L. (2009). Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (*). *Annual review of immunology* 27, 621-668.
- Matsuoka, M., Yoshida, K., Maeda, T., Usuda, S., and Sakano, H. (1990). Switch circular DNA formed in cytokine-treated mouse splenocytes: evidence for intramolecular DNA deletion in immunoglobulin class switching. *Cell* 62, 135-142.
- Mei, B., Ding, X., Xu, H.Z., and Wang, M.T. (2014). Global gene expression changes in human peripheral blood after H7N9 infection. *Gene* 551, 255-260.
- Melchers, F. (2015). Checkpoints that control B cell development. *The Journal of clinical investigation* 125, 2203-2210.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553-563.
- Murphy, K., Travers, P., Walport, M., and Janeway, C. (2012). *Janeway's immunobiology* (New York: Garland Science).
- Obukhanych, T.V., and Nussenzweig, M.C. (2006). T-independent type II immune responses generate memory B cells. *The Journal of experimental medicine* 203, 305.
- Odendahl, M., Mei, H., Hoyer, B.F., Jacobi, A.M., Hansen, A., Muehlinghaus, G., Berek, C., Hiepe, F., Manz, R., Radbruch, A., *et al.* (2005). Generation of migratory antigen-specific plasma blasts and mobilization of resident plasma cells in a secondary immune response. *Blood* 105, 1614-1621.
- Pascual, M., Roa, S., Garcia-Sanchez, A., Sanz, C., Hernandez-Hernandez, L., Grealley, J.M., Lorente, F., Davila, I., and Isidoro-Garcia, M. (2014). Genome-wide expression profiling of B lymphocytes reveals IL4R increase in allergic asthma. *The Journal of allergy and clinical immunology* 134, 972-975.
- Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proceedings of the National Academy of Sciences of the United States of America* 102, 14362-14367.
- Pitzalis, C., Jones, G.W., Bombardieri, M., and Jones, S.A. (2014). Ectopic lymphoid-like structures in infection, cancer and autoimmunity. *Nature reviews Immunology* 14, 447-462.
- Plaut, M., Sawyer, R.T., and Fenton, M.J. (2009). Summary of the 2008 National Institute of Allergy and Infectious Diseases–US Food and Drug Administration Workshop on Food Allergy Clinical Trial Design. *Journal of Allergy and Clinical Immunology* 124, 671-678.e671.
- Pulendran, B., and Ahmed, R. (2011). Immunological mechanisms of vaccination. *Nature immunology* 12, 509-517.

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathe, E., Benner, C., *et al.* (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159, 1524-1537.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 17, 47-62.

Rappuoli, R., Pizza, M., Del Giudice, G., and De Gregorio, E. (2014). Vaccines, new opportunities for a new society. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12288-12293.

Roost, M.S., van Iperen, L., Ariyurek, Y., Buermans, H.P., Arindrarto, W., Devalla, H.D., Passier, R., Mummery, C.L., Carlotti, F., de Koning, E.J., *et al.* (2015). KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. *Stem Cell Reports* 4, 1112-1124.

Rouaud, P., Saintamand, A., Saad, F., Carrion, C., Lecardeur, S., Cogne, M., and Denizot, Y. (2014). Elucidation of the enigmatic IgD class-switch recombination via germline deletion of the IgH 3' regulatory region. *The Journal of experimental medicine* 211, 975-985.

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., *et al.* (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell reports* 23, 181-193 e187.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nature reviews Immunology* 11, 251-263.

Schroeder, H.W., Jr., and Cavacini, L. (2010). Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology* 125, S41-S52.

Soon, W.W., Hariharan, M., and Snyder, M.P. (2013). High-throughput sequencing for biology and medicine. *Mol Syst Biol* 9, 640.

Stephens, Z.D., Lee, S.Y., Faghri, F., Campbell, R.H., Zhai, C., Efron, M.J., Iyer, R., Schatz, M.C., Sinha, S., and Robinson, G.E. (2015). Big Data: Astronomical or Genomical? *PLoS Biol* 13, e1002195.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* 2, e00534.

