



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS
AVANZADOS DEL INSTITUTO
POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE
FARMACOLOGÍA

**“ANÁLISIS TRANSCRIPTÓMICO DE RETINOBLASTOMA USANDO UN
PUNTAJE DE DETECCIÓN PARA CADA TRANSCRITO CON EL
MICROARREGLO HTA2.0 DE AFFYMETRIX.”**

T E S I S

Que presenta

M en C. DIANA EMILCE ALVAREZ SUAREZ

Para obtener el grado de

**DOCTORA EN CIENCIAS
EN LA ESPECIALIDAD DE FARMACOLOGÍA**

Directoras de la tesis: Dra. Liliana Favari Perozzi
Dra. Martha Verónica Ponce Castañeda

Ciudad de México.

Julio, 2020

Este trabajo se llevó a cabo en la Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias del Hospital de Pediatría del Centro Médico Nacional Siglo XXI del Instituto Mexicano del Seguro Social y en el departamento de Farmacología del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, bajo la dirección y tutoría de las doctoras Liliana Favari Perozzi y Martha Verónica Ponce Castañeda y colaboración de los doctores Hugo Antonio Tovar y Enrique Hernández Lemus del Instituto Nacional de Medicina Genómica INMEGEN.

Proyecto apoyado a través del donativo IMSS (Instituto Mexicano de Seguro Social) R2015-785-049 y el donativo del NIH (National Institutes of Health) CA167833, CA192662, CA98180

Agradezco al Consejo Nacional de Ciencia y Tecnología por la beca No 407869 otorgada durante la realización de mis estudios de Doctorado.

Agradezco al Doctor Dr. Javier Camacho Arroyo y a la Dra. Lilibiana Favary por su colaboración y apoyo constante.

Finalmente agradezco a la M. en C. Elisabeth Gallegos Hernández y a la auxiliar en investigación Martha Noyola Diaz por el apoyo técnico recibido.

CONTENIDO

RESUMEN	1
ABSTRACT	3
1. INTRODUCCIÓN	5
1.1 Anatomía celular de la retina	6
1.2 Retinoblastoma.	7
1.3 Análisis transcriptómicos en retinoblastoma.	10
1.4. Tipos de microarreglos.	11
Microarreglos de expresión de Affymetrix:	12
1.5 Antecedentes de análisis de detección.	14
2. PLANTEAMIENTO DEL PROBLEMA	17
3. JUSTIFICACIÓN E HIPÓTESIS	18
3.1 JUSTIFICACIÓN.	18
3.2 HIPÓTESIS.....	18
4. OBJETIVOS	19
4.1 OBJETIVO GENERAL.....	19
4.2 OBJETIVOS PARTICULARES.....	19
5. METODOLOGÍA	20
5.1 Características de las muestras.....	20
Cultivo primario:.....	20
Análisis Histopatológico.....	20
5.2 Extracción de RNA	21

5.3 Hibridación en el microarreglo HTA2.0	21
5.4. Análisis Bioinformático	22
5.4.1 Preprocesamiento:.....	22
Características importantes del chip HTA2.0 para el diseño del algoritmo de detección	23
5.4.2 Análisis de detección	24
Detección de exones y transcritos:.....	25
Detección de genes:	25
5.4.3. Análisis funcional de genes:	26
5.4.4 Expresión diferencial:	27
5.5 Validación.....	28
5.5.1 Validación de núcleo de genes detectados con sets de datos publicados.....	28
5.5.2 Validación de transcritos expresados diferencialmente por medio de qRT-PCR.	29
Determinación de la expresión de los genes problema.	29
Curva de disociación.....	30
6. RESULTADOS.....	31
6.1 Tamaño del transcriptoma y estructura.....	32
Descripción del transcriptoma en términos de genes y transcritos detectados.....	32
Grupos de genes detectados en las muestras.....	33
Descripción del transcriptoma en términos de transcritos detectados.....	35
6. 2. Enriquecimiento funcional en grupos de muestras de retinoblastoma.....	35
Análisis de enriquecimiento del grupo genes “presentes”.	35
Análisis de enriquecimiento “grupo genes ausentes”:	36
Análisis de enriquecimiento “grupo genes variables”:	37

Enriquecimiento en línea celular Y79.....	37
6. 3 Búsqueda de posibles blancos terapéuticos.	38
6.4 Hallazgos en el grupo de genes detectados de manera “Variable”.	41
Enriquecimiento Funcional de transcritos expresados diferencialmente	42
Análisis histopatológico de las muestras	44
6.5 Validación bioinformática.....	46
Detección de genes presentes.....	46
Cinasas en el grupo de genes “presentes”.....	47
Validación experimental de genes “variables”.	48
Expresión relativa de mRNA de GUCA1C por PCR tiempo real.....	48
Expresión relativa de mRNA de PDC por PCR tiempo real.....	49
Expresión relativa de mRNA de PDE6C por PCR tiempo real.	50
7. DISCUSIÓN	51
8. CONCLUSIÓN.....	56
9. PERSPECTIVAS.....	56
APÉNDICES.....	57
Comandos utilizados en RStudio para el análisis.	57
Transcritos expresados diferencialmente (Muestras Bilaterales Vs Muestras unilaterales) ...	66
Genes presentes exclusivamente en la línea celular Y79.....	68

INDICE DE FIGURAS

Figura 1. Estructura de la retina.	7
Figura 2. Patogenia del Retinoblastoma.	8
Figura 3. La vía canónica RB / E2F en cáncer.....	9
Figura 4. Tipos de microarreglos de expresión de Affymetrix	13
Figura 5. Diseño de microarreglos 3IVT y microarreglos de exones	14
Figura 6. Agrupamiento Jerárquico y análisis de blancos del miRNOMA de RB.....	15
Figura 7. Diseño experimental.	22
Figura 8. Diseño del microarreglo HTA2.0.	23
Figura 9. Determinación de la detección de probesets por encima del fondo (DABG).....	24
Figura 10. Estructura de la base de datos ENSEMBL.	25
Figura 11. Ejemplo de análisis de detección.	26
Figura 12. Análisis bioinformático.	28
Figura 13. Depuración de probesets PSR y JUC.	31
Figura 14. Número de transcritos por gen.....	33
Figura 15. Mapa transcriptómico de detección.....	34
Figura 16. Mapas de detección de genes y transcritos correspondientes	35
Figura 17. Nivel de expresión de Cinasas.....	41
Figura 18. Expresión diferencial por lateralidad.....	42
<i>Figura 19. Análisis Histopatológico de las muestras.</i>	<i>45</i>
Figura 20. Análisis Detección de 18 casos de retinoblastoma con Plataforma HUGENE1ST... ..	47
Figura 21. Heatmap del puntaje de detección cinasas en el grupo "presente".	47

Figura 22. Expresión de mRNA de GUCA1C en Rb.....	49
Figura 23. Expresión de mRNA de PDC en el Rb.....	49
Figura 24. Expresión de mRNA de PDE6C en el Rb.	50

INDICE DE TABLAS

Tabla 1. Características clínicas de las muestras.	21
Tabla 2. Protocolo de temperaturas PCR.	29
Tabla 3. Tamaño del transcriptoma detectado.	32
Tabla 4. Las 5 principales rutas enriquecidas en grupo de Core “genes presentes”.	36
Tabla 5. Las 5 principales rutas enriquecidas en grupo de “ genes ausentes”.	36
Tabla 6. Las 5 principales rutas enriquecidas en grupo de “genes variables”	37
Tabla 7. Enriquecimiento funcional de genes ausentes exclusivamente en la línea celular Y79.....	38
Tabla 8. Cinasas en el grupo de genes "presentes" e inhibidores aprobados por la FDA.	39
Tabla 9. Categorías Enriquecidas en transcritos expresados diferencialmente.....	43
Tabla 10. Procesos biológicos enriquecidos en muestras unilaterales.....	43
Tabla 11. Procesos Biológicos enriquecidos en muestras bilaterales.	44
Tabla 12. Primers para qRT-PCR	48

RESUMEN

Introducción: El análisis de transcriptoma a través de microarreglos ha permitido generar información relevante para el estudio molecular de diferentes enfermedades, particularmente en cáncer. Clásicamente, los análisis de microarreglos de expresión se han centrado en la identificación de genes que se expresan diferencialmente entre dos condiciones distintas (tejido enfermo contra sano). Este abordaje brinda información relevante, sin embargo, no permite responder interrogantes fundamentales como el tamaño del transcriptoma activo de un tejido o de una condición particular o si existe una maquinaria transcripcional común entre muestras de una misma categoría diagnóstica. Contar con un panorama transcriptómico de este tipo abre grandes posibilidades de entendimiento de una enfermedad como lo es el cáncer. El retinoblastoma (Rb) es un tumor maligno intraocular que se presenta durante la primera infancia (1 a 5 años de edad). Los estudios transcriptómicos de este tumor han sido de gran interés en las últimas décadas. Sin embargo, estos estudios han sido limitados por la falta de grandes números de muestras y poca cantidad de tejidos normales para contrastar por medio de expresión diferencial. Sumado a esto, el enfoque de estos estudios se ha basado en el análisis diferencial de los datos de expresión, ignorando los datos de detección que se pueden obtener con las tecnologías de alto rendimiento. El puntaje de detección permite obtener perfiles de elementos detectados y no detectados en todas las muestras sin la necesidad de una muestra o tejido control. En este estudio, nos propusimos analizar 8 tumores primarios cultivados de Rb y la línea celular Y79 con el objetivo de determinar el tamaño del transcriptoma, encontrar un conjunto de genes detectados y compartidos por todas las muestras, así como su variabilidad inter tumoral, sin la necesidad de tejido control.

Metodología: Las muestras se obtuvieron de ojos enucleados de pacientes afectados con Rb tratados en el Hospital de Pediatría en el CMN SXXI, del Instituto Mexicano de Seguro Social y el Hospital Infantil de México Federico Gómez, de la Secretaría de Salud en Ciudad de México. Los tejidos tumorales fueron cultivados por una semana y se extrajo RNA total. Se utilizó el chip de Affymetrix HTA2.0. El análisis bioinformático incluyó la normalización y análisis de detección de cada exón como presente o ausente y el desarrollo de un criterio o algoritmo para determinar si un gene o transcrito es detectable o no. Con estos puntajes los genes fueron analizados por medio de agrupamiento jerárquico y la variabilidad intertumoral se analizó por medio de análisis diferencial. Los genes detectados fueron enriquecidos funcionalmente con la herramienta Webgestalt, y con un abordaje heurístico encontramos en

el grupo de genes compartidos por todas las muestras, quinasas que cuentan con inhibidores específicos aprobados de la FDA y que potencialmente podrían redireccionarse en el tratamiento del retinoblastoma.

Resultados: Se encontró un grupo de 7513 genes detectados y 4321 genes no detectados compartidos en todas las muestras de Rb. Se encontró notable variabilidad en un grupo de 7681 genes. Las rutas metabólicas enriquecidas en el grupo de genes detectados están relacionadas con ciclo celular, transporte de RNA y replicación de DNA. Al realizar el análisis incluyendo la línea celular Y79 el número de genes detectado por todas las muestras disminuye. Finalmente, se encontraron 124 transcritos expresados diferencialmente entre muestras de tumores bilaterales frente a las muestras unilaterales de Rb.

Conclusión: Esta nueva propuesta de análisis permitió encontrar un grupo central de genes detectados y un grupo no detectado compartidos por todas las muestras, así como un grupo de genes correspondientes a la heterogeneidad entre estas muestras de Rb, es decir, con variabilidad a nivel de detección. Además este abordaje brinda una vía fácil y coherente para interpretaciones biológicas plausibles de datos masivos de RNA que no requiere un tejido control o sano

ABSTRACT

Introduction: Transcriptome analysis through microarrays has allowed the generation of relevant information for the molecular study of different diseases, specifically cancer. Microarray expression analysis have focused on the identification of genes that are differentially expressed between two different conditions (diseased versus healthy tissue). This approach provides relevant information; however, it does not allow answering fundamental questions like the size of the transcriptome in a particular tissue or condition or whether there is a common transcriptional machinery between samples of the same diagnostic category. Providing a broad molecular panorama at the transcriptomic level opens up great possibilities for understanding a disease such as cancer. Retinoblastoma (Rb) is a malignant intraocular tumor that occurs in early childhood (1-5 years old). Transcriptomic studies of this tumor have been of great interest in recent decades. However, these studies have been limited by the lack of large numbers of samples and a small amount of normal tissues to contrast by means of differential expression. In addition to this, the focus of these studies has been based on differential analysis of expression data, ignoring the detection data that high-performance technologies yield. Detection scoring allows to obtain profiles of detected or undetected elements in all samples without the need for a control sample. In this study, we set to analyze 8 cultured primary tumors of Rb and the Y79 cell line in order to find the transcriptomic size, a set of genes detected and shared by all the samples, as well as their inter-tumor variability, without the need for tissue control.

Methodology: The samples were obtained from enucleated eyes of patients with Rb treated at the Pediatric Hospital at CMN SXXI, from the Mexican Institute of Social Security and the Federico Gómez Children's Hospital of Mexico, of the Ministry of Health in Mexico City. Tumor tissues were cultured for only one week and total RNA was extracted, and we used Affymetrix HTA2.0 chip. Bioinformatic analysis included normalization and detection analysis of each exon as present or absent, and the development of an algorithm to determine if a gene or transcript is detectable or not. With these scores the genes were analyzed by means of hierarchical clustering and the intertumoral variability was analyzed by means of differential analysis. The detected genes were functionally enriched with the Webgestalt tool, and with an heuristic approach we found in the group of shared genes, kinases with specific FDA approved inhibitors that could be redirected in the treatment of retinoblastoma.

Results: A group of 7,513 detected genes and 4,321 undetected genes were found shared in all Rb samples. Remarkable variability was found in a group of 7,681 genes. The metabolic pathways enriched in the group of genes detected and shared by all samples are related to cell cycle, RNA transport, and DNA replication. When performing the analysis with the Y79 cell line, the number of genes detected by all the samples decreased. Finally, 124 differentially expressed transcripts were found between bilateral tumor samples versus unilateral Rb samples.

Conclusion: This new analysis proposed allows finding a central group of detected genes and undetected genes shared by all tumors studied, as well as tumor heterogeneity in Rb samples, that means, with variability at detection level. Furthermore, this approach provides a consistent route for plausible biological interpretations of massive RNA data that does not require healthy or control tissue.

1. INTRODUCCIÓN

Los microarreglos de expresión son una poderosa tecnología que permite el análisis a gran escala de perfiles de expresión de RNA en tejidos específicos (Cieślik & Chinnaiyan, 2018 ;Plomin & Schalkwyk, 2007). El análisis de este tipo de microarreglos se realiza por medio de análisis transcriptómico, que convencionalmente está basado en la expresión diferencial (Tarca, Romero, & Draghici, 2006). La ventaja que ofrece el análisis de expresión diferencial radica en que permite identificar patrones diferentes en la expresión de genes entre dos condiciones como tejido sano y tejido tumoral (Tarca *et al.*, 2006), sin embargo, este acercamiento puede dejar de lado importantes detalles para la reconstrucción de rutas metabólicas ya que solamente se identifica la porción de genes con mayor diferencia entre condiciones (Haynes, Higdon, Stanberry, Collins, & Kolker, 2013). Otra desventaja importante de este análisis en el caso del cáncer es la necesidad de contar con un tejido control o normal para contrastar. En muchos casos, obtener tejido normal que sea comparable con el tumoral presenta muchas dificultades, especialmente en retinoblastoma (Rb), ya que es un tumor intraocular de la infancia temprana que se origina en las células precursoras de foto-receptores en la retina (Dimaras & Corson, 2018).

Los análisis de microarreglos a nivel transcriptómico, se basaron originalmente en mediciones relativas. Inicialmente en microarreglos de doble canal, se incluían dos muestras marcadas con dos fluoróforos distintos en el mismo chip y se analizaban por medio de expresión relativa. Posteriormente, el diseño de microarreglos evolucionó a un solo canal lo cual ofrece la ventaja de generar datos de expresión absoluta de genes. El valor de expresión absoluta de estos microarreglos de un solo canal, supone una relación lineal entre la señal fluorescente detectada y la concentración RNA en la muestra de prueba. Sin embargo, debido al amplio rango de concentración para genes expresados en una muestra biológica, la intensidad de fluorescencia detectada no permanece necesariamente en el rango lineal para todos los genes (Shi *et al.*, 2005). Por esta razón, siempre se buscan y se han favorecido los análisis de expresión diferencial, que mitigan el error generado por las variaciones en concentración de genes en una muestra al obtener una medición relativa entre dos condiciones. Estos análisis son muy útiles, pero dejan de lado información relevante que impide obtener un panorama completo y descriptivo de las muestras analizadas por separado.

Con el fin de proponer una nueva aproximación para el estudio transcriptómico de microarreglos de expresión, tomando ventaja de las mediciones absolutas que las plataformas de un solo canal ofrecen, en este trabajo se propone un análisis novedoso para el estudio del transcriptoma de retinoblastoma usando el puntaje de detección de los genes y de los transcritos correspondientes que los microarreglos de expresión proporcionan. El puntaje de detección califica cada elemento o sonda del microarreglo como detectable o no detectable, esto en primer lugar permite determinar con precisión el tamaño del transcriptoma y en el contexto del cáncer, la importancia de este acercamiento deriva de la posibilidad de identificar los elementos compartidos por todas las muestras de un grupo, así como los elementos o mRNA asociados con la variabilidad inter tumoral, todo esto sin la necesidad de un tejido control. Teniendo en cuenta que el retinoblastoma se origina en las células de la retina, es pertinente conocer la morfología de la retina.

1.1 Anatomía celular de la retina

La retina es el órgano responsable de la transducción de la luz en señales neuronales que eventualmente son percibidas como imágenes. Específicamente, la luz incide en los fotorreceptores, una compleja capa de células especializadas, conocidas como conos y bastones, que contienen foto pigmentos que producen energía química ante la exposición a la luz. Esta energía es convertida en energía eléctrica que es procesada por la retina y transmitida por el nervio óptico al cerebro (Ofri, 2008). La retina es el único órgano que puede ser estudiado de manera no invasiva por medio de oftalmoscopio para evaluar detalles de procesos patológicos *in vivo*, que en otros órganos sólo se pueden hacer visibles por medio de técnicas quirúrgicas e histopatológicas.

La retina está compuesta por al menos seis tipos celulares, bastones, conos, células horizontales, amacrinas, bipolares y ganglionares y un tipo de células gliales denominadas de Müller, todas estas son derivadas de una célula progenitora retinal común. En la **Figura 1** se representa la estructura de la retina.

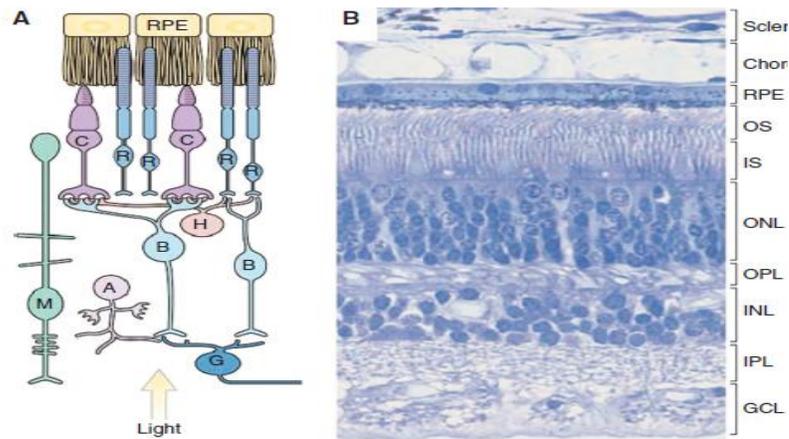


Figura 1. Estructura de la retina. A. Organización de la retina B. Tinción de Hematoxilina Eosina de la retina muestra los segmentos externos (OS-Outer Segments) e internos (IS - Inner Segments) de los bastones y cono receptores. El núcleo de los fotorreceptores forma la capa externa nuclear (ONL- Outer Nuclear Layer). Los núcleos de las células bipolares, amacrinas, horizontales y células de Müller forman la capa interna nuclear (INL- Inner Nuclear Layer) y el núcleo de las células ganglionares forma la capa de células ganglionares (GCL- Ganglion Cell Layer). La capa plexiforme externa (OPL- Outer Plexiform Layer) contiene las terminales sinápticas de los fotorreceptores, células horizontales y células bipolares. La capa plexiforme interna (IPL- Inner Plexiform Layer) contiene las terminales sinápticas de las células bipolares, amacrinas y células ganglionares. A, Células amacrinas; B, Células bipolares; C, conos; G, células ganglionares; H, células horizontales; M, células de Müller; R, bastones; epitelio de pigmento retinal (RPE- Retinal Pigment Epithelium). Reproducido de (Sung & Chuang, 2010).

1.2 Retinoblastoma.

El retinoblastoma es la neoplasia maligna intraocular más común en la infancia, con una incidencia entre 1 en 15,000 a 1 en 18,000 nacidos vivos. La mayor parte de casos afecta únicamente un ojo (unilateral) y constituye entre el 60 al 65% de los casos, mientras que los tumores bilaterales se presentan en aproximadamente el 25–35% de los casos, siendo estos últimos siempre heredables a la siguiente generación. Los casos unilaterales se diagnostican alrededor de los 24 meses y los casos bilaterales se suelen presentar antes de los 12 meses de edad (Pandey, 2014).

Para explicar la patogenia de la enfermedad y basándose en observaciones clínicas y epidemiológicas, en 1971, Knudson propuso la hipótesis del doble golpe (two hits). Propuso que para que se desarrolle el retinoblastoma, se necesitan dos eventos mutacionales. En los casos hereditarios, el golpe o alteración inicial consiste en una mutación germinal, que se hereda y se encuentra en todas las células, esto explicaría que los casos hereditarios están predispuestos al desarrollo de tumores secundarios como el osteosarcoma; en estos casos

afirmó que el segundo golpe se presenta en las células retinianas somáticas que conduce al desarrollo del retinoblastoma. Propuso además que en los casos esporádicos unilaterales, ambos eventos mutacionales ocurren durante el desarrollo de la retina y son mutaciones somáticas, explicando así que en estos casos, no hay riesgo de segundos tumores (Knudson, 1971). El desarrollo de herramientas de la biología molecular particularmente la secuenciación de DNA le dieron la razón. Los eventos mutacionales generan una inactivación bialélica del gen de susceptibilidad al retinoblastoma *RB1*, localizado en el cromosoma 13q14 que codifica para la proteína pRB. Esto sucede en momentos diferentes del desarrollo en la forma esporádica y hereditaria del retinoblastoma. En la forma hereditaria también llamada familiar, un alelo está mutado en la línea germinal y la otra mutación ocurre somáticamente en la retina en desarrollo. En la enfermedad no hereditaria, las mutaciones de ambos alelos son somáticas (Corson and Gallie 2007; Mairal *et al.*, 2000; Kumar, Abbas, and Aster 2013) (**Figura 2**)

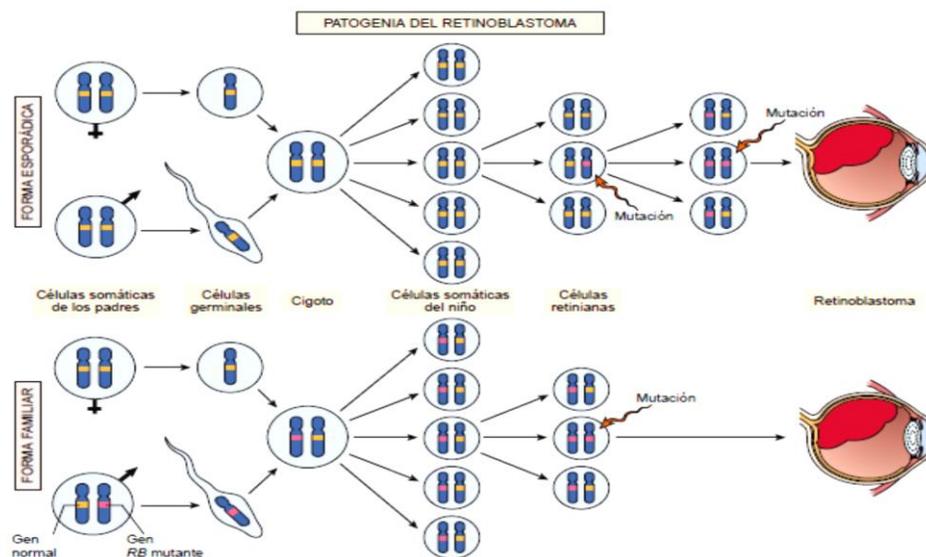


Figura 2. Patogénesis del Retinoblastoma. Se requieren dos mutaciones del locus *RB1* para la proliferación neoplásica de las células retinianas. **Forma esporádica:** Ambas mutaciones en el locus *RB1* son adquiridas por las células retinianas. **Forma familiar:** Las células heredan un gen *RB1* mutante de un padre (primer golpe). La segunda mutación afecta al locus *RB1* en una de las células de la retina (segundo golpe). Figura tomada de Kumar *et al.*, 2015

Las predicciones hechas por Knudson en 1971 a partir de la evidencia epidemiológica llevaron a la identificación 15 años después del gen *RB1*. La caracterización inicial de la función de la proteína pRb apuntaban al papel de pRb como un regulador de la transición del ciclo celular en la fase G1/S. Ahora se sabe que tanto las fases G1/S como G2/M del ciclo celular de los mamíferos están controladas por una ruta molecular compleja que involucra a

miembros de la Factor de unión al promotor E2 (E2F), pareja de dimerización (DP- Dimerization partner), la proteína pRb, quinasas dependiente de ciclinas (CDK- Cyclin-dependent Kinases), ciclinas y familias de inhibidores de CDK (CDKN)(Engel, Cress, & Santiago-Cardona, 2014). Esta vía es de gran importancia porque está interrumpida en la mayoría de los tumores sólidos (**Figura 3**).

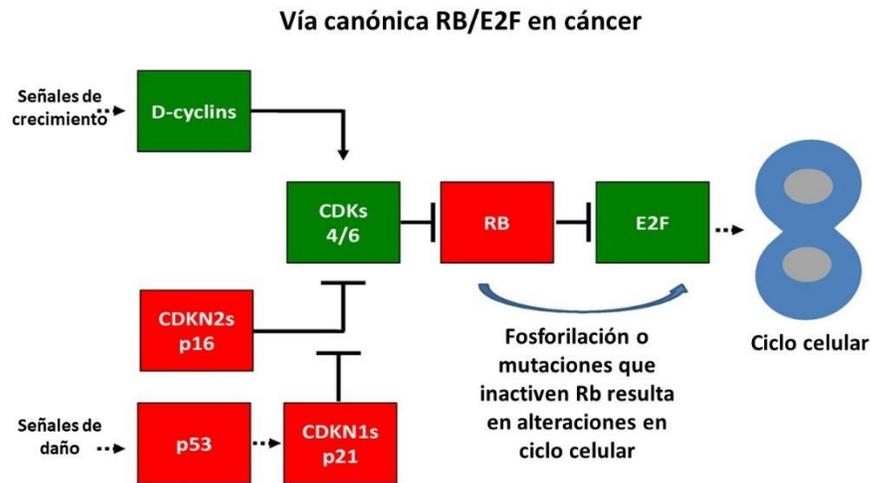


Figura 3. La vía canónica RB / E2F en cáncer: En este esquema, las proteínas conocidas como supresoras de tumores se resaltan en rojo, mientras que las proteínas promotoras de tumores se resaltan en verde. Las flechas puntiagudas sólidas indican activación directa. Las ciclinas se unen a CDK para activarlas. Las líneas romas indican la inactivación directa mediante. RB que une E2F para abolir su actividad transcripcional o mediante la modificación de proteínas. La fosforilación o mutaciones que inactiven pRB resulta en su incapacidad para unirse y reprimir E2F. Las líneas discontinuas indicaron regulación indirecta. Figura tomada y modificada de (Engel et al., 2014)

Aproximadamente el 15% de los retinoblastomas esporádicos unilaterales son causados por mutaciones germinales que afectan únicamente a un ojo (Yunis, 1983). Existen sin embargo un número pequeño de casos en los que no se han encontrado mutaciones y se han propuesto otro tipo de mecanismos que silencian el gene RB1 como la metilación del promotor o incluso la posibilidad de un agente viral, ya que se han reportado secuencias de Virus de Papiloma Humano en muestras de retinoblastoma por PCR aunque esto último es controvertido (Orjuela et al., 2000;Shetty et al., 2012), ya que otros grupos de investigación no han podido replicar estos hallazgos por lo que se requieren más estudios para aseverar la existencia de asociación entre este o algún virus y la aparición de retinoblastoma.

1.3 Análisis transcriptómicos en retinoblastoma.

El estudio de transcriptoma en retinoblastoma ha permitido identificar algunos mecanismos de crecimiento e invasión tumoral, así como el desarrollo de nuevas formas de atacar terapéuticamente estos tumores (Elchuri, Rajasekaran, & Miles, 2018). Los estudios genómicos y de microarreglos de expresión han permitido sugerir a los bastones y conos como las principales células de origen del retinoblastoma (Xu *et al.*, 2009; Xu *et al.*, 2014; Kapatai *et al.*, 2013). También se ha encontrado que aquellos tumores de retinoblastoma hereditarios tienen menos reordenamientos genómicos en comparación con los no hereditarios que suelen presentar mayores niveles de inestabilidad genómica (Corson y Gallie, 2007; Mol *et al.*, 2014). Estos cambios genómicos principalmente dan como resultado la ganancia de una serie de regiones cromosómicas que incluyen 1q, 2p, 6p y 13q y la pérdida de la región 16q (Kooi *et al.*, 2016).

Conjuntos adicionales de muestreo de genoma completo (WGS) han encontrado otras regiones de cambios genómicos en retinoblastoma e identificaron genes involucrados en la regulación de diferenciación (Ganguly *et al.*, 2009).

Sumado a estos antecedentes, resultados previos de nuestro grupo, a través de secuenciación masiva (NGS- Next Generation Sequencing) de baja cobertura, se encontraron patrones de ganancias y pérdidas génicas en muestras de Retinoblastoma asociadas a las bandas citogenéticas claras y oscuras respectivamente. En este abordaje, se encontraron pérdidas importantes en el brazo q del cromosoma 13 incluyendo el locus de retinoblastoma (13q14.2). Esta pérdida se extiende mucho más allá del locus de retinoblastoma, incluidas más de 11 bandas citogenéticas de 13q14.2 a 13q31.3. En este estudio, las ganancias están en cromosoma 1, 6, 7, 9, 11, 12, 16, 17, 19 y 22, aunque fueron de menor magnitud y más discontinuas en comparación con las pérdidas observadas en Chr13 (García-Chequer *et al.*, 2016).

En resumen, estas tecnologías han permitido enriquecer el entendimiento de la biología molecular del retinoblastoma (Pritchard *et al.*, 2014; Assayag *et al.*, 2016). Este tumor, se ha estudiado con prácticamente todas las herramientas moleculares disponibles, incluyendo las transcriptómicas (Bowles *et al.* 2007; Corson and Gallie 2007; Friend *et al.* 1986; Ganguly *et al.* 2009; Kapatai *et al.* 2013; Kooi *et al.* 2015; Lee *et al.* 1987; Mol *et al.* 2014; Xu *et al.* 2009, 2014). Sin embargo, de estos estudios, no se ha derivado blancos que formen parte del

tratamiento actual. Los estudios transcriptómicos previos, particularmente en microarreglos, han permitido conocer parte de su biología, pero los métodos tienen limitaciones. Una de las limitaciones es que se han realizado por medio de análisis de expresión diferencial dejando de lado datos absolutos. La identificación de qué genes se expresan en tejidos de retinoblastoma permitirá aumentar la comprensión sobre el papel de la expresión génica en la biología de este tipo de tumores y posiblemente sus blancos terapéuticos.

1.4. Tipos de microarreglos.

El diseño de los microarreglos de expresión ha evolucionado desde las plataformas de dos canales a plataformas de un canal. Las plataformas de dos canales cuentan con muestras de RNA de dos condiciones experimentales marcadas por separado con dos fluoróforos distintos e hibridadas competitivamente con las sondas en el mismo arreglo. (Smyth & Altman, 2013) Esto implica que los microarreglos de doble canal están diseñados para identificar la *expresión génica relativa* de una muestra de una condición experimental en comparación con otra muestra control o sana. Este tipo de arreglos cuenta con la ventaja de ajustar ambas muestras del microarreglo al efecto de fondo y generar los resultados en términos de las proporciones logarítmicas, lo que significa que cada unidad indica un cambio de al menos el doble entre las muestras. Sin embargo, estas plataformas no utilizan en su análisis toda la información disponible en las observaciones de cada canal o muestra por separado (Sedighi & Li, 2014). A diferencia de estos arreglos, los microarreglos de un canal y en específico los de Affymetrix, generan información de niveles de *expresión génica absoluta* de una muestra o condición experimental por cada microarreglo. Los datos generados por estos microarreglos están basados en cadenas cortas de oligonucleótidos (20 a 30 nt) denominados sondas, distribuidas a lo largo del gen. Estas sondas emiten muchas mediciones para cada gen que se combinan en una única medida de expresión génica. Esto genera el problema de ajuste de fondo en cada microarreglo, que se corrige en la normalización entre todas las condiciones experimentales. Los microarreglos de un canal permiten obtener un panorama amplio de cada muestra. Este tipo de arreglos facilitan la comparación a través de expresión diferencial, que permite encontrar los genes significativamente diferentes entre dos condiciones (Sîrbu, Kerr, Crane, & Ruskin, 2012), aun así, la información de detección que brinda esta tecnología no ha sido explorada.

Microarreglos de expresión de Affymetrix:

Los microarreglos de Affymetrix, son arreglos de sondas (oligonucleótidos de alrededor 25nt) de alta densidad. Para brindar mayor especificidad en los resultados, y según la información que se desee obtener, las sondas de estos microarreglos se pueden diseñar para detectar una región específica de transcritos o para detectar exones específicos y obtener alternativas de splicing. Teniendo en cuenta este acercamiento, los microarreglos de expresión de Affymetrix pueden tener los siguientes diseños (Dalma-Weiszhausz, Warrington, Tanimoto, & Miyada, 2006):

1. Microarreglos para estudio de expresión 3'UT: Estos son microarreglos denominados en inglés "3' Expression arrays". Cuentan en promedio con 11 sondas (que en grupo se denominan probesets) por gen. Este grupo de sondas o probesets detectan la región 3' de transcritos previamente identificados y anotados por diferentes bases de datos.
2. Microarreglos de exones: Con el avance de los nuevos microarreglos sensibles a transcritos, las plataformas Human gene 1.0 ST y Human Exon 1.0 ST fortalecieron el análisis de transcriptoma completo. En estos arreglos los probesets (con 2 a 4 sondas por probeset dependiendo la plataforma) se unen a los exones a lo largo del gen. Inicialmente este tipo de arreglos demostraron eficacia en la caracterización del splicing alternativo y la variación de isoformas a escala de todo el genoma, sin embargo, no cuentan con probesets para analizar el tipo de uniones que hacen los exones.
3. Microarreglos de transcriptoma completo: En este grupo está el microarreglo HTA2.0 (del inglés Human Transcriptome Array 2.0). Esta es una de las plataformas con mayor resolución de Affymetrix para el perfil de expresión génica. El diseño de esta plataforma permite detectar tanto ARN codificantes como no codificantes. El microarreglo HTA2.0 cuenta sondas distribuidas a lo largo de todos los exones y uniones interexonales. En este diseño, los probesets están conformados por 10 sondas que se unen específicamente a exones, denominadas PSR (Probe Selection Regions), y probesets conformados por 4 sondas

que hibridan con uniones interexonales, denominadas JUC (Junctions). Este arreglo cuenta con la ventaja de permitir una descripción detallada de las alternativas de splicing en el transcriptoma (**Figura 4**).

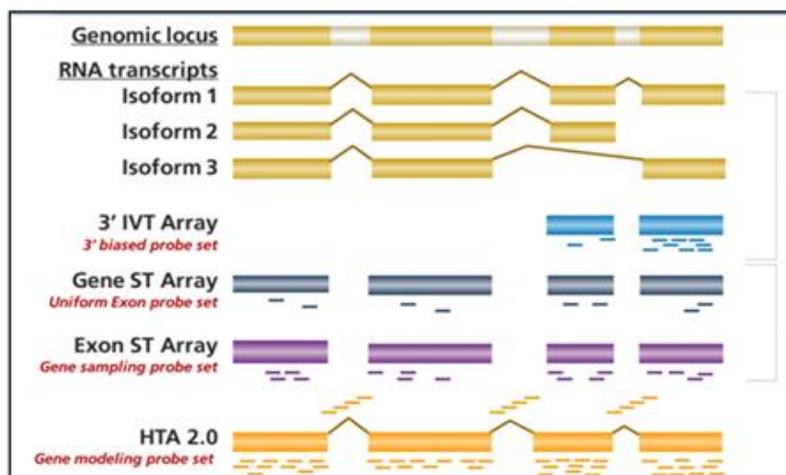


Figura 4. Tipos de microarreglos de expresión de Affymetrix: Representación de las diferentes estrategias de diseño para selección de la sondas. Un ejemplo de la secuencia del gen se muestra en la parte superior representa de los transcritos de destino. Los rectángulos amarillos representan exones, y los rectángulos blancos representan intrones. **3' IVT**: Las secuencias objetivo de las sondas de expresión de 3' IVT están en el extremo 3' y son adyacentes a la cola poli (A) del RNA. **Gene ST Y Exon ST**: Los conjuntos de sondas de las plataformas HuGene ST y Exon ST incluyen probesets que están dentro de los exones (probesets de 2 sondas para HugeneST y probesets de 4 sondas para Huexon ST). **HTA2.0**. Para el diseño de esta plataforma se cuentan con probesets de 10 sondas que se unen a exones y probesets de 4 sondas que están en uniones interexonales .

El primer diseño de los microarreglos 3'IVT consideró la expresión a nivel de genes individuales. Estos microarreglos no permitían determinar isoformas y las sondas estaba ubicada únicamente en el extremo 3' del transcrito. Sumado a esto, existía el potencial de perder ciertos transcritos completamente por degradación de RNA. Para mitigar estas limitaciones, se diseñaron microarreglos para investigar genes mediante el uso de muchas sondas colocadas a lo largo de todos los transcritos, para interrogar exones por separado, obteniendo la mayor cobertura posible y así poder identificar o descubrir transcritos.(Liu *et al.* 2002; Plomin and Schalkwyk 2007)

Los avances en la tecnología de microarreglos han hecho posible el diseño de chips cada vez más pequeños y con mayor cantidad de sondas. Los primeros microarreglos de exones, usan más de 6.5 millones de sondas, mientras que la generación anterior (3-IVT) tenía aproximadamente 600,000. En este diseño se eliminan la sondas MM (Mis-Match, sondas en

las que se cambia uno de los residuos del centro para determinar la especificidad de la hibridación), que se tenían en versiones previas y se emplean para tener en cuenta y corregir la señal de hibridación inespecífica en las mediciones de intensidad global. En generaciones posteriores se conservaron únicamente las sondas Perfect match (PM) (sondas que empalman perfectamente con la secuencia del transcrito) (Clark *et al.*, 2007) (**Figura 5**).

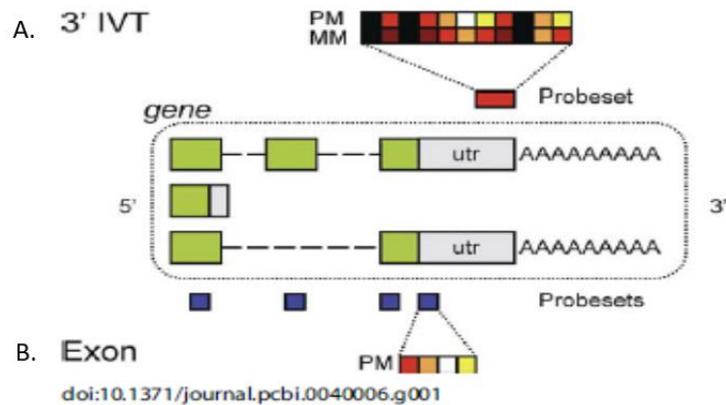


Figura 5. Diseño de microarreglos 3'IVT y microarreglos de exones. A. **Microarreglo 3'IVT:** cada gen es medido por un conjunto de sondas (probesets) colocadas en el extremo 3' de los transcritos. Estos probesets constan de 11 sondas Perfect Match (PM) y 11 sondas Mis-Match (MM). B. **Microarreglos de exones (Huexon ST):** En el diseño tienen probesets que hibridan con cada exón a lo largo del gen. Los probesets de lo microarreglos de exones no tienen puntos MM emparejados y tienen 4 sondas por probesets. Figura tomada de Clark *et al.*, 2007.

1.5 Antecedentes de análisis de detección.

En estudios previos, se ha realizado el análisis de puntajes de detección para obtener una descripción detallada del miRNOMA (microRNA) en 12 muestras de retinoblastoma. En ese análisis se obtuvo el valor o puntaje de detección de 2578 miRNAs codificados en el microarreglo miRNA 4.0 de Affymetrix (Castro-Magdonel *et al.*, 2017). El análisis por medio de puntaje de detección permitió obtener un grupo que representa la maquinaria central compuesta por 142 miRNAs que se encuentran compartida de manera uniforme por todas las muestras de retinoblastoma estudiadas, así como un grupo de 995 miRNAs considerados como "apagados" o no detectados en todas las muestras, y un tercer grupo de 1022 miRNAs detectados de manera variable entre muestras. En la **Figura 6a** se encuentra el agrupamiento jerárquico de las muestras que permitió encontrar los grupos de miRNAs.

El análisis funcional de 30 miRNAs con mayor nivel de expresión en el clúster de los 142 miRNAs presente en todas las muestras, reveló que sus blancos son 48 genes supresores de tumores entre ellos RB1. (Figura 6b)

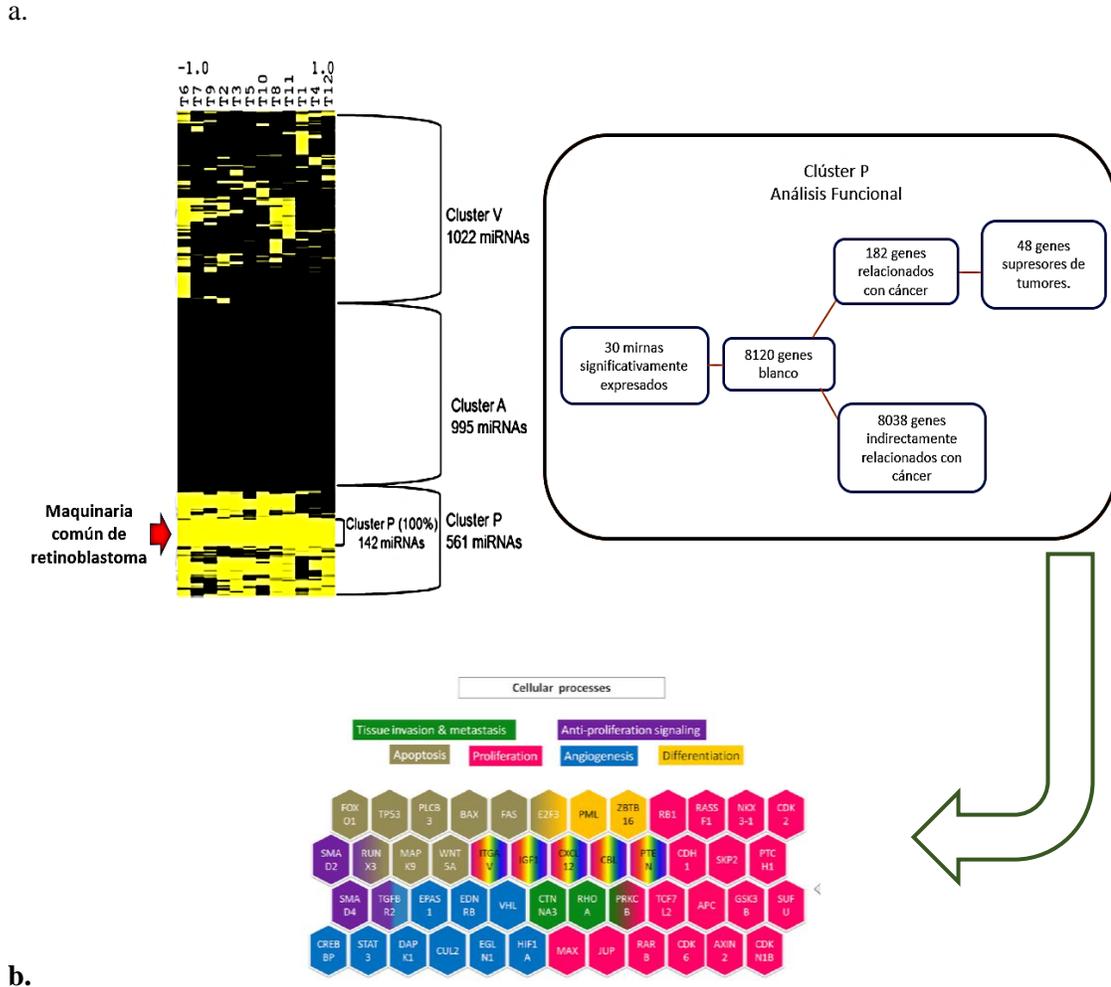


Figura 6. Agrupamiento Jerárquico y análisis de blancos del miRNOMA de RB a. Agrupamiento jerárquico de miRNAs detectados en 12 muestra de retinoblastoma. Se generan 3 grupos, Presentes, Ausentes y variables. (Color amarillo= miRNAs detectados, color negro= miRNAs no detectados) b. vías relevantes y procesos celulares probablemente afectados por el núcleo de 30 miRNA en Rb significativamente expresados.

Los resultados de este tipo de análisis en retinoblastoma permitieron determinar que las muestras de retinoblastoma comparten un perfil de expresión de miRNAs común a todas las muestras a pesar de su heterogeneidad. Este abordaje utilizó los puntajes de detección como una aproximación discretizada (encendido/apagado) al estado de expresión génica, y el agrupamiento jerárquico demostró que existen elementos compartidos entre todas las muestras

tanto en su estado expresado como no expresado. Este novedoso método de análisis brinda un nuevo modo de interpretar los datos de alta densidad que es plausible, consistente y clarifica de un modo intuitivo conceptos como la heterogeneidad tumoral asociándolo a elementos genéticos con una visión panorámica integrativa. Además, sin la necesidad de usar tejidos normales como controles.

2. PLANTEAMIENTO DEL PROBLEMA

Los microarreglos de expresión, en específico los de Affymetrix, ofrecen mediciones absolutas que no han sido explotadas en el campo del cáncer y la inmensa mayoría de análisis transcriptómicos se limitan a análisis diferenciales entre tejido tumoral y normal. Para ampliar los algoritmos de análisis de estas plataformas, se hace necesario aprovechar las sondas diseñadas para calibrar las señales inespecíficas o de fondo y analizar la información de detección que ofrece esta tecnología como una aproximación al estado de expresión de un gen. Este trabajo plantea un enfoque novedoso para obtener la expresión génica absoluta aplicando el concepto de detección de transcritos y este método se utilizará para determinar genes detectados (señal real) y no detectados (fluorescencia de fondo) en el transcriptoma. Esto permitirá conocer el tamaño real del transcriptoma que hasta el momento no se ha descrito, además de los componentes que son comunes a todas las muestras de Retinoblastoma en términos de detección.

3. JUSTIFICACIÓN E HIPÓTESIS

3.1 JUSTIFICACIÓN.

El empleo del puntaje de detección que ofrecen las plataformas de un canal de Affymetrix, permite identificar la expresión génica absoluta sin necesidad de un tejido control. De esta manera es de gran importancia la generación de un nuevo algoritmo para el análisis de datos de expresión que sea aplicable a diferentes muestras en este tipo de plataformas. Para el estudio de Retinoblastoma, es de gran interés la identificación de los elementos comunes entre un grupo de tumores tanto expresados como no expresados, así como los elementos heterogéneos. El análisis de la maquinaria común permite identificar potenciales vulnerabilidades de todos los tejidos estudiados y el análisis de los elementos variables abre la posibilidad de identificar elementos correspondientes a los rasgos de heterogeneidad tumoral que se observan en el área clínica. Además, este abordaje ofrece la posibilidad de definir con precisión el tamaño del transcriptoma en términos del número de genes o transcritos detectables.

3.2 HIPÓTESIS

La elaboración de un mapa transcriptómico en retinoblastoma con datos de detección (presentes/ausentes) usando agrupamiento jerárquico no supervisado, revelará un clúster medular de RNAm compartidos por todas las muestras, así como un grupo de RNAm no detectables también comunes y RNAm cuya detección es heterogénea.

4. OBJETIVOS

4.1 OBJETIVO GENERAL.

- Analizar el transcriptoma de retinoblastoma para determinar el grupo central de genes expresados y compartidos por todas las muestras, los genes no expresados comunes, así como los genes expresados de manera variable entre las muestras.

4.2 OBJETIVOS PARTICULARES

- 4.2.1** Desarrollar un algoritmo para asignar un puntaje de detección único a cada transcrito, a partir del puntaje de detección de cada exón que es posible obtener del chip.
- 4.2.2** Realizar análisis de agrupamiento jerárquico no supervisado con los datos de detección de genes con los datos de 8 retinoblastomas.
- 4.2.3** Realizar análisis de agrupamiento jerárquico no supervisado con los datos de detección de genes con los datos de 8 retinoblastomas y la línea celular Y79.
- 4.2.4** Identificar los procesos biológicos y rutas de señalización codificados en el grupo central de genes encontrado.
- 4.2.5** Validar mRNA(s) de interés por medio de RT-qPCR.

5. METODOLOGÍA

5.1 Características de las muestras.

Las muestras fueron obtenidas de 8 pacientes diagnosticados con Retinoblastoma que recibieron enucleación como parte de su tratamiento. Las muestras se recolectaron en el Hospital de Pediatría “Silvestre Frenk Freund”, del Centro Médico Nacional Siglo XXI del Instituto Mexicano de Seguro Social (IMSS) y en el Hospital Infantil de México Federico Gómez de la Secretaría de Salud. Se reclutaron los casos teniendo la aprobación del comité de ética de ambas instituciones y el consentimiento informado de los padres de familia de cada paciente.