Teng, M.W., Galon, J., Fridman, W.H., and Smyth, M.J. (2015). From mice to humans: developments in cancer immunoediting. *The Journal of clinical investigation* 125, 3338-3346.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., *et al.* (2018). The Immune Landscape of Cancer. *Immunity* 48, 812-830 e814.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178-192.

Tran, T.M., Jones, M.B., Ongoiba, A., Bijker, E.M., Schats, R., Venepally, P., Skinner, J., Doumbo, S., Quinten, E., Visser, L.G., *et al.* (2016). Transcriptomic evidence for modulation of host inflammatory responses during febrile *Plasmodium falciparum* malaria. *Scientific reports* 6, 31291.

Victoria, G.D., and Nussenzweig, M.C. (2012). Germinal centers. *Annual review of immunology* 30, 429-457.

Vidarsson, G., Dekkers, G., and Rispens, T. (2014). IgG subclasses and allotypes: from structure to effector functions. *Frontiers in immunology* 5, 520-520.

Winer, D.A., Winer, S., Chng, M.H., Shen, L., and Engleman, E.G. (2014). B Lymphocytes in obesity-related adipose tissue inflammation and insulin resistance. *Cellular and molecular life sciences : CMLS* 71, 1033-1043.

Wrammert, J., Onlamoon, N., Akondy, R.S., Perng, G.C., Polsrila, K., Chandele, A., Kwissa, M., Pulendran, B., Wilson, P.C., Wittawatmongkol, O., *et al.* (2012). Rapid and massive virus-specific plasmablast responses during acute dengue virus infection in humans. *Journal of virology* 86, 2911-2918.

Wrammert, J., Smith, K., Miller, J., Langley, W.A., Kokko, K., Larsen, C., Zheng, N.Y., Mays, I., Garman, L., Helms, C., *et al.* (2008). Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 453, 667-671.

Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nature reviews Immunology* 12, 517-531.

Yasui, H., Akahori, Y., Hirano, M., Yamada, K., and Kurosawa, Y. (1989). Class switch from mu to delta is mediated by homologous recombination between sigma mu and sigma mu sequences in human immunoglobulin gene loci. *European journal of immunology* 19, 1399-1403.

Yewdell, W.T., and Chaudhuri, J. (2017). A transcriptional serenAID: the role of noncoding RNAs in class switch recombination. *International immunology* 29, 183-196.

Yoshida, T., Mei, H., Dorner, T., Hiepe, F., Radbruch, A., Fillatreau, S., and Hoyer, B.F. (2010). Memory B and memory plasma cells. *Immunological reviews* 237, 117-139.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G., *et al.* (2018). Ensembl 2018. *Nucleic acids research* 46, D754-D761.

Zheng, N.Y., Wilson, K., Wang, X., Boston, A., Kolar, G., Jackson, S.M., Liu, Y.J., Pascual, V., Capra, J.D., and Wilson, P.C. (2004). Human immunoglobulin selection associated with class switch and possible tolerogenic origins for C delta class-switched B cells. *The Journal of clinical investigation* 113, 1188-1201.

Zou, Y.-R., Grimaldi, C., and Diamond, B. (2017). B Cells. In Kelley and Firestein's Textbook of Rheumatology (Elsevier), pp. 207-230.e203.