La línea Celular Y79 (ATCC® HTB-18™) fue establecida a partir de una muestra tumoral de una niña de 2.5 años con retinoblastoma. La donante tenía una fuerte historia familiar materna de retinoblastoma. El medio base para esta línea celular es el medio RPMI (ATCC, número de catálogo 30-2001) complementado con SFB 12%.

Cultivo primario: Una parte del tejido tumoral se cultivó por 7 días en medio RPMI complementado con SFB 12%, posteriormente se realizó la extracción de RNA y conservación a -70°C hasta su utilización experimental.

Análisis Histopatológico: Como parte del procedimiento diagnóstico microscópico, el globo ocular se procesó para su estudio y análisis histopatológico. Este procedimiento incluye la fijación del espécimen en formol y el embebido en parafina. Se realizaron secciones en cortes de aproximadamente 3 micrómetros de grosor que fueron fijados en laminillas empleadas para la tinción Hematoxilina/Eosina (HE). La tinción HE permite el análisis histopatológico de cada muestra de RB (Fischer, Jacobson, Rose, & Zeller, 2008) y se realiza rutinariamente para determinar factores histopatológicos de riesgo de diseminación de la enfermedad.

El análisis histopatológico fue realizado por un patólogo pediatra (Dra. M. de Lourdes Cabrera). Los cambios histopatológicos se describieron de acuerdo con la clasificación de Liu Yeu (Y. Liu *et al.*, 2014), basado en el grado de diferenciación celular y este nivel de diferenciación se definió como: indiferenciado (sin presencia de rosetas o flueretes) y diferenciado (con presencia de Rosetas de Homer Wright y/o Flexner-Wintersteiner en <1/3 del tumor, que son aglomeraciones cuboides arregladas alrededor de una luz central,

patognomónicas de tumores bien diferenciados). Las características clínicas e histopatológicas de cada muestra se describen en la **tabla 1**.

Tabla 1. Características clínicas de las muestras.

Cod	Género	Lateralidad	Estadio clínico ^a	Edad (meses)	Diferenciación histológica ^b	ICM	Invasión Postlaminar	NO	Grado de Anaplasia.
RB01	F	B	II	9.5	Diferenciado RHW RFW	Neg	NV	NV	Neg
RB02	M	B	I	36	Indiferenciado	Neg	Neg	Neg	Severa
RB03	F	B	II	5	Diferenciado RHW RFW	Neg	Neg	Neg	Neg
RB04	M	B	II	6	Diferenciado RHW RFW	Neg	Neg	Neg	Neg
RB05	M	B	II	12	Indiferenciado	Neg	Neg	Neg	Severa
RB06	M	U	II	33	Diferenciado RHW RFW	Neg	Neg	Neg	Severa
RB07	F	U	II	10	Moderadamente Diferenciado RHW RFW	Neg	Neg	Neg	Severa
RB08	F	U	III	36	Diferenciado RHW RFW	Neg	SI	Neg	Neg

B = Bilateral, U = Unilateral, F = Femenino, M = Masculino.

a = Sistema de tinción de St Jude.

b = división RB según el grado de diferenciación (Y. Liu *et al.*, 2014)

RHW = rosetas de Homer-Wright,

RFW = rosetas Flexner-Winterstein.

ICM = invasión coroidea masiva.

INVASIÓN POSTLAMINAR= Invasión post laminar del nervio óptico.

NO = margen quirúrgico del nervio óptico.

NV = no valorable

5.2 Extracción de RNA

La extracción de RNA de los cultivos primarios fue realizada con el reactivo TRIzol siguiendo las instrucciones del fabricante (Invitrogen, cat15596026 CA, USA). La concentración del RNA fue cuantificada en un espectrofotómetro nanodrop (Thermo Fisher Scientific), que permite medir micro volúmenes de ácidos nucleicos y proteínas purificadas. El RNA extraído fue resuspendido en agua libre de RNAsas y almacenado a -70 °C hasta su uso en los microarreglos.

5.3 Hibridación en el microarreglo HTA2.0

La integridad de las muestras RIN (RNA Integrity Score) se evaluó en un equipo Bioanalyzer. Las ocho muestra de RNA de cultivos primarios de retinoblastoma y de la línea

celular Y79 fueron marcadas con biotina e hibridadas en el microarreglo Human Transcriptome Array HTA2.0 (Affymetrix, Santa Clara, CA), siguiendo las indicaciones del manual GeneChip™ WT PLUS Reagent Kit Manual Target Preparation for GeneChip™ Whole Transcript (WT) Expression Arrays (Thermofisher, 2018). La hibridación fue realizada en la unidad de microarreglos del Instituto Nacional de Medicina Genómica (INMEGEN). El diseño experimental se representa en la **Figura 7**.

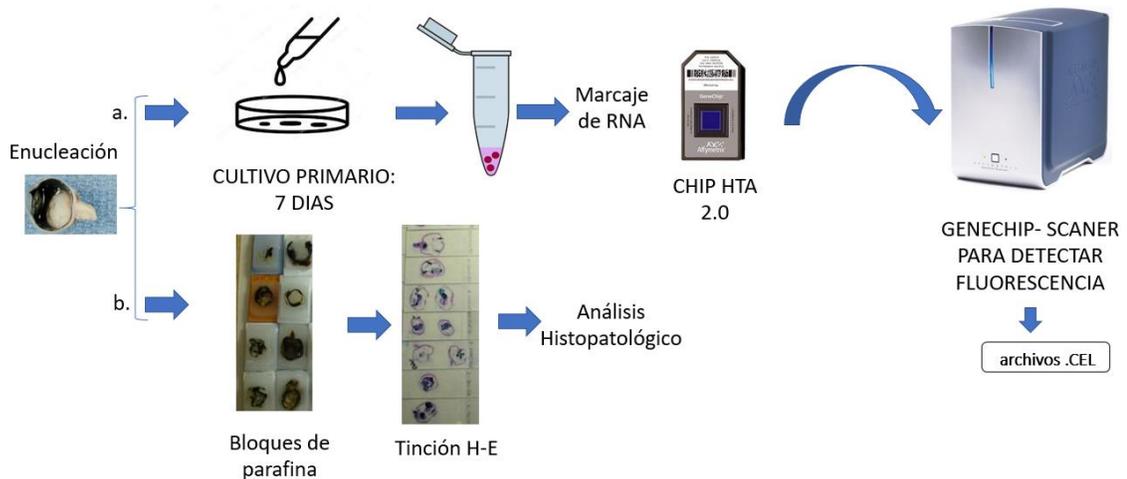


Figura 7. Diseño experimental. Se obtuvieron 8 muestra tumorales de ojos enucleados sin tratamiento previo. a. Análisis de RNA: Se cultivó la muestra por 7 días, se realizó extracción de RNA, marcaje de RNA con biotina, y la inyección de RNA marcado en microarreglo HTA2.0 (cada chip contiene una muestra) finalmente las muestras se analizaron en el equipo GENECHIP-SCANNER (Thermofisher. CA.) y se generaron los archivos de salida con la información de fluorescencia de cada sonda (archivos.CEL). b Análisis histopatológico: Cada muestra se fijó en formol y embebió en parafina y se realizaron cortes que fueron teñidos con HE y analizados por un patólogo pediatra

5.4. Análisis Bioinformático

El análisis bioinformático se realizó a través de las siguientes etapas:

5.4.1 Preprocesamiento: Los datos crudos (archivos .CEL) fueron preprocesados en el Software Expresión Console de la compañía Affymetrix. Este software toma los archivos de intensidad de fluorescencia (.CEL) y crea archivos numéricos de los niveles de expresión de las sondas (.CHP). Estos archivos pueden ser posteriormente analizados para identificar genes expresados diferencialmente en otros programas.

El preprocesamiento de los archivos .CEL incluyó la normalización que se realizó con el algoritmo set-rma-gene-full. Este algoritmo realiza corrección del fondo generado y la

transformación logarítmica de la señal (SST); también la normalización de sondas por medio de Robust Multichip Average (RMA). El recuento de citosina guanina y los algoritmos GCCN y SST, ajustan las intensidades del archivo .CEL . El algoritmo GCCN normaliza las intensidades con respecto a la diferencia en afinidad de las sondas con el contenido de guanina y citosina (GC). El SST estira la distribución de intensidad. Estos algoritmos permiten la comparación más directa de los resultados de expresión de microarreglos con los de otras plataformas (Okoniewski & Miller, 2008).

Características importantes del chip HTA2.0 para el diseño del algoritmo de detección

El microarreglo HTA2.0 cuenta con probesets que tienen sondas PM. (sondas diseñado para coincidir exactamente con la secuencia de interés). En este caso no cuentan con sondas MM (sondas con un desajuste de una sola base en la posición central) que detecten la unión inespecífica. En este microarreglo las sondas MM están reemplazadas por sondas con alta cantidad de C y G y se denominan sondas de fondo. En cuanto al diseño, los probesets PSR (del inglés Probe Selection Region) que cuentan con 10 sondas, se encuentran distribuidos a lo largo de exones de cada gene y los probesets JUC (del inglés Junction) que cuentan con 4 sondas, se pegan a uniones interexonales, permiten identificar específicamente los transcritos que se sustentan en las anotaciones así como transcritos nuevos (**Figura 8**). El diseño de este Chip está basado en las anotaciones de las bases de datos ENSEMBL y refSeq y la cantidad de sondas es mayor en comparación a plataformas previas de arreglos de exones.

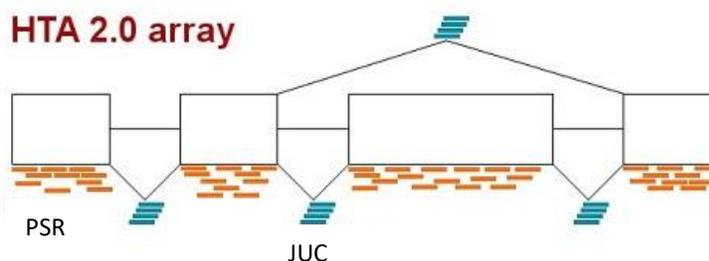


Figura 8. Diseño del microarreglo HTA2.0. El Diseño del chip HTA2.0 incluye probesets PSR (Color naranja) y probesets JUC (Color azul). Los probesets JUC permiten determinar variables de splicing en los transcritos. <https://www.thermofisher.com/>

El puntaje de detección de sondas PSR y JUC como detectados o no detectados fue calculado por el algoritmo Detection Above Background (DABG) o Detección por Encima del

Valor de Fondo (Affymetrix, 2005;Affymetrix, 2015) en la suite Expression Console. Este algoritmo compara la intensidad de cada sonda con la distribución de las sondas de fondo con la misma cantidad de G y C que están distribuidas en todo el chip HTA2. El valor de p obtenido determina la probabilidad de que la intensidad de la señal de cada sonda sea parte del ruido de fondo o no. Las sondas con valores bajos de p tienen señales que son significativamente distintas del fondo y, por lo tanto, se consideran **detectadas** por encima del fondo (DABG $p.value < 0.05$) (Clark *et al.*, 2007) (**Figura 9**). Este tratamiento de la señal constituye un calibrador y produce mediciones absolutas, lo que representa una diferencia tecnológica sustancial respecto a las plataformas de microarreglos originales de dos canales, que no contaban con calibradores y que producían mediciones relativas. El empleo de un calibrador permite definir un umbral de detección, así como la abundancia de moléculas de RNA en una solución. A partir de esta detección y su ponderación es posible inferir la expresión genética.

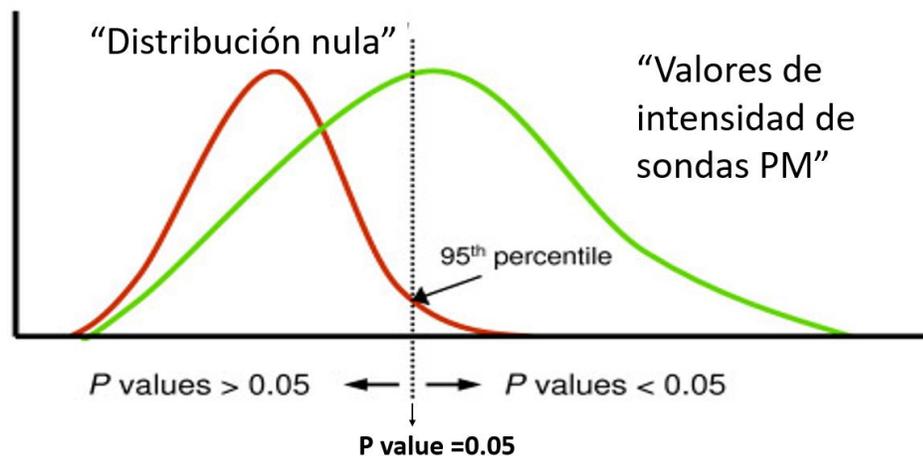


Figura 9. Determinación de la detección de probesets por encima del fondo (DABG). Para determinar si una señal de sonda dada se detecta por encima del fondo, la intensidad de las sondas PM (Perfect match) se compara con una distribución de sondas de fondo con el mismo contenido de G/C. Se calcula un valor p que representa la probabilidad de que la intensidad de la señal sea parte de la distribución nula. Las sondas con valores DABG con $p.value < 0,05$ se consideran detectadas por encima del fondo. En el diagrama se representa la comparación de la señal de fondo con la señal de PM. Las intensidades de señal de PM al 95% de las sondas de fondo con el mismo contenido de GC reciben un $p.value = 0.05$. Modificado de (Clark *et al.*, 2007)

5.4.2 Análisis de detección: Todos los probesets que se mapearon directamente con exones conocidos en la base de datos de Ensembl (Ensembl Genome Database)(Cunningham *et al.*, 2018) fueron incluidos en el análisis. Los probesets que se unieron a más de dos exones se eliminaron.

Detección de exones y transcritos: Para la integración e identificación de genes y transcritos se empleó la base de datos Ensembl que es uno de los principales recursos mundiales para la investigación genómica. En esta base de datos se puede obtener información completa del genoma humano, Ensembl también proporciona del transcriptoma muchos niveles de acceso a información basado en evidencia científica publicada. En esta base de datos se encuentra cada gene con la descripción de los respectivos transcritos descritos en publicaciones (identificación, localización cromosómica y número de exones) (**Figura 10**). En este trabajo se analizaron transcritos descritos previamente y no se tuvieron en cuenta transcritos novedosos ya que no cuentan con la posibilidad de análisis funcional.

Transcript: ABCD1-201 ENST00000218104.5 ← Ensembl gene id

Description: ATP binding cassette subfamily D member 1 [Source:HGNC Symbol;Acc:HGNC:616]

Gene Synonyms: ALD, ALDR, AMN, adrenoleukodystrophy

Location: Chromosome X: 153,724,868-153,744,782 forward strand.

About this transcript: This transcript has 10 exons, is annotated with 16 domains and features, is associated with 5017 variations and maps to 511 oligo probes.

Gene: This transcript is a product of gene ENSG00000101999 [Hide transcript table]

Name	Transcript ID	Gene	Biotype	CCDS	UniProt	RefSeq	Flags
ABCD1-201	ENST00000218104.5	ABCD1	Protein coding	CCDS14728.2	P30897.6	NM_000333.6 NP_000324.6	TSL:1 GENCODE basic APPRIS P1
ABCD1-202	ENST00000370128.4	ABCD1	222aa Protein coding	-	ARNEP9.6	-	TSL:2 GENCODE basic
ABCD1-203	ENST00000443694.2	ABCD1	No protein	-	-	-	TSL:3

Figura 10. Estructura de la base de datos ENSEMBL. La base de datos ofrece una interfaz amigable con información sobre el número de exones de cada transcritos. Para comparar la información con diferentes bases de datos existentes se cuenta con el código de identificación de cada gen (ENSG) y el código de cada transcritos de dicho gen (ENST) que son identificadores útiles para su posterior análisis funcional. Tomado de <https://www.ensembl.org/>.

Para este análisis, un exón fue considerado presente cuando al menos el 50% de sus probesets fueron detectados. Un transcritos se consideró detectado si el 100% de sus exones (PSR) y uniones interexonales (JUC) que los sustentan se encontraron detectados por el algoritmo DABG. (**Figura 11**).

Detección de genes: En el presente algoritmo, un gen fue considerado detectado o presente cuando uno o más de sus transcritos debidamente sustentados por Ensembl, fue detectado. Los datos categóricos de genes y transcritos (Detectados o no detectados) fueron agrupados por medio de agrupamiento jerárquico y graficados con el paquete gplots de R (Gregory R. *et. al.*,

2015; R Core Team, 2019). El análisis de detección se esquematiza en la siguiente figura de ejemplo:

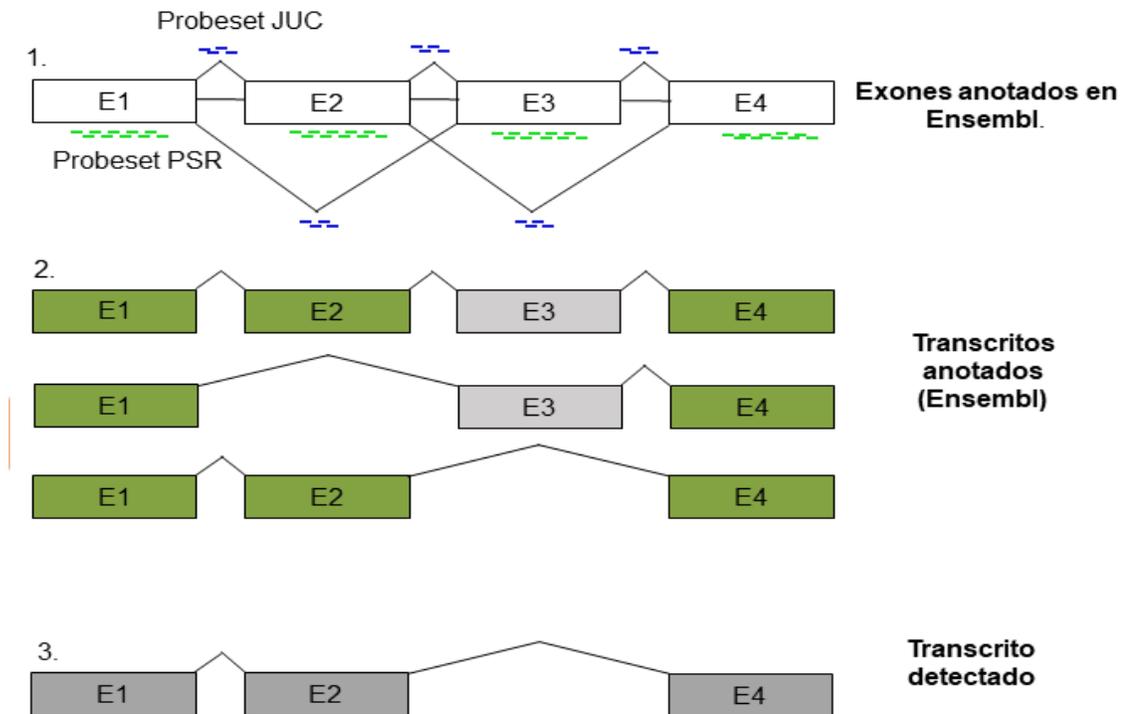


Figura 11. Ejemplo de análisis de detección. **1. Diseño del microarreglo HTA2.0.** El diseño del chip incluye probesets conformados por 10 sondas que hibridan con exones (PSR probesets = color verde) y probesets conformados por 4 sondas diseñados para hibridar con uniones interexonales (JUC probesets = color azul). **2. Transcritos anotados en Ensembl dentro del chip HTA2.0.** Se encuentran en este ejemplo, tres transcritos anotados por la base de datos Ensembl, con probesets para exones y las respectivas uniones interexonales. En este ejemplo, los exones detectados son rectángulos verdes y los no detectados son rectángulos de color gris. Todas las uniones interexonales se consideraron detectadas. **3. Transcritos detectados.** Un transcrito se considera detectado cuando el 100% de exones y uniones interexonales fueron detectadas por el algoritmo DABG. En este caso sólo uno de los transcritos cumplió la condición para considerarse detectado.

5.4.3. Análisis funcional de genes: Para interpretar los datos, el análisis funcional de los grupos de genes obtenidos por el agrupamiento jerárquico del análisis de detección, se realizó por medio de enriquecimiento funcional Overrepresentation Enrichment Analysis (ORA) en Webgestalt (Liao, Wang, Jaehnig, Shi, & Zhang, 2019). El análisis de enriquecimiento ORA es un método de análisis de enriquecimiento que mide la fracción de genes de interés (por ejemplo, genes detectados) que pertenecen a un grupo de genes en una determinada ruta

metabólica (genes de prueba), el grado de significancia de la superposición entre los genes de interés y el grupo de genes de prueba es evaluada mediante pruebas estadísticas, generalmente mediante pruebas de probabilidad hipergeométrica. En este análisis, las vías metabólicas que se superponen significativamente con el conjunto de genes de interés están relacionadas con algunos fenómenos biológicos que están asociados con estos genes (Tokar, Pastrello, & Jurisica, 2020).

Este tipo de análisis es ampliamente utilizado para determinar si las funciones o procesos biológicos conocidos están sobre representados o enriquecidos en una lista de genes. El valor de p (p -value) es calculado por la distribución hipergeométrica.

La aplicación de ORA requiere un conjunto de genes de interés y un conjunto de genes de referencia de rutas metabólicas o procesos biológicos a probar. En seguida se usa una prueba hipergeométrica para evaluar la significancia estadística. En el presente estudio las bases de datos de referencia fueron KEGG que está conformada por 296 rutas metabólicas con sus respectivos genes (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016) y Gene Ontology que cuenta con una gran fuente de información sobre la función de todos los genes (Carbon *et al.*, 2019). Para la representación de los mapas, se utilizó el programa Pathview (Weijun Luo and Cory Brouwer, 2013).

5.4.4 Expresión diferencial: A nivel de expresión, los transcritos expresados diferencialmente entre tumores unilaterales y bilaterales se identificaron mediante análisis de expresión diferencial. El análisis de expresión diferencial busca establecer si existen diferencias significativas entre la expresión de genes de una condición comparada con otro. Para este análisis se considera la estabilidad de la diferencial entre las muestras de ambos grupos y proporciona un valor de confianza estadística de esta diferencia. Este análisis se realizó comparando la expresión de mRNA del grupo de casos bilaterales en comparación con el grupo de casos unilaterales. Este análisis se realizó en el paquete de software R Bioconductor *limma* (Ritchie *et al.*, 2015), para visualizar los resultados se implementó el paquete *gplots* (Warnes *et al.*, 2015). El resultado del análisis arroja Log-fold change, expresión promedio, valor de p , valor de p ajustado y estadístico B. Los transcritos con fold change > 2 y valor p ajustado < 0.05 se consideraron expresados diferencialmente. El análisis funcional de los genes expresados diferencialmente se llevó a cabo a través del análisis ORA.

El flujo de trabajo se presenta en la **Figura 12**.

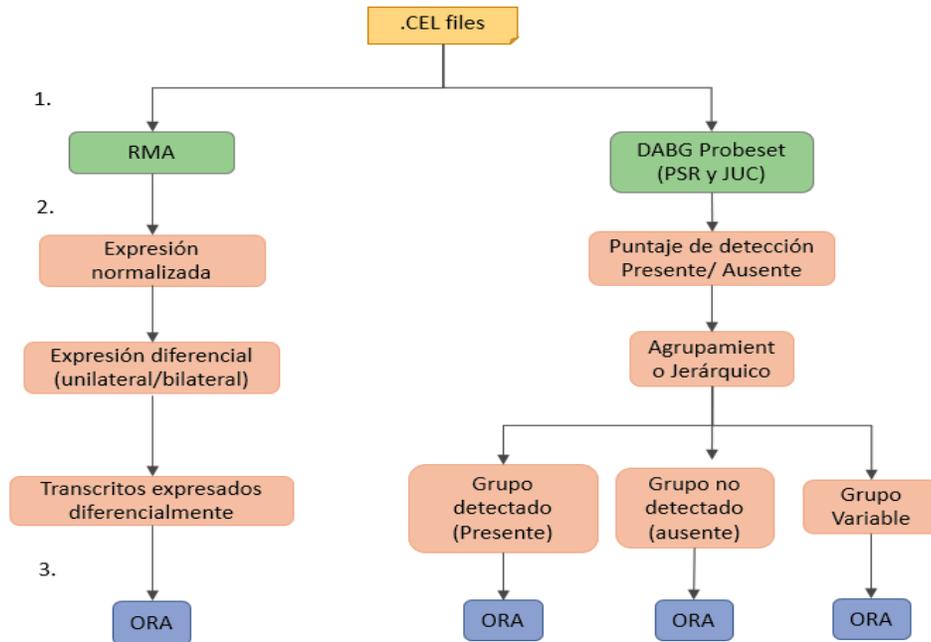


Figura 12. Análisis bioinformático. **a. Flujo de trabajo.** **1. Preprocesamiento de datos (cuadros verdes):** En esta etapa se incluye la normalización con RMA y detección de probesets con el algoritmo DABG. **2. Análisis estadístico (cuadros rosa):** La normalización de los datos de expresión se analizaron por medio de expresión diferencial entre muestras unilaterales y bilaterales. Los datos de detección fueron analizados con agrupamiento jerárquico. **3. Análisis de enriquecimiento.** Con el fin entender el contexto funcional de los diferentes sets de genes obtenidos, todos los grupos de genes presentes, variables y ausentes, además de genes diferencialmente expresados fueron enriquecido por separado con ORA.

5.5 Validación

5.5.1 Validación de núcleo de genes detectados con sets de datos publicados.

Como una estrategia de confirmación y para dar mayor sustento al “Núcleo central” de genes detectados en todas las muestras de retinoblastoma, se realizó un análisis de detección de transcritos en otro conjunto de datos de retinoblastoma. Este conjunto de datos se obtuvo de los resultados crudos de la publicación “*Gene expression profiling identifies different sub-types of retinoblastoma*” (Kapatai *et al.*, 2013). De este estudio se obtuvieron 18 muestras de retinoblastoma en el chip Hugene 1.0st. De Affymetrix. Este chip es la versión previa del chip HTA2 y se caracteriza por contar con sondas para la detección de exones, sin embargo, no cuenta con sondas para uniones interexonales. Por lo tanto, detecta isoformas de manera imprecisa y dificulta encontrar isoformas novedosas. El conjunto de datos fue analizado por medio de normalización con RMA y para la detección de exones se utilizó el algoritmo DABG.

En este caso un transcrito se consideró presente si todos sus exones se encontraron detectados en el análisis DABG. Como este chip no cuenta con sondas para uniones interexonales se omitió el análisis e identificación de diversos transcritos. Posteriormente, por medio de agrupamiento jerárquico, se obtuvo el grupo de genes compartidos por las 18 muestras de retinoblastoma y se compararon exclusivamente los genes en el grupo “Presente” de genes encontrados en ambos chips.

5.5.2 Validación de transcritos expresados diferencialmente por medio de qRT-PCR.

Los transcritos con mayor contraste (Fold Change) se seleccionaron para validarse experimentalmente. La validación de los genes expresados diferencialmente se realizó por medio de qRT-PCR. Se obtuvo mRNA de 8 muestras de RB unilateral y 8 muestras de retinoblastoma bilateral distintas a las analizadas por microarreglos. Se utilizó el kit Luna Universal One-step RT-qPCR Kit (New England BioLabs, UK) en el equipo applied Biosystems StepOnePlus de acuerdo con instrucciones del fabricante. La qRT-PCR se realizó por medio del Sistema SYBR-green con 300 ng of RNA y 0.4 μ M de iniciadores específicos para los genes problema (GUCA1C, PDC, PDE6C). El gen endógeno de referencia seleccionado fue HPRT y las diferencias fueron analizadas por medio del método de doble delta ($2^{-\Delta\Delta Ct}$).

El equipo utilizado en la PCR fue un StepOnePlus™ Real-Time PCR Systems y en el sistema de detección se utilizó el sistema SYBR green que es un colorante que se enlaza a la doble cadena sintetizada y produce fluorescencia. El protocolo de temperaturas que se siguió en la qRT-PCR se presenta en la **tabla 2**:

Tabla 2. Protocolo de temperaturas PCR.

Etapa	Temperatura	Tiempo	Ciclos
Transcripción Reversa	55°C*	10 minutos	1
Desnaturalización inicial	95°C	1 minuto	1
Desnaturalización	95°C	10 segundos	40-45
Extensión	60°C	30 segundos	

Determinación de la expresión de los genes problema. El análisis de Expresión de mRNA se realizó por cuantificación relativa, en este tipo de cuantificación se hace una comparación

cuantitativa del Ct del gen problema contra el control, que en este caso fue HPRT. El análisis se realizó por medio del método de $\Delta\Delta CT$ (Doble delta del threshold cycle o ciclo de umbral de detección), en este método comparativo se realiza una primera ΔCT que se calcula de la siguiente manera:

$$\Delta C_T = C_T \text{ gen de interés} - C_T \text{ gen endógeno}$$

C_T gen de interés = C_T de gen de interés (GUCA1C, PDC, PDE6C) de las muestras de la condición 1.

C_T gen endógeno = C_T gen de expresión constitutiva (HPRT), de las muestras de la condición 1.

$$\Delta\Delta C_T = \Delta C_T \text{ del tratamiento} - \Delta C_T \text{ control}$$

ΔC_T del tratamiento = C_T del gen de interés - C_T gen endógeno de cada una de las dos condiciones experimentales (Casos Unilaterales-Casos Bilaterales)

ΔC_T del calibrador = ΔC_T de las muestras control contra las cuales se desea comparar.

Curva de disociación. La reacción de PCR en tiempo real en la que se utiliza SYBR green debe validarse mediante la curva de disociación del producto generado durante los ciclos de amplificación y que debe presentar un único pico en la gráfica de disociación para garantizar la especificidad de los primers. En este método, la temperatura se incrementa gradualmente hasta 96°C, registrando la intensidad de la fluorescencia durante este proceso. Al incrementar la temperatura, las cadenas de DNA se desnaturalizan gradualmente, disminuyendo la intensidad de fluorescencia registrada. Al llegar a la T_m del amplicón, la velocidad de disociación se incrementa drásticamente dado que el 50% del producto de amplificación se encuentra desnaturalizado, la curva de disociación presentará un punto de inflexión a esta temperatura.

6. RESULTADOS

Para caracterizar el transcriptoma, se mapearon las sondas que se unen a los exones del genoma de referencia (GRCh37) (Harrow *et al.*, 2012). Los exones anotados se seleccionaron de la anotación Ensembl-Habana de todos los transcritos con evidencia conocida de un gen. En total, 465,070 exones correspondientes a 114,248 transcritos (19515 genes) fueron mapeados por el algoritmo. Para encontrar cada transcrito, El algoritmo mapeó según posición en el genoma, cada uno de los probesets PSR a su correspondiente exón y los probesets JUC a su correspondiente unión interexonal. La depuración de probesets PSR Y JUC se presenta en la siguiente figura (**Figura 13**).

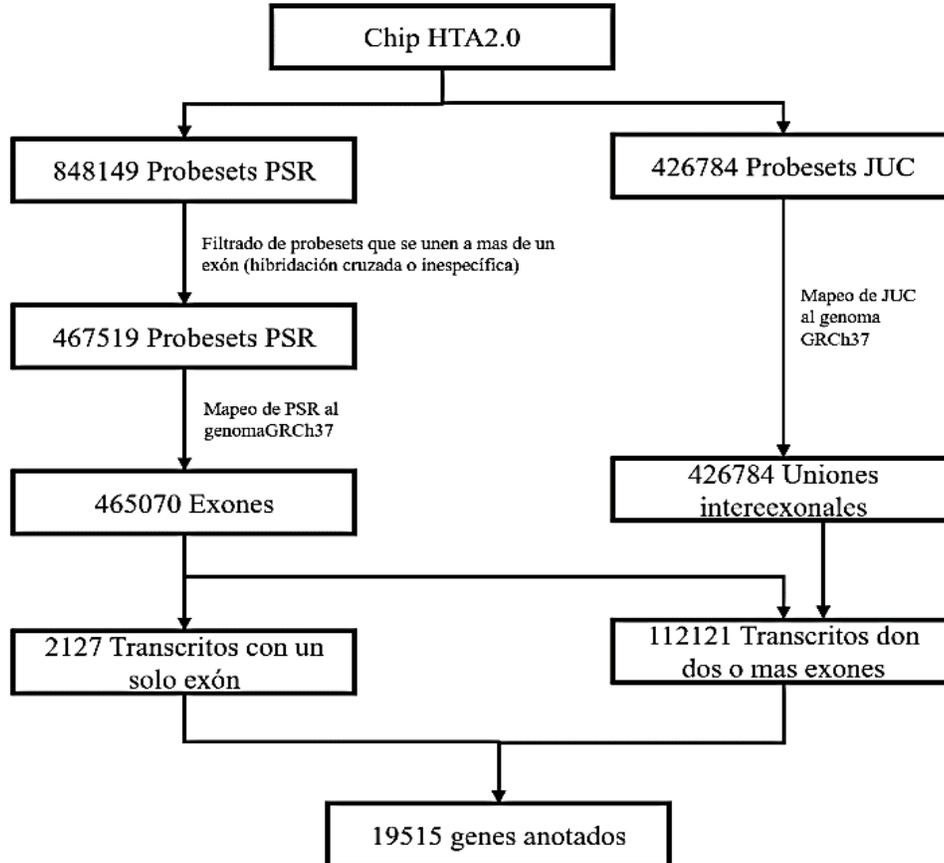


Figura 13. Depuración de probesets PSR y JUC. Se encontraron 467519 probesets PSR que mapearon a 465070 exones. Los probesets JUC se mapearon con 426784 uniones interexonales de transcritos de Ensembl. Estos exones y uniones interexonales corresponden a 114248 transcritos de 19515 genes anotados en el chip.

6.1 Tamaño del transcriptoma y estructura

Descripción del transcriptoma en términos de genes y transcritos detectados

En total se mapearon, identificaron y se retuvieron 19 515 genes que son aquellos genes que se encuentran en el chip HTA2 y que están sustentados con anotaciones en la base de datos ENSEMBL. Para el análisis de transcritos, se mapearon, identificaron y se retuvieron 114 248 transcritos anotados también en Ensembl y que se encuentran sustentados en su totalidad por probesets de exones y uniones interexonales del chip HTA2. Para facilitar la interpretación biológica y el enriquecimiento funcional se eliminaron los genes anotados en el chip como no codificantes como LNC RNA y miRNAs.

El tamaño del transcriptoma en términos de detección se determinó tanto a nivel de genes como de transcritos, de acuerdo a los genes y transcritos retenidos y descritos previamente. Es importante señalar que un gene se consideró detectado si por lo menos uno de sus transcritos fue detectado. Con este criterio, a nivel de genes activos transcripcionalmente, es decir detectados, se encontró en promedio 12 037 genes entre todas las muestras, esto corresponde al 61% de los genes anotados y retenidos del total (19 515 genes). A nivel de transcritos se realizó el mismo cálculo y se encontró en promedio 37 288 transcritos detectados entre todas las muestras, lo que corresponde al 32% de todos los transcritos anotados en el chip HTA2.0. Los valores mínimos y máximos del transcriptoma detectado se encuentran en la siguiente tabla:

Tabla 3. Tamaño del transcriptoma detectado.

	Genes detectados	Transcritos detectados
Promedio	12 037.2	37 288.6
Valor Mínimo detectado	9 292	18 438
Valor máximo detectado	13 776	46 896

En términos de la relación entre número de transcritos por gene del transcriptoma, se encontró por un lado que 5 873 genes cuentan con máximo dos transcritos anotados en el chip HTA2.0 y por otro lado hay un gene que tiene 66 transcritos anotados. Este análisis permitió

determinar cuáles transcritos de un gen se encuentran específicamente detectados en cada muestra, posteriormente se realizó el análisis funcional de estos genes y transcritos. La **Figura 14** presenta la gráfica de frecuencias de genes y la cantidad de transcritos encontrada.

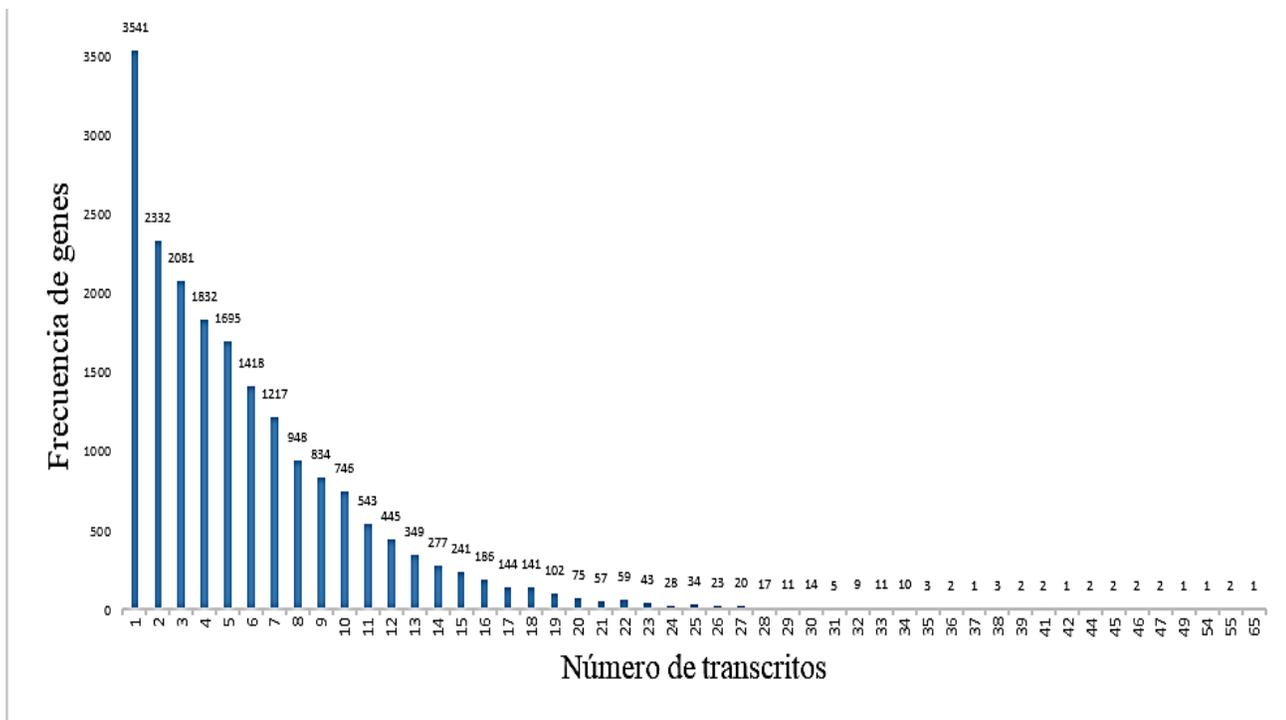


Figura 14. Número de transcritos por gen. Las barras indican la frecuencia de genes con diferente número de transcritos anotados. De tal manera, se observa al inicio de la gráfica que 3541 genes cuentan con un solo transcrito anotado y al final de la gráfica se observa que un solo gen cuenta con 65 transcritos. Eje x: indica la cantidad de transcritos anotados. Eje Y: indica la cantidad de genes para cada número de transcritos.

Grupos de genes detectados en las muestras: El análisis de detección, permitió encontrar cuales y cuantos genes se encuentran presentes y ausentes en cada muestra. Con los datos obtenidos de todas la muestras se realizó agrupamiento jerárquico y el resultado se graficó en un mapa de calor de dos colores. Este mapa está compuesto por los 19 515 genes retenidos, en color verde se representan los genes detectados y en negro los genes no detectados y constituye un perfil panorámico de detección transcriptómico (**Figura 15**).

Este agrupamiento jerárquico reveló una estructura de tres grupos o clusters de genes en las muestras de retinoblastoma. El primer grupo importante está compuesto por 7 513 (38%) genes detectados que son compartidos por todas las muestras, el segundo grupo tiene 4 321 (22%) genes que no se detectan en todas las muestras, el tercer grupo muestra variabilidad en la detección entre todas las muestras y está formado por 7 681 (40%) genes.

Cuando se realizó el agrupamiento jerárquico de las muestras de retinoblastoma junto con la línea celular Y79, los grupos encontrados, cambian en cuanto al número de genes en cada grupo o cluster descrito. En el primer grupo de genes compartidos por todas las muestras, se detectan 6,489 genes (33.3%), en el grupo de genes no detectados (o llamado ausente) se encontraron 4,294 genes (22%) y en el grupo variable se encontraron 8,732 genes (44.7%). En el caso de la línea celular Y79, cabe resaltar que este análisis encontró un grupo de 29 genes exclusivamente presentes en esta línea celular (ver listado en apéndices) y un grupo de 1024 genes que se encuentran exclusivamente ausentes en esta línea celular.

Los mapas de calor generados por este agrupamiento se representan en la siguiente figura:

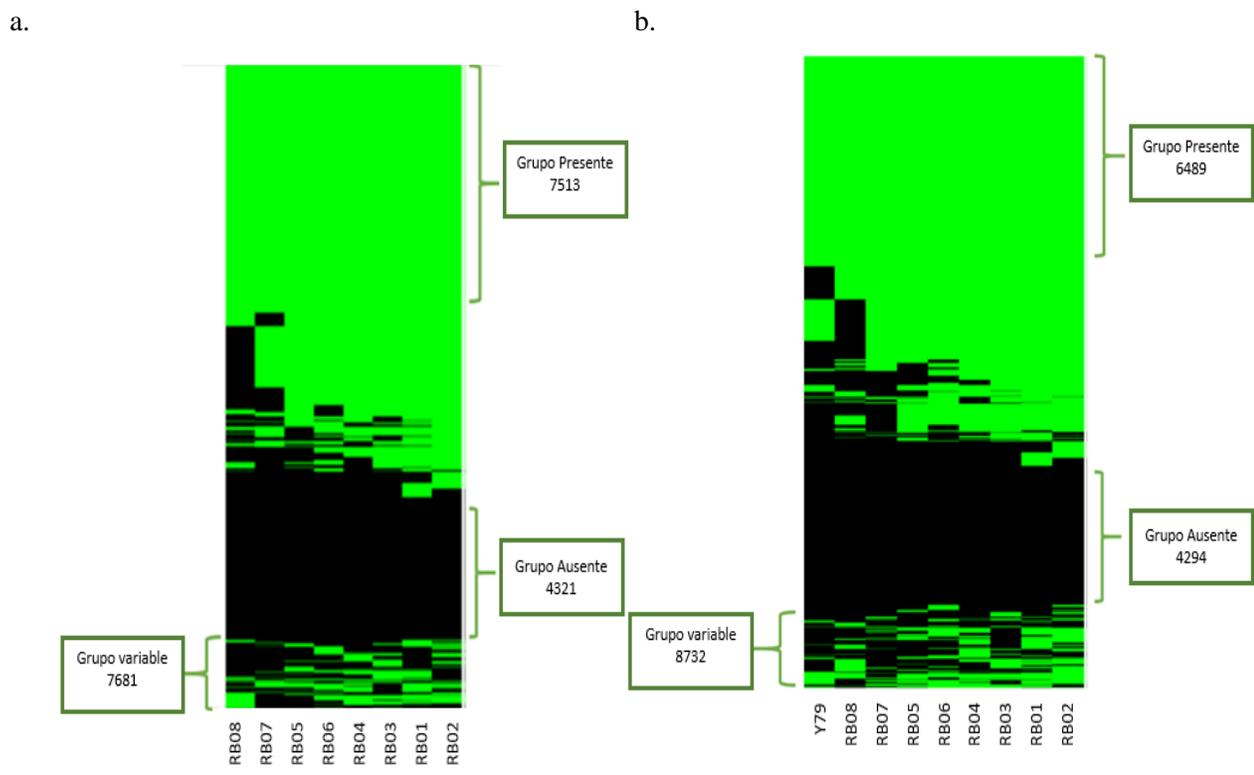


Figura 15. Mapa transcriptómico de detección. Heatmaps obtenidos de dos colores: color verde: genes detectados o presentes, color negro: genes no detectados o ausentes. a. Mapa transcriptómico con 8 muestras de retinoblastoma. b. Mapa transcriptómico 8 muestras de retinoblastoma y la línea celular Y79.

Descripción del transcriptoma en términos de transcritos detectados.

Se encontraron un total de 114,248 transcritos anotados que fueron graficados con Circosplot en relación al mapa transcriptómico de detección de genes (**Figura 16**, círculo exterior). De estos transcritos, 13,643 (11,8%) se detectan en todas las muestras, 55,886 (49,2%) no se detectan en ninguna de las muestras y 44,719 (39%) transcritos muestran variabilidad en la detección entre muestras. En el nivel de detección, un transcrito se consideró detectado o presente cuando el 100% de exones y uniones interexonales se detectaron.

En la siguiente figura se esquematizan de manera gráfica y general los genes detectados y sus transcritos correspondientes.

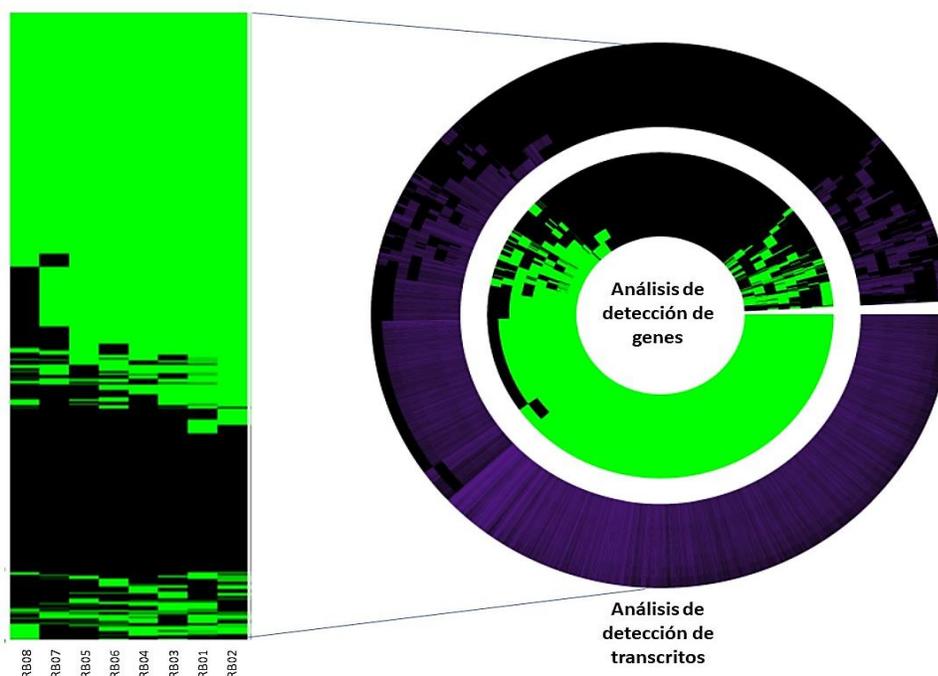


Figura 16. Mapas de detección de genes y transcritos correspondientes generados en Circosplot. a. mapa de calor de agrupamiento jerárquico de genes. b. mapa de calor circular. Representación de agrupamiento jerárquico de 8 muestras de retinoblastoma (filas) agrupadas por presencia y ausencia (verde: detectado, negro: no detectado) de genes. El círculo externo del panel b representa la proporción de transcritos en genes de la misma localización del mapa interno.

6. 2. Enriquecimiento funcional en grupos de muestras de retinoblastoma.

Análisis de enriquecimiento del grupo genes “presentes”. En el análisis de enriquecimiento empleando la base de datos KEGG, se identificaron 41 rutas metabólicas enriquecidas

significativamente en el core de “genes presentes”. Las vías enriquecidas están relacionadas con transcripción, traducción, crecimiento celular, metabolismo energético, y cáncer. En el grupo de genes detectados, un examen más detallado de las rutas de KEGG muestra que se enriquecen las rutas ciclo celular, regulación de la longevidad, la ruta de señalización de FoxO y la ruta de señalización de mTOR. Además, en este grupo, las vías metabólicas de glioma, el carcinoma de células renales, la carcinogénesis viral, el cáncer colorrectal y las vías del cáncer pancreático se enriquecieron significativamente. Estas rutas enriquecidas son de gran importancia porque están relacionadas con la proliferación. **GENEONTOLOGY (GO)**: El análisis de los procesos biológicos de GO reveló 45 categorías funcionales principales en el grupo de genes detectados. Las categorías en este grupo están relacionadas con la transcripción y traducción de proteínas, el procesamiento de RNA y el ciclo celular. Las 5 rutas mayormente enriquecidas (ordenadas por valor de False Discovery Rate FDR) se presentan en la **tabla 4**.

Tabla 4. Las 5 principales rutas enriquecidas en grupo de Core “genes presentes”.

Enriquecimiento KEGG				Enriquecimiento GeneOntology.			
Base	Descripción	genes	FDR	Base	Descripción	Genes	FDR
Rutas de Kegg	Espliceosoma	93	2.2E-10	GO_PB	Procesamiento de RNA	280	0
Rutas de Kegg	Ciclo celular	77	4.5E-07	GO_PB	Fase de transición mitótica en ciclo celular	275	0
Rutas de Kegg	Endocitosis	139	6.6E-07	GO_PB	Splicing de RNA	241	0
Rutas de Kegg	Transporte de RNA	101	6.6E-07	GO_PB	Localización de RNA	145	4.2E-13
Rutas de Kegg	Procesamiento de proteínas en retículo endoplasmático	93	6.2E-05	GO_PB	Procesos catabólicos de RNA	199	6.0E-13

GO_PB: Geneontology Procesos Biológicos.

Análisis de enriquecimiento “grupo genes ausentes”: KEGG. En el grupo de genes ausentes, las vías enriquecidas están involucradas principalmente en la adhesión focal y la migración. **GENEONTOLOGY:** En el grupo de genes no detectados se enriquecen categorías relacionadas con la comunicación extracelular, respuesta inmune e interacción de citocinas. Las 5 principales rutas enriquecidas se enumeran en la **tabla 5**.

Tabla 5. Las 5 principales rutas enriquecidas en grupo de “genes ausentes”.