Material Suplementario

Tabla S1: Proyectos de SRA seleccionados para mapear los límites de IH				
Estudio SRA	Muestras	Tipo de Célula	Nombre del Proyecto	Ref.
SRP009373				
	SRR364663	Células B CD19+	Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types	Salzman J, 2012
	SRR364679	Células B CD19+		
SRP020491				
	SRR805725	Células B circulantes (pre-vacunación)	Changes in gene expression profiles of circulating B cells after influenza vaccination in healthy human subjects	Henn AD, 2013
	SRR805736	Células B circulantes (pre-vacunación)		
	SRR805747	Células B circulantes (pre-vacunación)		
	SRR805758	Células B circulantes (pre-vacunación)		
	SRR805769	Células B circulantes (pre-vacunación)		
SRP021509				
	SRR834979	Células B vírgenes de tonsilas	EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation	Béguelin W, 2013
	SRR834980	Células B vírgenes de tonsilas		
	SRR834981	Células B vírgenes de tonsilas		
	SRR834982	Células B vírgenes de tonsilas		
	SRR834983	Células B de Centro Germinal Tonsilas		
	SRR834984	Células B de Centro Germinal Tonsilas		
	SRR834985	Células B de Centro Germinal Tonsilas		
	SRR834986	Células B de Centro Germinal Tonsilas		
SRP033335				
	SRR1038580	Células B Control CD19+	Genome-wide expression profiling of B Lymphocytes reveals IL4R increase in allergic asthma	Pascual M, 2014
	SRR1038581	Células B Control CD19+		
	SRR1038582	Células B Control CD19+		
SRP045500				
	SRR1550988	Células B de Control Sano	Next generation sequencing of human immune cell subsets across diseases	Linsley PS, 2014
	SRR1551042	Células B de Control Sano		
	SRR1551049	Células B de Control Sano		
	SRR1551070	Células B de Control Sano		
SRP048820				
	SRR1609989	Centrocito	Enhancer Sequence Variants and Transcription Factor Deregulation Synergize to Construct Pathogenic Regulatory Circuits in B Cell Lymphoma	Raju S, 2016
	SRR1609990	Centrocito		Koues OI et al., 2014
	SRR1609991	Células B activadas de sangre periférica		
	SRR1609992	Células B activadas de sangre periférica		
SRP048971				
	SRR1613932	Células B de Control Sano CD19+	Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell type specificity	Hrdlickova B et al., 2014
SRP051688				
	SRR1740034	Células B Primarias (pre-vacunación)	A Cell-based Systems Biology Assessment of Human Blood to Monitor Immune Responses After Influenza Vaccination	Hoek KL, 2013
	SRR1740062	Células B Primarias (pre-vacunación)		
SRP055390				
	SRR1812749	Células B Normales CD19+/CD27+	Transcriptome analysis in chronic lymphocytic leukemia cells using RNA sequencing (RNA-seq)	Kushwaha G, 2016
	SRR1812750	Células B vírgenes normales CD19+/ IgD+		
	SRR1812751	Células B Normales CD19+		
	SRR1812752	Células B Normales CD19+		
	SRR1812753	Células B Normales CD19+		
SRP060715				
	SRR2097512	Células B normales purificadas	Gene expression and splicing alterations analyzed by high throughput RNA sequencing of chronic lymphocytic leukemia specimens	Sin publicación
	SRR2097513	Células B normales purificadas		
	SRR2097514	Células B normales purificadas		
	SRR2097515	Células B normales purificadas		
	SRR2097516	Células B normales purificadas		