Enriquecimiento KEGG (Base: Rutas de Kegg)			Enriquecimiento Geneontology. (Base: GO_PB)		
Descripción	Genes	FDR	Descripción	Genes	FDR
Regulación receptores a citocinas	135	0	Regulación positiva de producción de citocinas.	135	5.9E-08
Interacción Neuroactiva ligando receptor	97	1.1E-10	Organización de la estructura extracelular	127	8.3E-07
Metabolismo de retinol	34	2.9E-10	Respuesta inmune humoral	82	1.6E-05
Ruta de señalización JAK-STAT	60	2.4E-07	Regulación de transporte de iones en membrana	128	3.1E-05
Metabolismo de xenobióticos por citocromo P450	30	5.0E-06	Migración de leucocitos	125	3.1E-05

Análisis de enriquecimiento “grupo genes variables”: En el enriquecimiento funcional de los genes del grupo variable, los valores de FDR son mayores a 0.05 en prácticamente todos los casos, es decir que no son significativas. Sin embargo, en este grupo, la vía de Adhesión celular de moléculas se enriqueció significativamente. En la **tabla 6** se enlistan las funciones más enriquecidas de este grupo.

Tabla 6. Las 5 principales rutas enriquecidas en grupo de “genes variables”

Enriquecimiento KEGG (Base: Rutas de Kegg)			Enriquecimiento GeneOntology. (Base: GO_PB)		
Descripción	genes	FDR	Descripción	genes	FDR
Adhesión celular de moléculas (CAMs)	88	0.007	Localización de lisosoma	41	0.07
Interacción receptor citocinas	153	0.27	Activación de mastocitos	36	0.07
Interacción Neuro activa ligando receptor	142	0.27	Inmunidad mediada por mastocitos	29	0.29
Metabolismo de ácido araquidónico.	38	0.27	Transporte de cofactor	29	0.29
Ruta de señalización de calcio	99	0.27	Desarrollo de lengua	14	0.30

Enriquecimiento en línea celular Y79: En el grupo de genes presentes exclusivamente en la línea celular Y79 (29 genes), no se enriquecen funciones de manera significativa, en embargo, algunas funciones que se enriquecen están relacionadas con funciones de intercambio gaseoso

respiratorio, detección de estímulo químico sensorial, percepción del estímulo químico, detección de estímulos involucrados en la percepción sensorial.

En el grupo de genes no detectados (1024 genes) exclusivamente en la línea celular Y79, pocas rutas se enriquecieron de manera significativa. En la tabla 7 se enlistan las funciones más enriquecidas de este grupo.

Tabla 7. Enriquecimiento funcional de genes ausentes exclusivamente en la línea celular Y79.

Enriquecimiento KEGG (Base: Rutas de Kegg)			Enriquecimiento GeneOntology. (Base: GO_PB)		
Descripción	genes	FDR	Descripción	genes	FDR
Trasducción olfatoria	427	1.22E-11	Detección de estímulos involucrados en la percepción sensorial	470	1.47E-11
Cardiomiopatía dilatada	90	0.18	Detección de estímulos químicos	449	1.47E-11
Cardiomiopatía Hipertófica	82	0.18	Percepción sensorial del estímulo químico	464	1.45E-09
Miocardopatía arritmogénica del ventrículo derecho	72	0.18	Percepción sensorial del estímulo de luz	208	0.001
Metabolismo de ácido araquidónico	63	0.18	organización de sinapsis	379	0.070

6. 3 Búsqueda de posibles blancos terapéuticos.

Con el objetivo de mostrar la utilidad del presente análisis, buscamos algún tipo de vulnerabilidad farmacológica que se encuentre en todas las muestras estudiadas. Teniendo en cuenta que el grupo de genes presentes compartidos por todas las muestras ofrece la oportunidad de encontrar potenciales blancos terapéuticos que puedan ser accionables en todas las muestras, y que los inhibidores de cinasas han demostrado ser útiles en el tratamiento del cáncer, exploramos la presencia genes de cinasas en este grupo. Las proteínas cinasas promueven la proliferación celular, la supervivencia, la migración y la oncogénesis (Fabbro, Cowan-Jacob, & Moebitz, 2015). El quinoma (Manning, Whyte, Martinez, Hunter, & Sudarsanam, 2002) es el conjunto completo de genes de cinasas codificadas en el genoma. Estos genes son muy importantes ya que se ha descubierto que su sobreexpresión tiene un papel esencial en diferentes enfermedades, y estas enzimas han sido una fuente significativa de blancos terapéuticos para el tratamiento del cáncer. La FDA de EE. UU. (Marzo de 2019) ha

aprobado 43 inhibidores de proteínas cinasas que son moléculas pequeñas para el tratamiento del cáncer (Roskoski, 2019). Esta búsqueda de potenciales blancos terapéuticos consistió en analizar en el cluster de genes “Presentes”, cuantas cinasas de la base de datos del quinoma se encuentran expresadas en el grupo central de genes “Presentes” (7513 genes). El objetivo de este análisis fue identificar si entre los genes detectados en todas las muestras estudiadas contienen mRNA de posibles blancos farmacológicos en cáncer es decir su inhibidor asociado. Con este análisis se encontraron 148 genes cinasas (6 con inhibidor aprobado por la FDA). Las cinasas detectadas y los inhibidores correspondientes utilizados en diferentes enfermedades se muestran en la **Tabla 8**. Cabe resaltar que no se encontraron genes de cinasas en el grupo de genes exclusivos de la línea celular Y79.

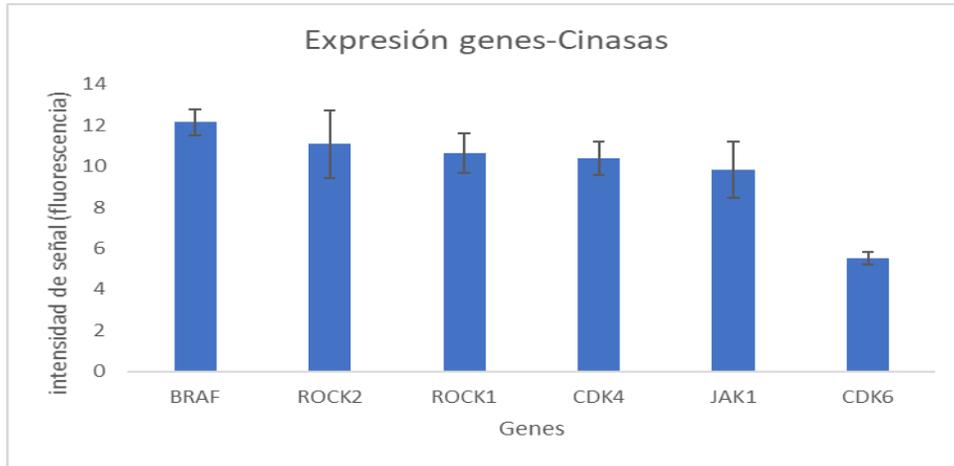
Tabla 8. Cinasas en el grupo de genes "presentes" e inhibidores aprobados por la FDA.

Genes de Quinasas	Inhibidor aprobado de FDA (Código)	Enfermedad	Año de aprobación
CDK4/6	Abemaciclib (LY2835219)	Terapia combinada y monoterapia para cáncer de mama.	2017
CDK4/6	Ribociclib (LEE011)	Terapia combinada y monoterapia para cáncer de mama.	2017
BRAF	Dabrafenib (GSK2118436)	Melanoma positivo a mutación en BRAF y carcinoma de pulmón de células no pequeñas	2013
BRAF	Encorafenib (LGX818)	Terapia combinada para mutaciones de BRAF ^{V600E/K} en melanoma	2018
BRAF	Vemurafenib (PLX-4032)	Mutación de BRAF ^{V600E} en melanomas	2011
ROCK1/ROCK2	Netarsudil (AR11324)	Glaucoma	2018
JAK1	Baricitinib (LY 3009104)	Artritis reumatoidea	2018
JAK1	Ruxolitinib (INCB-018424)	Mielofibrosis, policitemia vera	2011

Nivel de expresión de las Cinasas detectadas en el grupo de genes presentes.

Una vez identificadas estas cinasas, procedimos a evaluar los niveles de expresión correspondientes tanto de los genes como de los transcritos identificados. El valor promedio absoluto de expresión de cinasas indica que los genes de las cinasas BRAF presentan mayor nivel de expresión en todas las muestras de Rb. El gen que presenta menor nivel de expresión es CDK6. Al analizar los transcritos específicos expresados de este gen se observa que hay variabilidad en la expresión. (**Figura 17**)

a.



b.

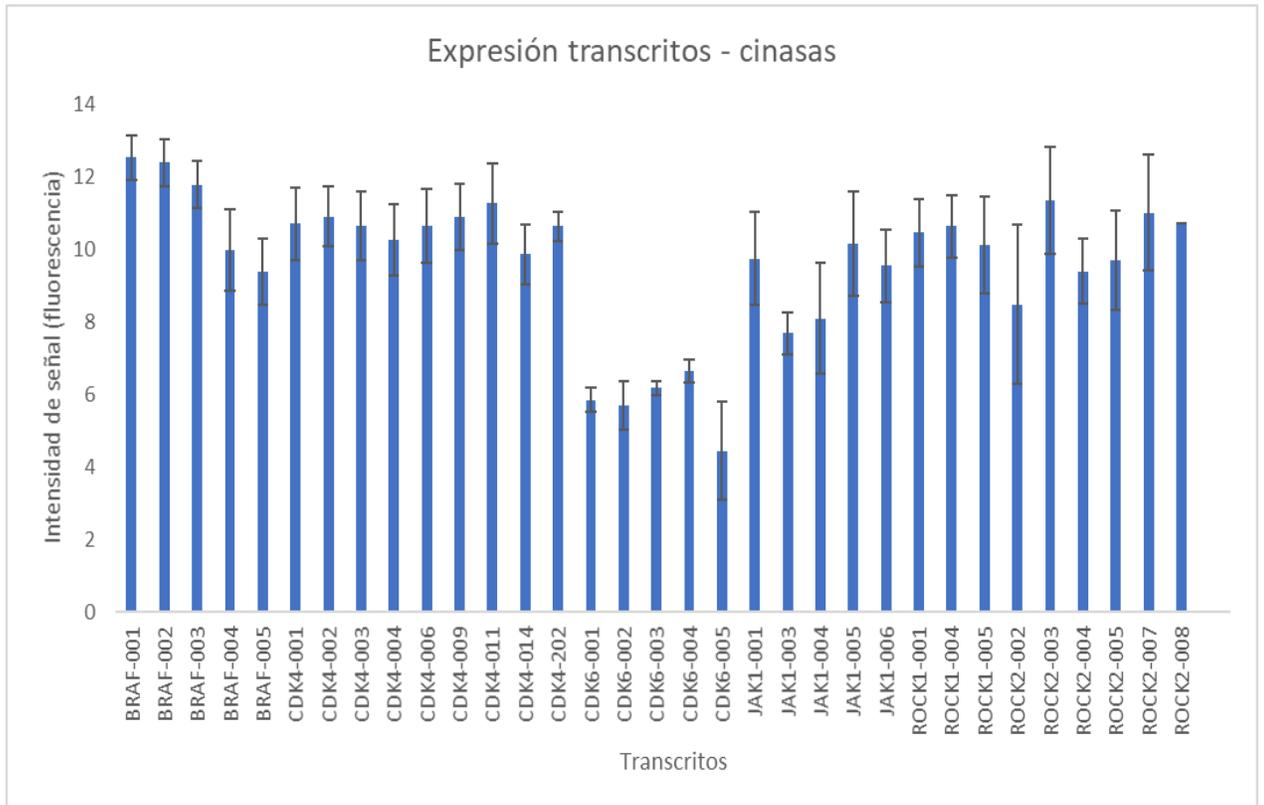


Figura 17. Nivel de expresión de Cinasas. El eje Y, representa la intensidad promedio de la fluorescencia en el microarreglo (8 muestras). El valor máximo detectado por el microarreglo es 19. El eje X en los paneles a y b representa los genes de cinasas y sus transcritos respectivos. a. Representa la intensidad de señal o expresión absoluta de los genes de cinasas en el grupo de genes presentes. B. Representa el nivel de expresión absoluta de los transcritos de los genes cinasas.

6.4 Hallazgos en el grupo de genes detectados de manera “Variable”.

En el análisis de detección de genes, se encontró un patrón de genes detectados y no detectados compartidos por todas las muestras de retinoblastoma, y también un grupo de genes variables. Una de las características clínicas que forma parte de la variabilidad entre estos tumores es la lateralidad. Para probar si existen diferencias en el perfil transcriptómico del retinoblastoma según la lateralidad, se realizó un análisis de expresión diferencial de todo el transcriptoma entre muestras bilaterales en comparación con muestras unilaterales.

El análisis de expresión diferencial empleando los valores de expresión de los genes no arrojó elementos diferenciales, sin embargo el análisis diferencial entre los transcritos arrojó 124 transcritos expresados diferencialmente entre las muestras de retinoblastoma unilaterales y bilaterales. 115 transcritos fueron regulados a la alta en el grupo de bilaterales y 9 regulados a la alta en el grupo de unilaterales. El agrupamiento jerárquico de los transcritos expresados diferencialmente muestra la separación correspondiente del grupo en dos ramas que separan las muestras unilaterales de las bilaterales (**Figura 18**).

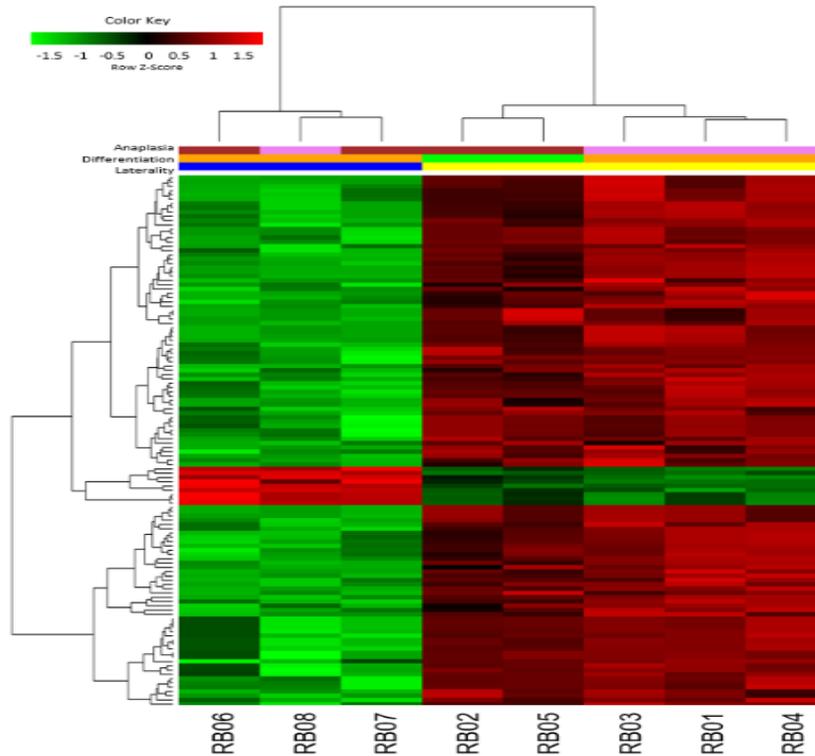


Figura 18. Expresión diferencial por lateralidad. Mapa de calor de 124 transcritos expresados diferencialmente, el dendrograma superior muestra la agrupación de las muestras en las dos grandes ramas correspondientes. Los transcritos sobreexpresados se encuentran en el mapa en rojo. Los transcritos subexpresados se representan en el mapa en color verde. Etiquetas de color: en la parte superior del mapa las etiquetas de color representan características de las lateralidad, diferenciación y anaplasia. Lateralidad: unilateral (azul), bilateral (amarillo), diferenciación celular: indiferenciada (verde), diferenciada (naranja), Anaplasia: anaplasia severa (marrón), sin anaplasia (violeta).

Enriquecimiento Funcional de transcritos expresados diferencialmente: Los 124 transcritos expresados diferencialmente corresponden a 52 genes (Ver apéndice: listado de transcritos expresados diferencialmente).

El análisis funcional de todos los transcritos diferenciales empleando GeneOntology muestra principalmente dos procesos biológicos enriquecidos: percepción visual y percepción sensorial del estímulo de luz, los genes que están relacionados con estas categorías son GUCA1C, GABRR2, PDC, PDE6C, EYS, MYO3B y KIFC3, los cuales se sobre expresan en las muestras bilaterales de este estudio. Las 10 principales categorías o funciones enriquecidas están relacionadas con la detección de estímulos externos y la organización celular y se encuentran en la siguiente tabla:

Tabla 9. Categorías Enriquecidas en transcritos expresados diferencialmente.

Base	Descripción	No de genes	FDR	Genes	Lateralidad
GO:0007601	Percepción visual	7	0.00573	GUCA1C, GABRR2, PDC, PDE6C, EYS, MYO3B, KIFC3	Bilateral
GO:0050953	Percepción sensorial de la luz	7	0.00573	GUCA1C, GABRR2, PDC, PDE6C, EYS, MYO3B, KIFC3	Bilateral
GO:0009583	Detección de estímulos de luz	4	0.11749	GUCA1C; PDC; PDE6C; EYS	Bilateral
GO:0008627	Vía intrínseca de la apoptosis en respuesta a estrés osmótico	2	0.36249	PTGS2; SCN2A	Bilateral
GO:0009581	Detección de estímulos externos	4	0.58566	GUCA1C; PDC; PDE6C; EYS	Bilateral
GO:0009582	Detección de estímulos abióticos	4	0.58566	GUCA1C; PDC; PDE6C; EYS	Bilateral
GO:0007602	Fototransducción	3	0.58566	GUCA1C; PDC; PDE6C	Bilateral
GO:0009584	Detección de luz visible	3	0.58566	GUCA1C; PDE6C; EYS	Bilateral
GO:0045216	Organización de uniones célula-célula	4	0.80329	CDH20; ANK2; FRMPD2; KIFC3	Bilateral

FDR: False Discovery Rate (Significativo: FDR < 0.05)

Para evaluar las funciones que se encuentran relacionadas con los transcritos sobre expresados específicamente en tumores bilaterales y unilaterales, se repitió el análisis, separando genes sobre expresados en cada uno de los grupos. En este análisis se encontró que los genes sobre expresados en tumores bilaterales están relacionados de manera significativa con percepción visual y a estímulos de la luz. En contraste los genes sobre expresados en tumores unilaterales se relacionaron con desarrollo embrionario y diferenciación a nivel sistémico como desarrollo de vías respiratorias o desarrollo gonadal. Las principales funciones enriquecidas en cada grupo se encuentran en las **Tablas 10 y 11**.

Tabla 10. Procesos biológicos enriquecidos en muestras unilaterales

Descripción	pValue	FDR	Genes
Desarrollo pulmonar	0.0016	1	ZFPM2;MAN2A1
Desarrollo de vías respiratorias	0.0016	1	ZFPM2;MAN2A1
Regulación negativa del Desarrollo gonadal femenino	0.0019	1	ZFPM2
Desarrollo de sistema respiratorio	0.0020	1	ZFPM2;MAN2A1

Tabla 11. Procesos Biológicos enriquecidos en muestras bilaterales.

Descripción	FDR	Genes
Percepción visual	0.0021	GUCA1C, GABRR2, PDC, PDE6C, EYS, MYO3B, KIFC3
Percepción sensorial a estímulos de luz	0.0021	GUCA1C, GABRR2, PDC, PDE6C, EYS, MYO3B, KIFC3
Detección de estímulos de luz	0.066	GUCA1C; PDC; PDE6C; EYS
Vía de señalización apoptótica intrínseca en respuesta al estrés osmótico	0.27	PTGS2; SCN2A
Detección de estímulos externos	0.38	GUCA1C; PDC; PDE6C; EYS

En retinoblastoma, a nivel Histopatológico se evalúa el grado de diferenciación y la presencia de anaplasia. Los tumores indiferenciados, considerados de alto riesgo, están compuestos por células pequeñas y redondas con núcleos hipercromáticos. Los tumores diferenciados, se caracterizan por presencia de rosetas de Homer-Wright y rosetas de Flexner-Wintersteiner, estas últimas indican una forma específica de diferenciación retinal. La presencia de anaplasia, caracterizada por células pleomórficas y núcleo hipercromático, se considera un indicador de alto riesgo en Retinoblastoma ya que indica alto grado de división celular. En este trabajo, se realizó una evaluación histopatológica minuciosa para explorar la relación de estos rasgos morfológicos con los hallazgos moleculares diferenciales entre tumores unilaterales y bilaterales.

Análisis histopatológico de las muestras: El análisis histopatológico de las muestras obtenidas de las muestras RB02 y RB01 reveló la ausencia de rosetas lo que indica que el tejido es indiferenciado, además presentan zonas de anaplasia severa. Las muestras RB04, RB03 y RB01 presentan rosetas, es decir que son tejidos bien diferenciados y no presentan áreas con anaplasia. Las muestras RB07 y RB06 están bien diferenciadas, pero tienen zonas de anaplasia. A nivel histopatológico la descripción de la anaplasia es de gran importancia porque la anaplasia celular fue identificada recientemente como un indicador histopatológico complementario a la diferenciación en la detección de retinoblastoma de alto riesgo (Mendoza *et al.*, 2015). Existen tejidos bien diferenciados que pueden ser de alto riesgo si presentan zonas con anaplasia. Las imágenes con la descripción histopatológica se encuentran en la **Figura 19**.

Cabe resaltar que estos hallazgos histopatológicos son parcialmente consistentes con los hallazgos funcionales, un ejemplo que indica consistencia entre los hallazgos histopatológicos y los hallazgos moleculares se da en las muestras RB02 y RB05, que histológicamente son indiferenciadas, ambas presentan zonas de anaplasia (**Figura 19**) y además se agrupan en la misma rama del heatmap de genes diferenciales (**Figura 18**). Esto no sucede con las muestras RB08 Y RB07, que histológicamente se describen como tumores bien diferenciados, sin embargo, la muestra RB07 presenta zonas de anaplasia y la muestra RB08 no (**Figura 19**) y molecularmente son similares, al agruparse en la misma rama del heatmap (**Figura 18**).

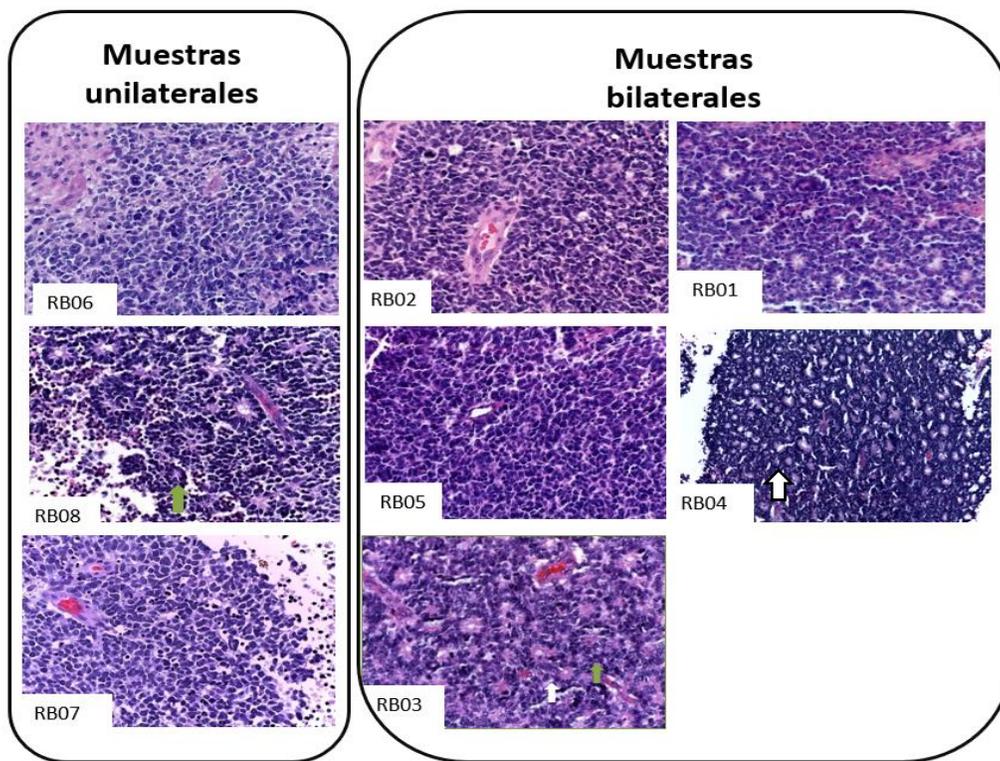


Figura 19. Análisis Histopatológico de las muestras. Tinción con hematoxilina eosina de secciones de retinoblastoma: las muestras RB08, RB03, RB04 muestran áreas diferenciadas con rosetas Wintersteiner (flecha blanca) y rosetas Homer Wright (flecha verde). Las muestras RB05, RB06, RB07, RB02 muestran anaplasia.