*Los datos son mostrados como aparecen en SRA para cada uno de los proyectos

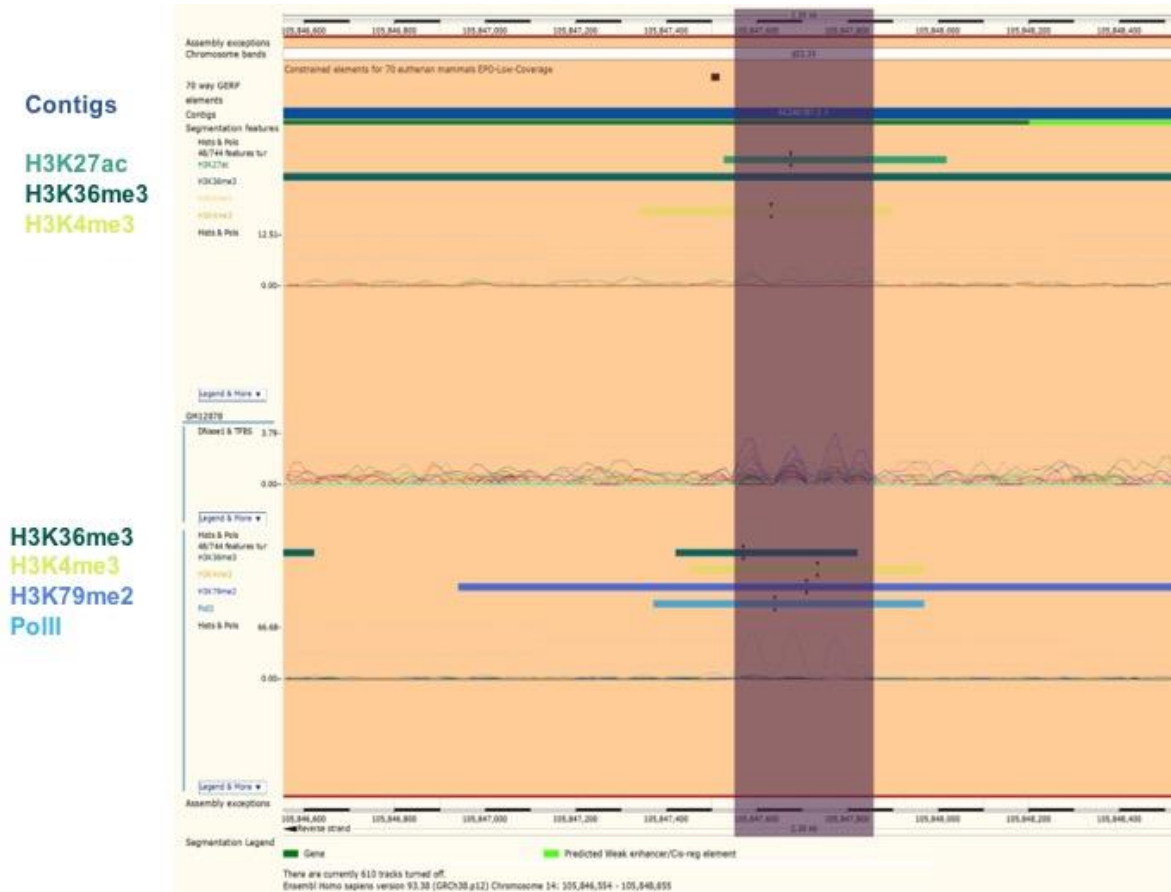


Figura S1. Nueva región transcrita en el locus del intrón IGHM-IGHD (I_{δ}). Se muestran las marcas epigenéticas asociadas con la transcripción activa en las células B. El buscador de genomas ENSEMBL se usó para buscar marcas epigenéticas en la HI identificada del lugar de IgH en cromosoma humano 14 (Zerbino et al., 2018). La pista superior corresponde a la anotación del proyecto Roadmap Epigenomics para células B de sangre periférica (Roadmap Epigenomics et al., 2015). Señales de H3K36me3 (traza verde oscuro) y H3K4me3 (traza ocre) y H3K27ac (traza verde). Todas las marcas de transcripción activa se superponen con el exón no codificante I_{δ} (rectángulo sombreado). La pista inferior corresponde a los datos del proyecto ENCODE que comprende a la línea de células linfoblastoides transformadas con EBV: GM12878 (Consortio, 2012). además de H3K36me3 (traza verde oscuro) y H3K4me3 (traza ocre), se identificó enriquecimiento de H3K79me2 (azul oscuro) y PolII (cielo azul).