6.5 Validación bioinformática

Detección de genes presentes

El interés de esta validación fue evaluar si el grupo de genes “Presentes” coincide con el mismo grupo obtenidos en la plataforma previa de Affymetrix en muestras distintas de retinoblastoma; para realizar este análisis, se replicó la metodología de análisis de detección de probesets de exones en el chip HUGENE 1.0 ST (el chip previo al HTA2) en 18 muestras de retinoblastoma de pacientes reportados previamente (Kapatai *et al.*, 2013). El análisis se ajustó al diseño del chip. En este análisis se encontró un grupo de genes “Presentes” en todas las muestras conformado por 11881 genes, un grupo que no fue detectado o “Ausentes” conformado por 1100 genes y un grupo de genes “Variables” compuesto por 5545 genes. El diseño y naturaleza de las muestras generó que se presentaran cambios importantes en el número de genes detectados, a pesar de esto, cuando se realizó la comparación entre los dos grupos de genes presentes (del chip HTA2.0 y del chip HUGENE1.0St) se encontró que 5959 genes se encuentran compartidos entre ambos chips lo que corresponde a 85.3% de coincidencia. Esto es de gran importancia porque indica que estos genes están compartidos no sólo por las 8 muestras de retinoblastoma analizadas con en el microarreglo HTA2.0 sino también por las 18 muestras analizadas el chip HUGENE1.0 ST que es una plataforma previa y con menor cantidad de sondas. En la **Figura 20** se encuentra el mapa de detección generado y la cantidad de genes que coinciden en cada plataforma en el grupo de genes “presentes”.

a.

b.

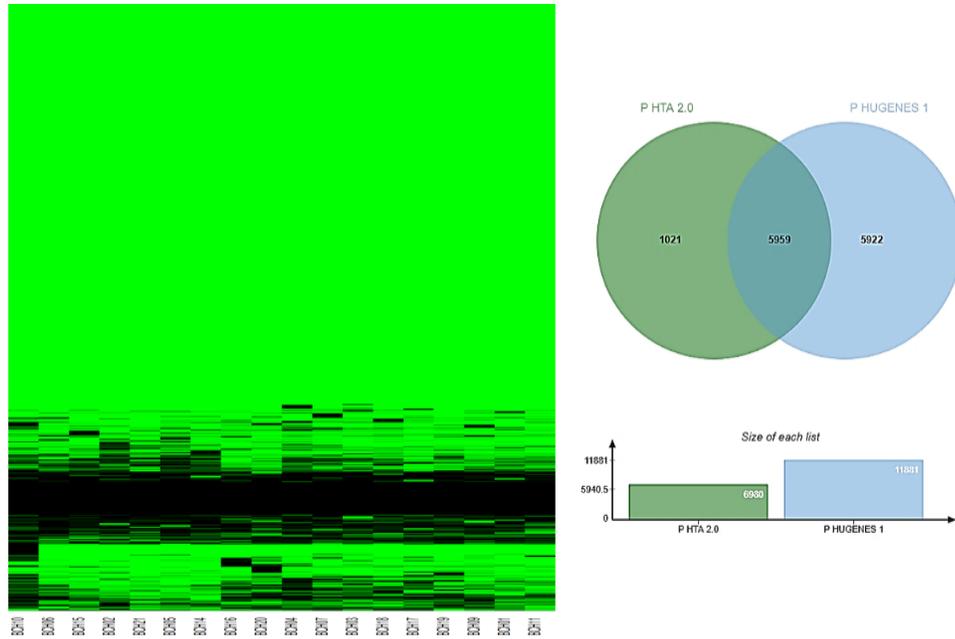


Figura 20. Análisis Detección de 18 casos de retinoblastoma con Plataforma Huguene1ST. a. Agrupamiento jerárquico de muestras. Cada fila representa un gen y cada columna representa una muestra. En total se graficaron 18 muestras. clave de colores: Color verde: detectado, color negro: no detectado. b. Diagrama de Venn del grupo del Core “presente” entre genes del chip HTA2.0 y el chip HUGENE 1.0 ST.

Cinastas en el grupo de genes “presentes”.

Para validar las cinastas detectadas en el grupo “presentes” empleando el chip HTA2 y que cuentan con inhibidor aprobado por la FDA, se buscaron los genes correspondientes en el grupo de genes “presentes” de las 18 muestras de Rb obtenidas por el grupo de Kapatai *et al.* (Kapatai *et al.*, 2013). Los genes de las cinastas CDK6, CDK4, BRAF, ROCK2 Y JAK1, se encuentran detectados en todas las muestras de RB. ROCK1 se encuentra presente en 17 de las 18 muestras (Figura 21).

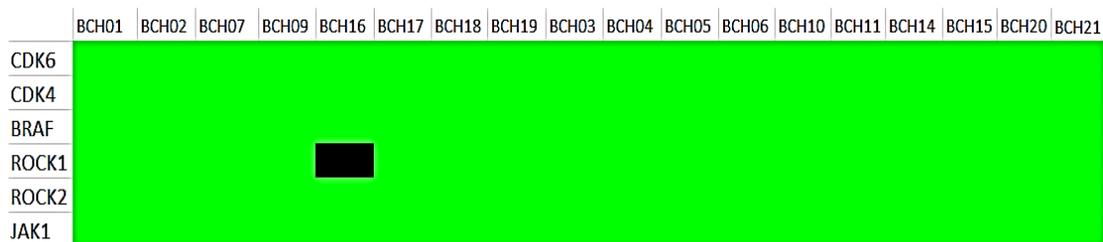


Figura 21. Heatmap del puntaje de detección cinastas en el grupo “presente”. En la gráfica se representa la detección de los genes (filas) para cada muestra (columnas). El color verde indica que el gen se encuentra detectado o “presente. El color negro indica que el gen se encuentra no detectado o “ausente”.

Validación experimental de genes “variables”.

qRT-PCR de genes expresados diferencialmente. Para la validación experimental de los resultados, se realizó qRT-PCR para los transcritos de los genes GUCA1C, PDC y PDE6C. Estos genes fueron seleccionados para la validación porque cuentan con FC >2 entre muestras bilaterales y unilaterales, es decir, se sobreexpresan en muestras bilaterales sumado a esto son genes importantes en la percepción visual. Los iniciadores diseñados para cada uno de los transcritos se presentan en la siguiente tabla:

Tabla 12. Primers para qRT-PCR

Gene Blanco	Primer-FW	Primer- RV	Trancrito
GUCA1C	GCCAGCAAACCTCTGAGTCCT	TCCCTTTGTAGGAAAAGTGATCCA	ENST00000393963
PDC	ACCATTGAAGAAAAGGACCCAA	TGAGGAGAAGACATTTGCCTGA	ENST00000497198
PDE6C	CCTAAGGCTTCATCACACCAGC	GCGCTTTGTGAAACTGTCGCTC	ENST00000371447

Expresión relativa de mRNA de GUCA1C por PCR tiempo real.

La expresión del mRNA de GUCA1C en retinoblastoma bilateral es significativamente mayor en comparación con las muestras unilaterales. Las muestras unilaterales presentan alta variabilidad (**Figura 22**).

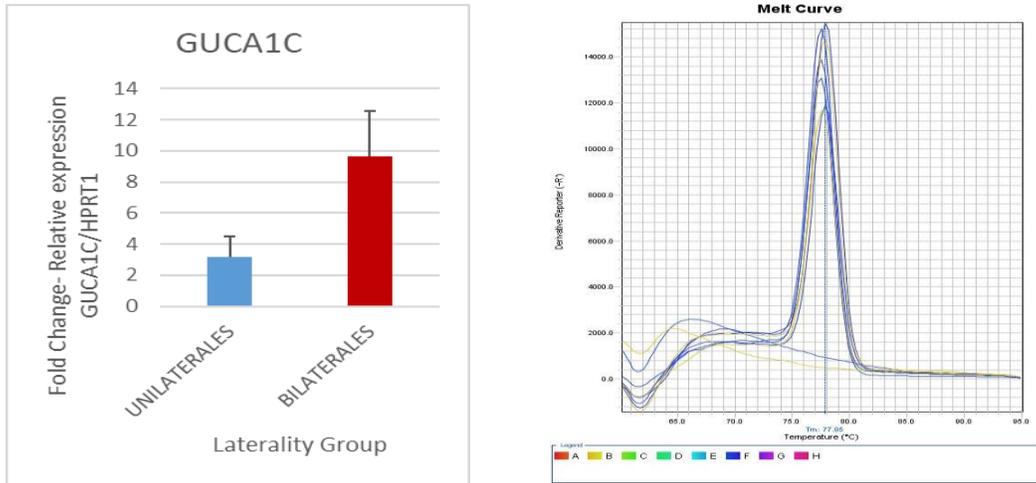


Figura 22. Expresión de mRNA de GUCA1C en Rb. Se presenta la expresión relativa del mensajero de GUAC1C en el Retinoblastoma Unilateral y bilateral, empleando como gen endógeno HPRT. En las barras se muestra promedio y barras de error, (n=8 en cada grupo). Curva de disociación de GUCA1C que muestra un producto único de amplificación.

Expresión relativa de mRNA de PDC por PCR tiempo real.

La expresión del mRNA de PDC en retinoblastoma bilateral es mayor en comparación con las muestras unilaterales. (Figura 23).

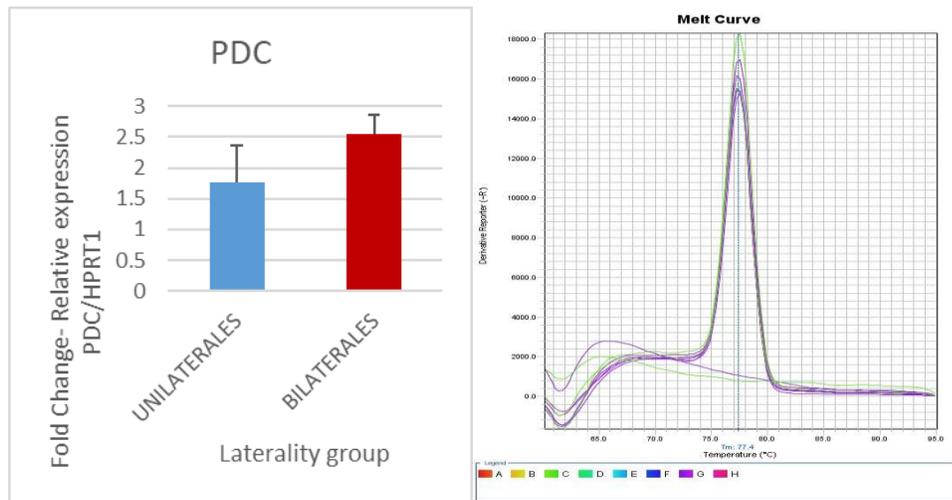


Figura 23. Expresión de mRNA de PDC en el Rb. Se presenta la expresión relativa del mensajero de PDC en el Retinoblastoma Unilateral y bilateral, empleando como gen endógeno HPRT. En las barras se muestra promedio y barras de error. (n=8). Curva de disociación de PDC.

Expresión relativa de mRNA de PDE6C por PCR tiempo real.

La expresión del mRNA de PDE6C en retinoblastoma bilateral es mayor en comparación con las muestras unilaterales pero la desviación es alta y hace que la diferencia no sea estadísticamente significativa. Las muestras unilaterales y bilaterales presentan alta variabilidad y dispersión (**Figura 24**).

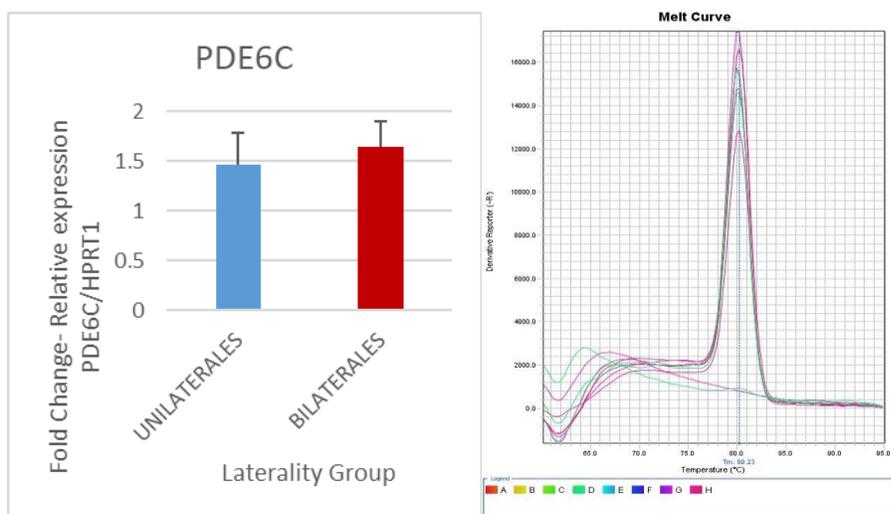


Figura 24. Expresión de mRNA de PDE6C en el Rb. Se presenta la expresión relativa del mensajero de PDE6C en el Retinoblastoma Unilateral y bilateral, empleando como gen endógeno HPRT. En las barras se muestra promedio y barras de error , (n=8). Curva de disociación de PDE6C.

7. DISCUSIÓN

Los microarreglos de expresión para análisis de transcriptoma se consideran una tecnología madura, cuyos resultados son comparables a otras tecnologías para el análisis de grandes datos como RNAseq (Mejía-Pedroza *et al.* 2018) que es una tecnología con la ventaja de examinar el nivel de expresión y la estructura fina del transcriptoma a partir de secuencias, sin embargo, los métodos de análisis son muy complejos. En comparación con RNAseq, los microarreglos cuentan con herramientas amigables para el análisis bioinformático, incluido el desarrollo de un puntaje de detección que constituye un concepto clave de este trabajo. Los microarreglos se han utilizado en la investigación clínica del cáncer por su capacidad de analizar de manera ágil miles de transcritos obteniendo resultados precisos para estimar los niveles de RNA (Romero *et al.*, 2018; Zhang *et al.*, 2015). Las necesidades a nivel clínico y la dificultad de obtener en muchos casos tejido normal para contrastar resultados, nos ha llevado a desarrollar este análisis para describir el perfil transcriptómico de Rb

Este trabajo es el primer acercamiento a un análisis de expresión absoluta de transcritos en términos de detección utilizando el microarreglo HTA2.0. En estudios previos del transcriptoma de Rb, Chakraborty (Chakraborty *et al.*, 2007) utilizó una plataforma doméstica con pocas sondas. Ganguly (Ganguly & Shields, 2010), McEvoy (McEvoy *et al.*, 2011) y Kooi *et al.* (Kooi *et al.*, 2015b) utilizaron microarreglos 3 TVT que tienen solamente sondas para el extremo 3' de la cadena de RNA. En otro estudio realizado por Kapatai *et al.* (Kapatai *et al.*, 2013) se realizó el análisis de expresión utilizando la plataforma Human Gene 1ST, que se caracteriza por ser la primera versión de microarreglos de exones de Affymetrix y la versión inmediata anterior al microarreglo HTA2 empleado en el presente trabajo. Cabe resaltar que Kapatai y Kooi, realizaron agrupamiento no supervisado de las muestras de Rb, y encontraron dos subgrupos que tienen funciones diferentes a nivel de diferenciación, percepción visual y fotoreceptores. Esto es similar a los resultados que encontramos en este estudio en el grupo de genes variables.

En este estudio, se desarrolló por medio de herramientas bioinformáticas, una estrategia novedosa para el análisis de datos transcriptómicos utilizando como referencial el puntaje de detección de sondas-transcritos-genes como primera aproximación de la expresión genética. Los microarreglos son plataformas muy sensibles, pero generan diversas y

considerables señales de fondo. El avance en el diseño de microarreglos ha llevado al desarrollo de plataformas de un canal y con estas plataformas se diseñaron puntajes de detección que pueden simplificar este ruido de fondo. El puntaje de detección es un procedimiento robusto pero subexplotado que se ignora en la mayoría de los análisis de expresión. Con el análisis de detección se reveló qué elementos se detectan y cuáles no se detectan con alta precisión, mostrando un enfoque práctico para aprovechar estas mediciones absolutas que proporcionan los microarreglos de un canal. Este tipo de análisis tiene diversas ventajas, hace que los tejidos control sean innecesarios, y permite la identificación de posibles blancos terapéuticos accionables en aquellos genes detectados comunes para todas las muestras estudiadas.

El puntaje de detección permite la determinación inequívoca del tamaño del transcriptoma definido como el número de genes expresados o no expresados. Es importante destacar que el tamaño preciso del transcriptoma no se ha reportado previamente para ninguna célula o tejido. Con este método se logró determinar el tamaño transcriptómico de manera numérica y descriptiva. Primero se obtuvo el número total de genes detectados / genes expresados. El procedimiento propuesto permite calificar como no detectados los genes que no cubren las condiciones del algoritmo de detección, aunque cuenten con señal de expresión. Esto evita considerar la expresión de falsos positivos que pueden emitir señal causada por ruido de fondo localizado en el chip. Posteriormente, se describió el tamaño transcriptómico en términos del número de genes expresados por grupos definidos por sus puntajes de detección a lo largo de las muestras que denominamos “presentes”, “ausentes” y “variable”. Esto abre la oportunidad a nuevos estudios en diferentes dimensiones del transcriptoma como podrían ser las relacionadas con la ubicación física (p. Ej., Ubicaciones cromosómicas) de genes en grupos presentes o ausentes, y la distribución espacial de estos loci de cromosomas dentro del núcleo, con esto a su vez abriendo nuevas formas de describir, comprender y modelar el transcriptoma.

El análisis funcional del grupo de genes “presentes” valida funciones centrales conocidas como la proliferación celular que es un proceso activo y compartida por todos los tumores malignos. La interpretación del análisis funcional en el grupo “variable” es más complicada. La heterogeneidad tumoral es una forma de variabilidad biológica, sin embargo, el grupo es pequeño y no incluyó suficientes casos de diferentes etapas oncológicas, por ejemplo, réplicas biológicas de diferentes etapas oncológicas. La presencia de funciones relacionadas con las células inmunes sugiere que el tumor puede estar adoptando programas

genéticos relacionados con la motilidad y la invasión, que son funciones normales en algunas células inmunes, o que el cultivo tumoral original tenía células inmunes. Otras fuentes de variabilidad pueden originarse en componentes oscilantes del ciclo celular. Estas células no estaban sincronizadas en el ciclo celular y, por lo tanto, podría ser posible encontrar diversos genes relacionados con el ciclo celular en los grupos detectados de forma variable. En el grupo de genes “no detectados”, el análisis funcional sugiere que la comunicación extracelular y procesos de diferenciación no están presentes en Rb. Esto es consistente con las células pequeñas e indiferenciadas que en su mayoría componen estos tumores.

El análisis Funcional del grupo de genes presentes y la búsqueda del quinoma en este grupo de genes detectados, se convierten en el primer paso hacia la posibilidad de reutilizar los inhibidores de quinasas ya aprobados como tratamientos alternativos en el retinoblastoma. Esto se debe a que las proteínas quinasas pueden promover la proliferación celular, la supervivencia, la migración y varios están involucrados en la oncogénesis. (Fabbro *et al.*, 2015). Estos genes son muy importantes ya que se ha descubierto que la sobreexpresión de algunos de ellos tiene funciones esenciales en diferentes enfermedades y constituye una fuente de blancos terapéuticos para el tratamiento del cáncer. La FDA de E. U., en marzo de 2019, aprobó 43 moléculas pequeñas que son inhibidores de proteínas quinasas para el tratamiento del cáncer (Roskoski, 2019). En retinoblastomas esto es particularmente relevante en tumores extraoculares avanzados, donde las opciones terapéuticas actuales están asociadas con importantes eventos de toxicidad y morbilidad, y a pequeños subgrupos de tumores pueden volverse refractarios a la terapia estándar y provocar mortalidad prematura. Los siguientes pasos necesarios incluyen validar si estas quinasas se traducen en proteínas en Rbs y si es diferente a la expresión proteica en tejido de retina normal. Además, el hecho de que 148 quinasas puedan estar presentes en todos los casos de Rb, indica que hay varios candidatos potenciales a considerar para una evaluación adicional como posibles blancos terapéuticos para los inhibidores existentes.

Al realizar el análisis de variabilidad exploramos y descubrimos que la lateralidad puede influir en la expresión genética. Es notable y sorprendente encontrar transcritos expresados diferencialmente entre Rb uni y bilateral en un número tan pequeño de muestras. Existen al menos dos tipos de retinoblastoma, unilateral y bilateral. El Rb bilateral es el modelo clínico más robusto de predisposición genética para desarrollar cáncer y siempre es hereditario,

sin embargo, la mayoría de Rb son unilaterales y no son hereditarios (Rodríguez-Galindo, Orbach, and VanderVeen 2015; Dimaras *et al.*, 2012). El análisis diferencial y funcional relacionado con la lateralidad puede ser informativo en el contexto del cáncer como un problema de desarrollo. Este análisis funcional reveló que los tumores unilaterales codifican funciones propias del desarrollo embrionario, mientras que los tumores bilaterales codifican las funciones de percepción visual y sensorial. Una forma de interpretar esta información es que la interrupción del desarrollo de la retina ocurrió en diferentes momentos del proceso de diferenciación retiniana; en una célula durante una etapa de desarrollo más primitiva en casos unilaterales, y en una célula más cercana a la etapa de diferenciación completa en casos bilaterales. Otra explicación de la existencia de transcritos expresados diferencialmente por lateralidad es que estas diferencias podrían ser causadas por la aparición de variantes constitutivas de RB1 en diferentes momentos de la embriogénesis, desafortunadamente no hay información disponible sobre las variantes de RB1. No obstante, se puede considerar que estos hallazgos de lateralidad son plausibles por las siguientes razones: el análisis de la lateralidad se ha pasado por alto en estudios transcriptómicos anteriores y, en segundo lugar, el transcriptoma es la capa funcional de los genes ya que representa todos genes expresados en una célula o tejido. y constituye las bases moleculares funcionales de los fenómenos biológicos, y estos hallazgos en estas muestras sugieren que la lateralidad de Rb es un fenómeno que puede tener bases moleculares. En este punto, estos hallazgos son iniciales y limitados, y pueden mejorarse con mayor conocimiento en las anotaciones genéticas, mejoras tecnológicas de microarreglos y madurez bioinformática. Independientemente de la explicación del origen de la expresión diferencial de la lateralidad, la posibilidad de utilizar la expresión transcripcional como un aproximamiento para identificar la lateralidad de los casos de Rb es ciertamente alentadora para su aplicabilidad e indica que vale la pena buscar más estudios y análisis dirigidos a probar esto.

Con respecto a los hallazgos transcriptómicos y las características histopatológicas, se encontró coincidencia en las ramas de agrupamiento jerárquico entre los casos bilaterales pero no en los unilaterales. Una posibilidad que puede explicar esto es que el examen histopatológico se basa en un número limitado de secciones que, según los criterios de diagnóstico estándar, deben incluir el nervio óptico (para una evaluación adecuada del riesgo

de diseminación metastásica). Sin embargo, dada la heterogeneidad del tumor, tales secciones histológicas podrían no incluir la porción más representativa del tumor.

El enfoque de este estudio es de gran utilidad para describir el panorama transcriptómico en tumores del sistema nervioso central para los que también es complicado obtener los tejidos normales correspondientes. La biología de sistemas también puede beneficiarse de este enfoque. El cáncer es un estado celular perturbado, y las células malignas como células normales también muestran plasticidad, la capacidad de cambiar en respuesta a los estímulos. El modelado funcional de genes activos en el grupo “presente” puede restringir el número de genes a utilizar en los esfuerzos por desarrollar modelos de redes genéticas, con el objetivo de predecir las respuestas de las células cancerosas y fortalecer el conocimiento que existe en el campo de análisis masivos de datos genómicos.

Por último, hay que reconocer las limitaciones presentadas en la validación experimental. Los resultados son consistentes en 2 de 3 genes validados no prueban ni refutan los hallazgos encontrados en este estudio. Sin embargo, las ideas principales de este trabajo se relacionan con la aplicabilidad del puntaje de detección, principalmente en aquellos casos en los que no se cuenta con tejido control, y el análisis funcional de los tres tipos de grupos de detección resultantes: genes “presentes”, “ausentes” y “variables”. Estos hallazgos abren nuevas posibilidades orientadas a explorar y encontrar relaciones entre el grupo variable y la lateralidad en Rb ya que tiene implicaciones importantes en el caso concreto de los casos unilaterales. Existe un número pequeño de casos unilaterales que con el tiempo desarrollan la enfermedad en el otro ojo, por lo que contar con una firma tumoral de lateralidad en principio, haría posible el desarrollo de una prueba de laboratorio que ayude a identificar estos casos unilaterales en riesgo de convertirse en bilaterales. Este trabajo abre un camino para validar estos resultados no sólo con lateralidad sino con más características clínicas propias de cada paciente y también propias de estadíos más avanzados de la enfermedad..

8. CONCLUSIÓN.

Este trabajo ofrece un nuevo enfoque para obtener resultados analíticos accionables utilizando tecnologías transcriptómicas de alto rendimiento que pueden aplicarse a otros tumores malignos y sistemas celulares mucho más allá del retinoblastoma.

9. PERSPECTIVAS

Este trabajo abre puertas a futuros estudios que se proponen a continuación:

- Analizar el proteoma de retinoblastoma, evaluando los transcritos encontrados presentes en todas las muestras.
- Evaluar si los transcritos encontrados de cinasas codifican para proteínas o no, y si estas proteínas contienen su dominio catalítico para proponer potenciales blancos terapéuticos accionables en el corto plazo.
- Utilizar el algoritmo para analizar diferentes muestras tumorales en el microarreglo HTA2.0 y determinar firmas moleculares en grandes cantidades de muestras tumorales.
- Analizar oncogenes en el grupo de genes detectados por todas las muestras y genes supresores de tumores en el grupo ausente o no detectado.
- Evaluar en los microarreglos HTA2.0 los datos de detección de miRNAs y genes no codificante que se obtienen de las anotación del Chip.

APÉNDICES

Comandos utilizados en RStudio para el análisis.

```
##-----
#ANOTACIONES DEL CHIP HTA2.0
##-----
library(data.table)
library(stringi)
library(Rgb)
library(dplyr)
transcripts<-fread("HTA-2_0.na36.hg19.transcript.csv",
                  skip=19,header=TRUE,sep=",")
transcripts$probeset_id<-gsub('.hg.1', '', transcripts$probeset_id)
transcripts=as.data.frame(transcripts)
transcriptData=transcripts[,c(2,4,5,6,7,8)]
transcriptData_ENST<-transcriptData[grep("ENST", transcriptData$gene_assignment),]
colnames(transcriptData_ENST)[1]<-"TC_ID"
TRANSCRIPTS<-stri_extract_all(transcriptData_ENST$gene_assignment, regex=("ENST[0-9]{11}"))
anot_tr_dataframe<-data.frame(TC_ID= rep(transcriptData_ENST$TC_ID,
sapply(TRANSCRIPTS,length)), ensembl_transcript_id= unlist(TRANSCRIPTS))
transc_ENST1<-anot_tr_dataframe[!duplicated(anot_tr_dataframe[c(1,2)]),]
transcriptData<-merge(transc_ENST1, transcriptData_ENST, by = "TC_ID", sort = FALSE)

##-----
## ANOTACIONES DE LOS PROBESET ID
##-----
probesets<- fread("HTA-2_0.na36.hg19.probeset.csv",
                 skip=14,header=TRUE,sep=",")
probesets <-as.data.frame(probesets)
Positiondata_probesets<-probesets[probesets[,13] == "main",c(1,2,4,5,6,9,21,22)]
Positiondata_probesets$probeset_id <- gsub('.hg.1', '',
Positiondata_probesets$probeset_id)
Probeset_to_TC_ID<-read.table("Transcript_clusterid_TO_probeset_HTA2.txt",sep = "\t",
header = FALSE)
colnames(Probeset_to_TC_ID)<-c("TC_ID", "probeset_id")
Positiondata_probesets_all<-merge(Probeset_to_TC_ID, Positiondata_probesets, by =
"probeset_id")
Positiondata_PSR<-Positiondata_probesets_all[grep("PSR",
Positiondata_probesets_all$probeset_id),]
Positiondata_JUC<-Positiondata_probesets_all[grep("JUC",
Positiondata_probesets_all$probeset_id),]

##-----
##OBTENER LOS GENES ASIGNADOS PARA CADA IDENTIFICADOR DE AFFYMETRIX.
##PRIMERO: ABRIR GTF GENOMA DE REFERENCIAL HG19.
##-----
gtf<-read.gtf("gencode.v19.annotation.gtf_withproteinids", attr = c("split", "intact",
"skip"), features = NULL, quiet = FALSE)
gtf_exon<-gtf[grep("exon", gtf$feature),]
gtf_exon1<-gtf_exon[grep("protein_coding", gtf_exon$gene_type),]

##-----
## ARCHIVOS QUE SE UTILIZAN PARA OBTENER LOS PROBESETS Y SU
##CORRESPONDIENTE EXON.
##-----
Positiondata_JUC1<- Positiondata_JUC[,c(1,3,8,9)]
Positiondata_PSR1<-Positiondata_PSR[,c(1,4,5,3)]
gtf_exon2<-gtf_exon1[,c(22,4,5,1,10,9)]

##-----
##DE LOS ARCHIVOS GTF_EXON2 Y POSITIONDATA_PSR SE GENERA UN ARCHIVO CON ##LOS EXONES Y
LOS CORRESPONDIENTES PROBESETS.
##-----
```

```

PSR_EXON1<-read.table("psr_exon.txt", sep = "\t", header = FALSE)
colnames(PSR_EXON1)<-c("probeset_id", "ensembl_exon_id", "ensembl_transcript_id",
"ensembl_gene_id", "chr_name", "psr_start", "psr_end",
"exon_chrom_start", "exon_chrom_end")
PSR_EXON3<-read.table("psr_exon1.txt", sep = "\t", header = FALSE)
colnames(PSR_EXON3)<-c("probeset_id", "ensembl_exon_id", "ensembl_transcript_id",
"ensembl_gene_id", "chr_name", "psr_start", "psr_end",
"exon_chrom_start", "exon_chrom_end")
PSR_EXON5<-read.table("psr_exon2.txt", sep = "\t", header = FALSE)
colnames(PSR_EXON5)<-c("probeset_id", "ensembl_exon_id", "ensembl_transcript_id",
"ensembl_gene_id", "chr_name", "psr_start", "psr_end",
"exon_chrom_start", "exon_chrom_end")

##-----
##UNIR LOS DATOS
##-----
PSR_EXON_ALL<-rbind(PSR_EXON1, PSR_EXON3, PSR_EXON5)
PSR_EXON_ALL<-
PSR_EXON_ALL[order(PSR_EXON_ALL$chr_name,PSR_EXON_ALL$ensembl_transcript_id,
PSR_EXON_ALL$exon_chrom_start, PSR_EXON_ALL$psr_start),]
PSR_EXON<-PSR_EXON_ALL

##-----
## SE BUSCAN LAS UNIONES DE CADA EXON PARA DETERMINAR SI EL
##TRANSCRITO ESTÁ COMPLETO Y SE GENERA UN DOCUMENTO CON LOS DATOS
##UNICOS DE EXONES Y SU CORRESPONDIENTE TRANSCRITO
##-----
PSR_EXON_UNIQUE<-PSR_EXON[!duplicated(PSR_EXON[c(2,3)]),]
PSR_EXON_UNIQUE$ensembl_transcript_id<-
as.character(PSR_EXON_UNIQUE$ensembl_transcript_id)
PSR_EXON_UNIQUE$ensembl_gene_id<-as.character(PSR_EXON_UNIQUE$ensembl_gene_id)

##-----
## ORDENAR LOS EXONES DE CADA TRANSCRITO
##-----
PSR_EXON_SORT<-
PSR_EXON_UNIQUE[order(PSR_EXON_UNIQUE$chr_name,PSR_EXON_UNIQUE$ensembl_transcript_id,
PSR_EXON_UNIQUE$ensembl_gene_id, PSR_EXON_UNIQUE$exon_chrom_start),]

##-----
## EDICION DE PSR_EXON_SORT PARA OBTENER LOS TRANSCRITOS QUE TIENEN UN EXON
##-----
TR_ONLY_ONE_EXON_CANDIDATE<-PSR_EXON_SORT [%>%
  group_by(ensembl_transcript_id) [%>%
    filter(n() == 1)
PSR_EXON_SORT<-PSR_EXON_SORT[which(!PSR_EXON_SORT$ensembl_transcript_id %in%
TR_ONLY_ONE_EXON_CANDIDATE$ensembl_transcript_id),]

##-----
## EDICION DE PSR_EXON_UNIQUE PARA OBTENER LOS EXONES CONTINUOS
##-----

PSR_EXON_UNIQUE1<-PSR_EXON_SORT[-1,c(2,3,4,8,9)]

##-----
##SE ADICINA UNA FILA VACIA AL NUEVO DATAFRAME
##-----

PSR_EXON_UNIQUE1[nrow(PSR_EXON_UNIQUE1)+1,] <- NA

##-----
##CONCATENAR LAS DOS TABLAS Y EDITAR NOMBRES
##-----

TR_EXON_JUN<- cbind(PSR_EXON_SORT,PSR_EXON_UNIQUE1)
TR_EXON_JUN<- TR_EXON_JUN[,c(5,4,3,2,8,9,12,11,10,13,14)]
colnames(TR_EXON_JUN)<-
c("chr_name", "ensembl_gene_id", "ensembl_transcript_id", "start_exon_id", "start_exon_sta

```

```

rt","start_exon_end","ensembl_gene_id.1","ensembl_transcript_id.1","stop_exon_id","sto
p_exon_start","stop_exon_end")

##-----
## SE SELECCIONAN LOS EXONES CONTINUOS PARA OBTENER LOS PROBESETS DE ##UNIONES
##-----

TR_EXON_JUN$test <- ifelse(TR_EXON_JUN$ensembl_transcript_id ==
TR_EXON_JUN$ensembl_transcript_id.1,"same","diff")
newdataset_clean0 <- head(TR_EXON_JUN,-1)
TR_CANDIDATE_SPLICING<- newdataset_clean0 %>% filter(test == "same")
dim(TR_CANDIDATE_SPLICING)
length(unique(TR_CANDIDATE_SPLICING$ensembl_transcript_id))
colnames(Positiondata_JUC1)[2]<-"chr_name"
Positiondata_JUC1<-Positiondata_JUC1[which(Positiondata_JUC1$chr_name %in%
TR_CANDIDATE_SPLICING$chr_name),]
Positiondata_JUC1$probeset_id<-as.character(Positiondata_JUC1$probeset_id)
Positiondata_JUC1$junction_start_edge<-
as.integer(Positiondata_JUC1$junction_start_edge)
Positiondata_JUC1$junction_stop_edge<-as.integer(Positiondata_JUC1$junction_stop_edge)

JUN_FUN<-function(x, JUC_DF, tr_chrl){
  exon<-tr_chrl[x,]
  probeset_seq<-subset(JUC_DF, chr_name == exon$chr_name)
  probeset<-probeset_seq$probeset_id[probeset_seq$junction_start_edge>=
  exon$start_exon_start
  & exon$start_exon_end >=
  probeset_seq$junction_start_edge
  & exon$stop_exon_start <=
  probeset_seq$junction_stop_edge
  & exon$stop_exon_end >=
  probeset_seq$junction_stop_edge]
  if (length(probeset) == 0){
    result <- NA
  }
  else{
    result<-as.character(probeset)
  }
  result
}
exon_all_list<-lapply(seq(1:nrow(TR_CANDIDATE_SPLICING)),JUN_FUN, tr_chrl =
TR_CANDIDATE_SPLICING, JUC_DF=Positiondata_JUC1 )

df_exon_all_start<-data.frame(start_exon_id= rep(TR_CANDIDATE_SPLICING$start_exon_id,
sapply(exon_all_list,length)), JUN_ID= unlist(exon_all_list))
df_exon_all_start<-df_exon_all_start[!is.na(df_exon_all_start$JUN_ID),]
df_exon_all_stop<-data.frame(stop_exon_id= rep(TR_CANDIDATE_SPLICING$stop_exon_id,
sapply(exon_all_list,length)), JUN_ID= unlist(exon_all_list))
df_exon_all_stop<-df_exon_all_stop[!is.na(df_exon_all_stop$JUN_ID),]
df_exon_start_uniq<-df_exon_all_start[!duplicated(df_exon_all_start[c(1,2)]),]
df_exon_stop_uniq<-df_exon_all_stop [!duplicated(df_exon_all_stop[c(1,2)]),]
df_junctions<-merge(df_exon_start_uniq, df_exon_stop_uniq, by = "JUN_ID", sort =
FALSE)
colnames(Positiondata_JUC1)[1]<-"JUN_ID"
df_junctions1<-merge(df_junctions, Positiondata_JUC1, by = "JUN_ID", sort = FALSE)
colnames(df_junctions1)<-c("JUN_ID", "start_exon_id", "stop_JUN_exon_id","chr_name",
"junction_start_edge","junction_stop_edge")
df_junctions<-df_junctions1[,-4]
PSR_EXON_A<-read.table("psr_exon1.txt",sep = "\t", header = FALSE)
colnames(PSR_EXON_A)<-c("probeset_id", "ensembl_exon_id", "ensembl_transcript_id",
"ensembl_gene_id","chr_name", "psr_start", "psr_end",
"exon_chrom_start","exon_chrom_end")
PSR_EXON_B<-read.table("psr_exon2.txt",sep = "\t", header = FALSE)
colnames(PSR_EXON_B)<-c("probeset_id", "ensembl_exon_id", "ensembl_transcript_id",
"ensembl_gene_id","chr_name", "psr_start", "psr_end",
"exon_chrom_start","exon_chrom_end")
PSR_EXON_AB<-rbind(PSR_EXON_A, PSR_EXON_B)
PSR_EXON_AB<-PSR_EXON_AB[order(PSR_EXON_AB$chr_name,PSR_EXON_AB$ensembl_transcript_id,
PSR_EXON_AB$exon_chrom_start, PSR_EXON_AB$psr_start),]

```

```

PSR_JUC_START<-read.table("psr_juc_edge.txt",sep = "\t", header = FALSE)
colnames(PSR_JUC_START)<-c("juc_id", "probeset_id", "seqname", "juc_st",
"juc_stop","psr_st", "psr_stop")
PSR_JUC_START1<-merge(PSR_EXON_AB, PSR_JUC_START, by= "probeset_id", sort=FALSE)
PSR_JUC_STOP<-read.table("psr_juc_edge_stop.txt",sep = "\t", header = FALSE)
colnames(PSR_JUC_STOP)<-c("juc_id", "probeset_id", "seqname", "juc_st_end",
"juc_stop_end","psr_st_end", "psr_stop_end")
PSR_JUC_STOP1<-merge(PSR_EXON_AB, PSR_JUC_STOP, by= "probeset_id", sort = FALSE)
junctions_psr<-merge(PSR_JUC_START1, PSR_JUC_STOP1, by="juc_id", sort= FALSE)
same_tr_junctions_psr<- junctions_psr %>% filter(junctions_psr$ensembl_transcript_id.x
== junctions_psr$ensembl_transcript_id.y)
same_tr_junctions_psr1<-same_tr_junctions_psr[,c(1,3,17,12,13)]
colnames(same_tr_junctions_psr1)<-c("JUN_ID","start_exon_id",
"stop_JUN_exon_id","junction_start_edge" ,"junction_stop_edge" )
df_junctions1<-rbind(df_junctions, same_tr_junctions_psr1)
df_junctions1<-
df_junctions1[order(df_junctions1$start_exon_id,df_junctions1$junction_start_edge),]
df_junctions1<-df_junctions1[!duplicated(df_junctions1),]
df_junctions<-df_junctions1
POSITIONDATA_JUC_EXON<-merge(TR_CANDIDATE_SPLICING, df_junctions, by =
"start_exon_id", sort = FALSE)
JUC_EXON_sort<-
POSITIONDATA_JUC_EXON[order(POSITIONDATA_JUC_EXON$chr_name,POSITIONDATA_JUC_EXON$ensem
bl_transcript_id, POSITIONDATA_JUC_EXON$start_exon_id),]

##-----
## SE DESCARTAN LAS UNIONES QUE NO PERTENECEN A EXONES CONTINUOS.
##-----
JUC_EXON_sort$continuous<- ifelse(JUC_EXON_sort$stop_exon_id ==
JUC_EXON_sort$stop_JUN_exon_id,"same","diff")
JUC_Continuous_exon<- JUC_EXON_sort %>% filter(continuous == "same")
JUC_Continuous_exon1<-JUC_Continuous_exon[,c(12,2,3,4,1,5,6,9,10,11,13,15,16)]
JUC_Continuous_exon1_SORT<-
JUC_Continuous_exon1[order(JUC_Continuous_exon1$chr_name,JUC_Continuous_exon1$ensembl
transcript_id, JUC_Continuous_exon1$start_exon_start),]
PSR_SORT_FILTER<-PSR_EXON_SORT[which(PSR_EXON_SORT$ensembl_transcript_id %in%
JUC_Continuous_exon1$ensembl_transcript_id),]
JUC_Found<-JUC_Continuous_exon1_SORT[!duplicated(JUC_Continuous_exon1_SORT[c(4,5)]),]
JUC_Found<-JUC_Found[order(JUC_Found$chr_name,JUC_Found$ensembl_transcript_id,
JUC_Found$start_exon_start),]
JUN_FUN1<-function(x, JUC_F, TRANS_COMPL){
  exon<-TRANS_COMPL[x,]
  TR<-subset(JUC_F, ensembl_transcript_id == exon$ensembl_transcript_id)
  EXON_OUT<-TR$start_exon_id[TR$start_exon_id %in% exon$start_exon_id]
  if (length(EXON_OUT) == 0){
    result <- NA
  }
  else{
    result<-as.character(EXON_OUT)
  }
  result
}

comp_exon_list<-lapply(seq(1:nrow(TR_CANDIDATE_SPLICING)),JUN_FUN1, TRANS_COMPL=
TR_CANDIDATE_SPLICING, JUC_F=JUC_Found)
df_compl<-data.frame(start_exon_id= rep(TR_CANDIDATE_SPLICING$start_exon_id,
sapply(comp_exon_list,length)), EXON_ID= unlist(comp_exon_list))
df_compl1<-cbind(df_compl,TR_CANDIDATE_SPLICING[,3])
df_compl1.1<- df_compl1[is.na(df_compl1$EXON_ID),]
PSR_EXON_SORT1<-PSR_EXON_SORT[which(!PSR_EXON_SORT$ensembl_transcript_id %in%
df_compl1.1$`TR_CANDIDATE_SPLICING[, 3]`),]
length(unique(PSR_EXON_SORT1$ensembl_transcript_id))
JUC_Continuous_exon1_SORT1<-
JUC_Continuous_exon1_SORT[which(!JUC_Continuous_exon1_SORT$ensembl_transcript_id %in%
df_compl1.1$`TR_CANDIDATE_SPLICING[, 3]`),]
length(unique(JUC_Continuous_exon1_SORT1$ensembl_transcript_id))
TR_ONLY_ONE_EXON_CANDIDATE<-as.data.frame(TR_ONLY_ONE_EXON_CANDIDATE)
TR_ALL<-rbind(PSR_EXON_SORT1, TR_ONLY_ONE_EXON_CANDIDATE)

```

```

TR_ALL<-TR_ALL[order(TR_ALL$chr_name, TR_ALL$ensembl_transcript_id,
TR_ALL$exon_chrom_start),]
GTF_TR_ALL<-gtf_exon1[which(gtf_exon1$transcript_id %in%
TR_ALL$ensembl_transcript_id),c(1,22,10,9,4,5)]
GTF_TR_ALL<-GTF_TR_ALL[order(GTF_TR_ALL$seqname, GTF_TR_ALL$transcript_id,
GTF_TR_ALL$start),]
GTF_TR_FILT<-GTF_TR_ALL[,2:3]
GTF_TR_FILT<-GTF_TR_FILT[!duplicated(GTF_TR_FILT[,c(1,2)]),]
TR_ALL_FILT<-TR_ALL[,2:3]
TR_ALL_FILT<-TR_ALL_FILT[!duplicated(TR_ALL_FILT[,c(1,2)]),]
TR_ALL_FILT$ensembl_exon_id<-as.character(TR_ALL_FILT$ensembl_exon_id)
temp_GTF_TR_FILT<-GTF_TR_FILT %>%
  group_by(transcript_id) %>%
  mutate(exon_count=n())
temp_GTF_TR_FILT<-temp_GTF_TR_FILT[!duplicated(temp_GTF_TR_FILT[,2]),c(2,3)]
temp_TR_ALL_FILT<-TR_ALL_FILT%>%
  group_by(ensembl_transcript_id) %>%
  mutate(exon_count_detected=n())
temp_TR_ALL_FILT<-temp_TR_ALL_FILT[!duplicated(temp_TR_ALL_FILT[,2]), c(2,3)]
annot_filt<-merge(temp_GTF_TR_FILT,temp_TR_ALL_FILT, by.x = "transcript_id", by.y =
"ensembl_transcript_id", sort = FALSE)
annot_filt$test <- ifelse(annot_filt$exon_count ==
annot_filt$exon_count_detected,"same","diff")
annot_filt_dif<- annot_filt %>% filter(test == "diff")
annot_filt<-annot_filt %>% filter(test == "same")
Juc_filt<-
JUC_Continuous_exon1_SORT1[which(JUC_Continuous_exon1_SORT1$ensembl_transcript_id %in%
annot_filt$transcript_id),]
PSR_EXON_SORT_FILT<-PSR_EXON_SORT1[which(PSR_EXON_SORT1$ensembl_transcript_id %in%
Juc_filt$ensembl_transcript_id),]
length(unique(PSR_EXON_SORT_FILT$ensembl_transcript_id))
TR_ONLY_ONE_EXON_CANDIDATE<-
TR_ONLY_ONE_EXON_CANDIDATE[which(TR_ONLY_ONE_EXON_CANDIDATE$ensembl_transcript_id %in%
annot_filt$transcript_id),]

##-----
#ANÁLISIS DE DETECCIÓN GENES Y TRANSCRITOS
##-----
##-----
##UNION DE ANOTACIONES CON DATOS GENERADOS EN EXPRESSION CONSOLE DE
##AFFYMETRIX PARA PRESENCIAS Y AUSENCIAS DE PROBESETS.
##-----

PSR_ALL<-rbind(PSR_EXON_SORT_FILT, TR_ONLY_ONE_EXON_CANDIDATE)
EXONLEVEL_DABG<- read.table("dabg.summary_ALL_APT_AFFYMETRIX.txt",row.names = 1,
header = TRUE, sep = "\t")
PYA_DETECTION<-EXONLEVEL_DABG[,c(6,7,8,1,2,3,4,5)]

#-----
### DETERMINAR AQUELLOS PROBESETS CON p<0.05 COMO PRESENTES Y EL RESTO SON AUSENTES.
#-----

PYADETECTION_TUMOR<-ifelse(PYA_DETECTION[,1:8]<=0.05,T,F)
PYADETECTION_TUMOR<-as.data.frame(PYADETECTION_TUMOR)
PYADETECTION_TUMOR$probeset_id<-rownames(PYADETECTION_TUMOR)
PYADETECTION_TUMOR$probeset_id<- gsub('.hg.1', '', PYADETECTION_TUMOR$probeset_id)

#-----
# SE ANEXAN ANOTACIONES DE LOS PROBESETS.
#-----

PSR_DABG<-PYADETECTION_TUMOR[which(PYADETECTION_TUMOR$probeset_id %in%
PSR_ALL$probeset_id),]
PSR_PYA_ANNOT<-merge(PSR_ALL[,c(1,2)], PSR_DABG, by = "probeset_id", sort = FALSE)

#-----

```

```

#OBTENER EL VALOR DE RESENCIA Y AUSENCIA A PARTIR DE LOS PROBESETS QUE SE UNEN A
#CADA EXON. SI EL 50% DE PROBESETS ES PRESENTE EL EXON SE CONSIDERA PRESENTE
#-----

PSR_PYA_ANNOT$ensembl_exon_id<-as.vector(PSR_PYA_ANNOT$ensembl_exon_id)
PSR_PYA_ANNOT<-PSR_PYA_ANNOT[order(PSR_PYA_ANNOT$ensembl_exon_id),]
EXON_FOR_COLLAPSE<-PSR_PYA_ANNOT[!duplicated(PSR_PYA_ANNOT[c(1,2)]),]

##-----
### COLAPSAR LOS PROBESETS PARA OBTENER EL NIVEL DE DETECCIÓN ÚNICO DE CADA EXÓN.
##-----

EXON_FOR_COLLAPSE<-EXON_FOR_COLLAPSE[,2:10]
MTZ_EXON_PYA_TUMORES<-EXON_FOR_COLLAPSE %>%
  group_by(ensembl_exon_id) %>%
  mutate_all(funs(sum.)/length.)) %>%
  distinct(ensembl_exon_id, .keep_all= TRUE)

MTZ_EXON_PYA_TUMORES<-as.data.frame(MTZ_EXON_PYA_TUMORES)
MTZ_EXON_PYA_TUMORES[,2:9]<-ifelse(MTZ_EXON_PYA_TUMORES[,2:9] >= 0.5 ,T, F)

#-----
# SE RECUPERA EL LISTADO DE TRANSCRITOS CORRESPONDIENTES A LOS EXONES
#-----

EXON_TR<-PSR_ALL[!duplicated(PSR_ALL[c(2,3)]),]
EXON_TR<-EXON_TR[,-c(1,4)]
MTZ_EXON_PYA_TUMORES_TR<-merge(EXON_TR, MTZ_EXON_PYA_TUMORES, by = "ensembl_exon_id")
MTZ_EXON_PYA_TUMORES_TR1<-
MTZ_EXON_PYA_TUMORES_TR[order(MTZ_EXON_PYA_TUMORES_TR$ensembl_transcript_id),]
MTZ_EXON_PYA_TUMORES_TR<-MTZ_EXON_PYA_TUMORES_TR1[,c(2,8:15)]

##-----
## COLAPSAR LOS EXONES A NIVEL DE TRANSCRITO
##-----

MTZ_EXON_PYA_TUMORES_TR$ensembl_transcript_id<-
as.vector(MTZ_EXON_PYA_TUMORES_TR$ensembl_transcript_id)
MTZ_TR_DETECCION<-MTZ_EXON_PYA_TUMORES_TR %>%
  group_by(ensembl_transcript_id) %>%
  mutate_all(funs(sum.)/length.)) %>%
  distinct(ensembl_transcript_id, .keep_all= TRUE)

MTZ_TR_DETECCION<-as.data.frame(MTZ_TR_DETECCION)
MTZ_TR_DETECCION[,2:9]<-ifelse(MTZ_TR_DETECCION[,2:9] == 1 ,T, F)

#-----
##UNIR ANOTACIONES DE LOS PROBESETS JUC CON LAS PRESENCIAS Y AUSENCIAS DE ESTOS.
##SE UNEN LAS ANOTACIONES CON LOS RESULTADOS DE DETECCIÓN
#-----
colnames(Juc_filt)[11]<-"probeset_id"
PROBESET_JUC_EXON<-Juc_filt[,c(3,4,5,8,11)]
JUC_DABG<-PYADETECCION_TUMOR[which(PYADETECCION_TUMOR$probeset_id %in%
PROBESET_JUC_EXON$probeset_id),]
JUC_PYA_ANNOT<-merge(PROBESET_JUC_EXON, JUC_DABG, by = "probeset_id", sort = FALSE)
JUC_PYA_ANNOT<-JUC_PYA_ANNOT[order(JUC_PYA_ANNOT$ensembl_transcript_id),]
colnames(JUC_PYA_ANNOT)<-
c("probeset_id","ensembl_gene_id","ensembl_transcript_id","start_exon_id","stop_exon_i
d","F141", "F161", "F278", "F73", "F74", "F78", "F85", "F86")
##-----
##### SE OBTIENEN LOS DATOS DE PRESENCIA Y AUSENCIA DE LOS TRANSCRITOS TENIENDO EN
CUENTA LOS PROBESETS QUE SE UNEN A LOS EXONES Y POSTERIORMENTE LAS UNIONES ENTRE
EXONES. SI LOS EXONES ESTÁN PRESENTES PERO LAS UNIONES ESTAN AUSENTES EL TRANSCRITO SE
CONSIDERA AUSENTE
##-----

JUC_PYA_ANNOT_COMPL<-within(JUC_PYA_ANNOT, EXON_ID<- paste(start_exon_id,
stop_exon_id, sep=' to '))
JUC_FOR_COLLAPSE<-JUC_PYA_ANNOT_COMPL[,c(6:14)]