Tabla S2 Transcripción de RCCnc en líneas celulares de cancer					
	Muestras con expresión RCCnc	Muestras sin expresión RCCnc	p_val	fdr	
Leucemia Linfocítica Aguda	6	111	0.4732423	0.6309897	
Leucemia Monocítica Aguda	12	142	0.7539633	0.9279548	
Leucemia Mieloide Aguda	2	185	0.0001564	0.0003128	*
Cancer de próstata	13	319	0.0128775	0.020604	
Cancer Pancreático	0	58	0.0206602	0.0300513	
Cáncer colorectal	9	580	9.83E-11	2.62E-10	*
Cáncer de mama	32	1417	3.84E-20	2.05E-19	*
Cáncer de pulmón	0	462	4.98E-16	1.59E-15	*
Cáncer gástrico	3	46	1	1	
Cáncer Adrenal	0	3	1	1	
Linfoma de Células B	77	38	6.16E-61	9.86E-60	*
Astrocitoma	1	534	7.04E-17	2.82E-16	*
Cáncer Cervical	7	1613	2.77E-48	2.21E-47	*
Leucemia Mieloide Crónica	8	522	7.16E-10	1.64E-09	*
Osteosarcoma	3	159	0.0033804	0.0060096	*
Colangiosarcoma	0	7	1	1	
Prueba exacta de Fisher Enriquecida Depletada					
**Corrección post hoc Benjamini-Hochberg para comparaciones multiples					

Proyecto	Descripción	Comparación (Exp vs Control)	Valor de p*										
			Muestra	IgM	IgD	IgG3	IgG1	IgA1	IgG2	IgG4	IgE	IgA2	
SRP032775	Malaria	Post-infección vs Pre-infección	Sangre	1	0.007	1	1	1	1	1	1	1	1
SRP033696	Influenza	Infección por H7N9 vs control	Sangre	2.1E-03	1	0.228	0.382	3.2E-04	1.7E-03	4.4E-02	1	0.052	
SRP059039	Diarrea pediátrica	<i>Shigella</i> vs Control	Sangre	0.977	0.032	1	1	1	1	1	1	1	1
SRP059039	Diarrea pediátrica	<i>Salmonella</i> vs Control	Sangre	1	1	1	1	1	0.546	1	1	1	1
SRP059039	Diarrea pediátrica	<i>Rotavirus</i> vs Control	Sangre	1	0.534	1	1	1	0.422	1	1	1	1
SRP059172	Infección Brucellosis/Leishmaniasis	<i>Brucellosis</i> vs Control	Sangre	1	1	1	0.853	1	1	1	1	1	1
SRP059172	Infección Brucellosis/Leishmaniasis	<i>Leishmaniasis</i> vs Control	Sangre	1	0.002	0.073	0.007	0.762	0.565	0.053	1	0.339	
SRP042228	Crohn pediátrico	Crohn vs No-IBD	Biopsia ileal	1.0E-03	1	0.013	0.011	5.0E-06	3.0E-03	2.0E-03	1	2.1E-06	
SRP042228	Crohn pediátrico	Crohn crónico vs No-IBD	Biopsia ileal	0.321	1	0.136	0.139	0.065	1.9E-02	2.4E-02	1	3.6E-02	
SRP042228	Crohn pediátrico	Colitis Ulcerativa vs No-IBD	Biopsia ileal	1	1	1	1	1	1	1	1	1	1
SRP051688	Influenza	Post-inmunizado(día 7) vs preimmune	Linfocitos B	1	1	1	0.071	1	1	1	1	0.893	
SRP020491	Influenza	Post-inmunizado(día 7) vs preimmune	CMSP (PBM)	1	1	1	0.267	1	1	0.954	1	1	1
SRP062966	Lupus Eritomatoso Generalizado(S	Pacientes con Lupus vs control	Sangre	1	1	0.988	0.555	0.190	1	1	1	1	0.308

*Valor de p ajustado por Bonferroni. En color verde, reducción significativa de la expresión con respecto al control sano. En color rojo, aumento significativo de la expresión con respecto al control sano.

Tabla S3. Análisis de expresión diferencial de C_HS