```

```

JUC_FOR_COLLAPSE$EXON_ID<-as.vector(JUC_FOR_COLLAPSE$EXON_ID)

##-----
#### SE BUSCAN LAS UNIONES QUE ENLAZAN LOS MISMOS EXONES Y SE ASIGNA EL VALOR DE
#### PPRESENCIA O AUSENCIA DE CADA UNION.
##-----

MTZ_JUC_DETECTION<-JUC_FOR_COLLAPSE %>%
  group_by(EXON_ID) %>%
  mutate_all(funs(sum(.)/length(.))) %>%
  distinct(EXON_ID, .keep_all= TRUE)

MTZ_JUC_DETECTION<-as.data.frame(MTZ_JUC_DETECTION)
MTZ_JUC_DETECTION[,1:8]<-ifelse(MTZ_JUC_DETECTION[,1:8] > 0 ,T, F)
dim(MTZ_JUC_DETECTION)

##-----
#### SE BUSCAN LAS UNIONES DE CADA TRANSCRITOS Y SE ASIGNA EL VALOR DE PRESENCIA
#### O AUSENCIA.
##-----

JUC_TRANSCRIPTS<-JUC_PYA_ANNOT_COMPL[!duplicated(JUC_PYA_ANNOT_COMPL[c(3,14)]),]
JUC_TRANSCRIPTS<-JUC_TRANSCRIPTS[,c(3,14)]
JUC_TR_DETECTION_COLLAPSE<-merge(JUC_TRANSCRIPTS,MTZ_JUC_DETECTION, by = "EXON_ID")
JUC_TR_DETECTION_COLLAPSE<-
JUC_TR_DETECTION_COLLAPSE[order(JUC_TR_DETECTION_COLLAPSE$ensembl_transcript_id),]
JUC_TR_DETECTION_COLLAPSE$ensembl_transcript_id<-
as.vector(JUC_TR_DETECTION_COLLAPSE$ensembl_transcript_id)

##-----
## SE COLAPSAN LAS UNIONES DE CADA TRANSCRITO
##-----

JUC_TR_DETECTION_COLLAPSE<-JUC_TR_DETECTION_COLLAPSE[,-1]
MTZ_TR_JUCS_DETECTION<-JUC_TR_DETECTION_COLLAPSE %>%
  group_by(ensembl_transcript_id) %>%
  mutate_all(funs(sum(.)/length(.))) %>%
  distinct(ensembl_transcript_id, .keep_all= TRUE)

MTZ_TR_JUCS_DETECTION<-as.data.frame(MTZ_TR_JUCS_DETECTION)
MTZ_TR_JUCS_DETECTION[,2:9]<-ifelse(MTZ_TR_JUCS_DETECTION[,2:9] == 1 ,T, F)

##-----
##AHORA SE UNEN LOS VALORES DE DETECCIÓN DE PSR Y DE JUCS.
##UN TRANSCRITO ESTÁ PRESENTE SI TIENE TODOS SUS EXONES Y UNIONES INTEREXONALES ##
## PRESENTES O DETECTABLES
##-----

colnames(MTZ_TR_DETECTION)<-c("ensembl_transcript_id","F141_exon", "F161_exon",
"F278_exon", "F073_exon", "F074_exon", "F078_exon", "F085_exon", "F086_exon")
colnames(MTZ_TR_JUCS_DETECTION)<-c("ensembl_transcript_id","F141_jun", "F161_jun",
"F278_jun", "F073_jun", "F074_jun", "F078_jun", "F085_jun", "F086_jun")
transcripts_detecton_exon_jun<-merge(x=MTZ_TR_DETECTION, y=MTZ_TR_JUCS_DETECTION, by =
"ensembl_transcript_id")
transcripts_detecton_exon_jun<-
transcripts_detecton_exon_jun[,c(1,2,10,3,11,4,12,5,13,6,14,7,15,8,16,9,17)]
transcripts_detecton_exon_jun[,2:17]<-ifelse(transcripts_detecton_exon_jun[,2:17] == T
,1, 0)

n<- ncol(transcripts_detecton_exon_jun)
ind <- split(2:n, rep(2:n, each = 2, length = n - 1))
Transcript_det_2<- do.call(cbind, lapply(ind, function(i)
rowSums(transcripts_detecton_exon_jun[, i])))
transcripts_detecton_exon_jun1<-cbind(transcripts_detecton_exon_jun,Transcript_det_2)
colnames(transcripts_detecton_exon_jun1)[18:25]<-c("TR_F141", "TR_F161", "TR_F278",
"TR_F073", "TR_F074", "TR_F078", "TR_F085", "TR_F086")

Transcript_det_exon_jun_all<-transcripts_detecton_exon_jun1[,c(1,18:25)]
rownames(Transcript_det_exon_jun_all)<-Transcript_det_exon_jun_all[,1]

```

```

Transcript_det_exon_jun_all1<-ifelse(Transcript_det_exon_jun_all[,2:9] == 2, 1, 0)

transcript_one_exon_detection<-
MTZ_TR_DETECTION[which(!MTZ_TR_DETECTION$ensembl_transcript_id %in%
MTZ_TR_JUCS_DETECTION$ensembl_transcript_id),]
colnames(transcript_one_exon_detection)<-c("ensembl_transcript_id",
"TR_F141","TR_F161","TR_F278","TR_F073","TR_F074",
"TR_F078","TR_F085","TR_F086")
row.names(transcript_one_exon_detection)<-transcript_one_exon_detection[,1]
transcript_one_exon_detection<-transcript_one_exon_detection[,-1]
transcript_one_exon_detection<-ifelse(transcript_one_exon_detection == T ,1, 0)

Transcript_det_exon_jun_all11<-
rbind(Transcript_det_exon_jun_all1,transcript_one_exon_detection)

##-----
### OBTENER TRANSCRITOS PRESENTES, VARIABLES Y AUSENTES
##-----

TR_ALL_PRESENTS<-
Transcript_det_exon_jun_all11[rowSums(Transcript_det_exon_jun_all11[,1:8]) == 8,]
TR_ALL_VARIABLES<-
Transcript_det_exon_jun_all11[rowSums(Transcript_det_exon_jun_all11[,1:8]) > 0 &
rowSums(Transcript_det_exon_jun_all11[,1:8]) < 8,]
TR_ALL_ABSENTS<-
Transcript_det_exon_jun_all11[rowSums(Transcript_det_exon_jun_all11[,1:8]) == 0,]

##-----
##### AHORA SE BUSCA LA PRESENCIA Y AUSENCIA DE LOS GENES A PARTIR DE
##### LA PRESENCIA O DETECCIÓN DE CUALQUIERA DE SUS TRANSCRITOS
##-----

GENE_AND_TR_ANNOT<-PSR_ALL[!duplicated(PSR_ALL[,c(3,4)]),c(3,4)]
Transcript_det_exon_jun_all111<-
as.data.frame(ifelse(Transcript_det_exon_jun_all11[,1:8] == 1,T,F))
Transcript_det_exon_jun_all111$ensembl_transcript_id<-
rownames(Transcript_det_exon_jun_all111)
GENE_DETECTION<-merge(GENE_AND_TR_ANNOT, Transcript_det_exon_jun_all111, by =
"ensembl_transcript_id")
GENE_DETECTION<-GENE_DETECTION[order(GENE_DETECTION$ensembl_gene_id),]
GENE_DETECTION$ensembl_gene_id<-as.vector(GENE_DETECTION$ensembl_gene_id)

##-----
##### SE BUSCA SI ALGUNO DE LOS TRANSCRITOS DEL GENE ESTAN PRESENTES ENTONCES
##### EL GENE SE CONSIDERA COMO DETECTADO.
##-----

GENE_DETECTION<-GENE_DETECTION[,-1]
MTZ_GENE_PAY_ALL_DETECTION<-GENE_DETECTION %>%
  group_by(ensembl_gene_id) %>%
  mutate_all(funs(sum.)/length.)) %>%
  distinct(ensembl_gene_id, .keep_all= TRUE)

MTZ_GENE_PAY_ALL_DETECTION<-as.data.frame(MTZ_GENE_PAY_ALL_DETECTION)
MTZ_GENE_PAY_ALL_DETECTION[,2:9]<-ifelse(MTZ_GENE_PAY_ALL_DETECTION[,2:9] > 0 ,1, 0)

##-----
### OBTENER GRUPO DE GENES PRESENTES, VARIABLES Y AUSENTES.
##-----

GENE_ALL_PRESENTS<-
MTZ_GENE_PAY_ALL_DETECTION[rowSums(MTZ_GENE_PAY_ALL_DETECTION[,2:9]) == 8,]
GENE_ALL_VARIABLES<-
MTZ_GENE_PAY_ALL_DETECTION[rowSums(MTZ_GENE_PAY_ALL_DETECTION[,2:9]) > 0 &
rowSums(MTZ_GENE_PAY_ALL_DETECTION[,2:9]) < 8,]
GENE_ALL_ABSENTS<-MTZ_GENE_PAY_ALL_DETECTION[rowSums(MTZ_GENE_PAY_ALL_DETECTION[,2:9])
== 0,]

```

Archivos de entrada.

-HTA-2_0.na36.hg19.probeset.csv y Transcript_clusterid_TO_probeset_HTA2.txt: archivos de anotación para probesets del chip HTA2.0 de affymetrix.

-gencode.v19.annotation.gtf_withproteinids: Archivo de anotación de exones, transcritos y genes del genoma versión 19 de GENECODE.

-psr_exon.txt: Archivo con probesets mapeados para exones según el genoma versión 19. (generado en python)

-psr_exon1.txt: Archivo con probesets mapeados para exones según el genoma versión 19. (generado en python)

-psr_exon2.txt: Archivo con probesets mapeados para exones según el genoma versión 19. (generado en python)

-psr_juc_edge.txt: Archivo con probesets de uniones interexonales mapeados para exones según el genoma versión 19. (generado en python)

-psr_juc_edge_stop.txt: Archivo con probesets de uniones interexonales mapeados para exones según el genoma versión 19. (generado en python)

-dabg.summary_ALL_APT_AFFYMETRIX.txt: Archivo de Resultados análisis DABG obtenido de expresión console.

Transcritos expresados diferencialmente (Muestras Bilaterales Vs Muestras unilaterales)

Transcrito	Gen	Símbolo	logFC (Bilateral Vs Unilateral)	adj.P.Val
ENST00000393963.3	ENSG00000138472.6	GUCA1C	8.32	0.047
ENST00000602432.1	ENSG00000111886.9	GABRR2	7.32	0.044
ENST00000497198.1	ENSG00000116703.12	PDC	6.45	0.046
ENST00000508086.1	ENSG00000184305.10	CCSER1	6.43	0.020
ENST00000513522.1	ENSG00000184305.10	CCSER1	5.50	0.031
ENST00000482017.1	ENSG00000053524.7	MCF2L2	5.46	0.044
ENST00000466691.1	ENSG00000073756.7	PTGS2	5.35	0.020
ENST00000371447.3	ENSG00000095464.9	PDE6C	5.18	0.044
ENST00000432775.2	ENSG00000184305.10	CCSER1	4.90	0.031
ENST00000535430.1	ENSG00000175727.9	MLXIP	4.88	0.044
ENST00000330816.5	ENSG00000188107.9	EYS	4.78	0.021
ENST00000366272.2	ENSG00000175727.9	MLXIP	4.76	0.031
ENST00000378225.1	ENSG00000152463.10	OLAH	4.74	0.020
ENST00000539861.1	ENSG00000175727.9	MLXIP	4.72	0.046
ENST00000538698.1	ENSG00000175727.9	MLXIP	4.69	0.044
ENST00000461074.1	ENSG00000053524.7	MCF2L2	4.66	0.031
ENST00000459750.1	ENSG00000053524.7	MCF2L2	4.63	0.024
ENST00000319080.7	ENSG00000175727.9	MLXIP	4.60	0.044
ENST00000538061.1	ENSG00000175727.9	MLXIP	4.58	0.048
ENST00000509176.1	ENSG00000184305.10	CCSER1	4.56	0.045
ENST00000370616.2	ENSG00000188107.9	EYS	4.52	0.021
ENST00000333691.8	ENSG00000184305.10	CCSER1	4.43	0.045
ENST00000370621.3	ENSG00000188107.9	EYS	4.43	0.020
ENST00000503581.1	ENSG00000188107.9	EYS	4.41	0.020
ENST00000559627.1	ENSG00000073756.7	PTGS2	4.40	0.044
ENST00000367468.5	ENSG00000073756.7	PTGS2	4.40	0.044
ENST00000472196.1	ENSG00000108018.11	SORCS1	4.39	0.048
ENST00000505073.1	ENSG00000184305.10	CCSER1	4.39	0.044
ENST00000580666.1	ENSG00000108932.7	SLC16A6	4.36	0.047
ENST00000490885.2	ENSG00000073756.7	PTGS2	4.36	0.044
ENST00000424833.1	ENSG00000136531.9	SCN2A	4.31	0.046
ENST00000320634.3	ENSG00000135472.4	FAIM2	4.31	0.050
ENST00000442690.1	ENSG00000071909.14	MYO3B	4.30	0.048
ENST00000429028.1	ENSG00000152463.10	OLAH	4.29	0.021
ENST00000378228.3	ENSG00000152463.10	OLAH	4.28	0.031
ENST00000514352.1	ENSG00000184305.10	CCSER1	4.24	0.046
ENST00000428897.1	ENSG00000152463.10	OLAH	4.23	0.021
ENST00000466171.1	ENSG00000125633.6	CCDC93	4.20	0.045
ENST00000508550.1	ENSG00000184305.10	CCSER1	4.15	0.044
ENST00000470650.1	ENSG00000125633.6	CCDC93	4.14	0.044
ENST00000523377.1	ENSG00000083099.6	LYRM2	4.12	0.020
ENST00000369698.1	ENSG00000108018.11	SORCS1	4.11	0.045
ENST00000414362.2	ENSG00000053524.7	MCF2L2	4.11	0.036
ENST00000488149.1	ENSG00000053524.7	MCF2L2	4.10	0.031
ENST00000342274.4	ENSG00000183337.12	BCOR	4.09	0.044
ENST00000378444.4	ENSG00000183337.12	BCOR	4.09	0.044
ENST00000397354.3	ENSG00000183337.12	BCOR	4.09	0.044
ENST00000473233.1	ENSG00000053524.7	MCF2L2	4.05	0.038
ENST00000447025.2	ENSG00000053524.7	MCF2L2	4.00	0.044
ENST00000328913.3	ENSG00000053524.7	MCF2L2	4.00	0.044
ENST00000378455.4	ENSG00000183337.12	BCOR	3.96	0.045
ENST00000492331.1	ENSG00000053524.7	MCF2L2	3.94	0.031
ENST00000489327.1	ENSG00000115977.14	AAK1	3.93	0.046
ENST00000541750.1	ENSG00000175727.9	MLXIP	3.86	0.046
ENST00000344440.6	ENSG00000108018.11	SORCS1	3.85	0.044
ENST00000437999.1	ENSG00000125633.6	CCDC93	3.84	0.046
ENST00000427012.1	ENSG00000183337.12	BCOR	3.84	0.044
ENST00000442018.1	ENSG00000183337.12	BCOR	3.84	0.044
ENST00000263054.6	ENSG00000108018.11	SORCS1	3.83	0.045
ENST00000509848.1	ENSG00000214944.5	ARHGEF28	3.83	0.046

ENST00000463289.1	ENSG00000188467.6	SLC24A5	3.81	0.046
ENST00000413905.1	ENSG00000183337.12	BCOR	3.79	0.044
ENST00000496611.1	ENSG00000130226.12	DPP6	-3.79	0.048
ENST00000485637.1	ENSG00000135299.12	ANKRD6	3.73	0.045
ENST00000475427.2	ENSG00000095464.9	PDE6C	3.69	0.044
ENST00000536675.2	ENSG00000101542.5	CDH20	-3.68	0.031
ENST00000509996.1	ENSG00000138798.7	EGF	3.68	0.046
ENST00000262717.4	ENSG00000101542.5	CDH20	-3.63	0.031
ENST00000319432.5	ENSG00000125633.6	CCDC93	3.63	0.045
ENST00000376300.2	ENSG00000125633.6	CCDC93	3.63	0.045
ENST00000449382.2	ENSG00000188467.6	SLC24A5	3.57	0.046
ENST00000538374.1	ENSG00000101542.5	CDH20	-3.56	0.031
ENST00000377037.2	ENSG00000175727.9	MLXIP	3.54	0.046
ENST00000533703.1	ENSG00000166801.11	FAM111A	3.54	0.046
ENST00000531147.1	ENSG00000166801.11	FAM111A	3.52	0.031
ENST00000378463.1	ENSG00000183337.12	BCOR	3.52	0.047
ENST00000524235.1	ENSG00000169946.9	ZFPM2	-3.47	0.044
ENST00000502701.1	ENSG00000145362.12	ANK2	3.44	0.044
ENST00000357731.5	ENSG00000172260.9	NEGR1	3.44	0.046
ENST00000416412.1	ENSG00000066923.13	STAG3	3.44	0.045
ENST00000359945.3	ENSG00000180336.13	C17orf104	3.40	0.046
ENST00000341115.4	ENSG00000055732.8	MCOLN3	3.38	0.047
ENST00000318160.4	ENSG00000180875.4	GREM2	3.38	0.044
ENST00000475312.1	ENSG00000055732.8	MCOLN3	3.32	0.045
ENST00000512298.1	ENSG00000145362.12	ANK2	3.28	0.044
ENST00000515034.1	ENSG00000145362.12	ANK2	3.28	0.046
ENST00000475664.1	ENSG00000053524.7	MCF2L2	3.11	0.049
ENST00000481161.1	ENSG00000115271.6	GCA	3.07	0.044
ENST00000229634.9	ENSG00000111912.14	NCOA7	3.00	0.046
ENST00000368357.3	ENSG00000111912.14	NCOA7	2.99	0.049
ENST00000261483.4	ENSG00000112893.5	MAN2A1	-2.96	0.041
ENST00000479199.1	ENSG00000115271.6	GCA	2.93	0.044
ENST00000389048.3	ENSG00000171094.11	ALK	2.91	0.049
ENST00000468889.1	ENSG00000029639.6	TFB1M	2.87	0.044
ENST00000429691.2	ENSG00000115271.6	GCA	2.83	0.047
ENST00000409309.3	ENSG00000198075.5	SULT1C4	2.80	0.031
ENST00000486151.1	ENSG00000170324.15	FRMPD2	2.79	0.044
ENST00000524796.1	ENSG00000177570.9	SAMD12	2.76	0.046
ENST00000503169.1	ENSG00000087274.12	ADD1	2.74	0.048
ENST00000461259.1	ENSG00000135205.10	CCDC146	2.69	0.049
ENST00000360971.2	ENSG00000197594.7	ENPP1	2.65	0.044
ENST00000373803.2	ENSG00000101079.16	NDRG3	2.64	0.047
ENST00000513998.1	ENSG00000197594.7	ENPP1	2.59	0.046
ENST00000305531.3	ENSG00000170324.15	FRMPD2	2.59	0.045
ENST00000349004.1	ENSG00000101079.16	NDRG3	2.59	0.046
ENST00000407470.4	ENSG00000170324.15	FRMPD2	2.57	0.046
ENST00000526765.1	ENSG00000177570.9	SAMD12	2.57	0.044
ENST00000459624.1	ENSG00000197594.7	ENPP1	2.56	0.045
ENST00000374201.3	ENSG00000170324.15	FRMPD2	2.56	0.046
ENST00000393380.2	ENSG00000188107.9	EYS	2.54	0.044
ENST00000359675.2	ENSG00000101079.16	NDRG3	2.54	0.046
ENST00000437150.2	ENSG00000115271.6	GCA	2.54	0.049
ENST00000565351.1	ENSG00000140859.11	KIFC3	2.48	0.044
ENST00000463842.1	ENSG00000156500.10	FAM122C	2.47	0.044
ENST00000446271.1	ENSG00000115271.6	GCA	2.45	0.048
ENST00000487087.1	ENSG00000168386.14	FILIP1L	-2.43	0.031
ENST00000409003.4	ENSG00000177570.9	SAMD12	2.41	0.047
ENST00000370618.3	ENSG00000188107.9	EYS	2.29	0.046
ENST00000504484.1	ENSG00000163697.12	APBB2	2.24	0.046
ENST00000495147.1	ENSG00000156500.10	FAM122C	2.14	0.044
ENST00000459699.1	ENSG00000066923.13	STAG3	2.13	0.046
ENST00000486708.1	ENSG00000198089.10	SFI1	2.05	0.046
ENST00000494013.1	ENSG00000100095.14	SEZ6L	-1.89	0.047
ENST00000478726.1	ENSG00000128594.3	LRRC4	-1.66	0.047

Genes presentes exclusivamente en la línea celular Y79.

Ensembl	nombre	descripción
ENSG00000101204	CHRNA4	cholinergic receptor nicotinic alpha 4 subunit
ENSG00000105605	CACNG7	calcium voltage-gated channel auxiliary subunit gamma 7
ENSG00000124939	SCGB2A1	secretoglobin family 2A member 1
ENSG00000137834	SMAD6	SMAD family member 6
ENSG00000143333	RGS16	regulator of G protein signaling 16
ENSG00000143772	ITPKB	inositol-trisphosphate 3-kinase B
ENSG00000149654	CDH22	cadherin 22
ENSG00000151892	GFRA1	GDNF family receptor alpha 1
ENSG00000152242	C18orf25	chromosome 18 open reading frame 25
ENSG00000152284	TCF7L1	transcription factor 7 like 1
ENSG00000158480	SPATA2	spermatogenesis associated 2
ENSG00000164897	TMUB1	transmembrane and ubiquitin like domain containing 1
ENSG00000165905	LARGE2	LARGE xylosyl- and glucuronyltransferase 2
ENSG00000166482	MFAP4	microfibril associated protein 4
ENSG00000167646	DNAAF3	dynein axonemal assembly factor 3
ENSG00000169220	RGS14	regulator of G protein signaling 14
ENSG00000171551	ECEL1	endothelin converting enzyme like 1
ENSG00000171786	NHLH1	nescient helix-loop-helix 1
ENSG00000172840	PDP2	pyruvate dehydrogenase phosphatase catalytic subunit 2
ENSG00000181698	OR5T1	olfactory receptor family 5 subfamily T member 1
ENSG00000183688	RFLNB	refilin B
ENSG00000184221	OLIG1	oligodendrocyte transcription factor 1
ENSG00000198965	OR10R2	olfactory receptor family 10 subfamily R member 2
ENSG00000231861	OR5K2	olfactory receptor family 5 subfamily K member 2
ENSG00000237388	OR4A47	olfactory receptor family 4 subfamily A member 47
ENSG00000237412	PRSS56	serine protease 56
ENSG00000257019	OR37K	olfactory receptor, family 13, subfamily C, member 2

REFERENCIAS

- Affymetrix. (2015). Microarray normalization using Signal Space Transformation with probe Guanine Cytosine Count Correction. Retrieved from http://tools.thermofisher.com/content/sfs/brochures/sst_gccn_whitepaper.pdf
- Affymetrix, I. (2005). Microarray normalization using Signal Space Transformation with probe Guanine Cytosine Count Correction. Retrieved from https://tools.thermofisher.com/content/sfs/brochures/sst_gccn_whitepaper.pdf
- Bowles, E., Corson, T. W., Bayani, J., Squire, J. A., Wong, N., Lai, P. B. S., & Gallie, B. L. (2007). Profiling genomic copy number changes in retinoblastoma beyond loss of RB1. *Genes Chromosomes and Cancer*, *46*(2), 118–129. <https://doi.org/10.1002/gcc.20383>
- Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., Lewis, S. E., ... Westerfield, M. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. <https://doi.org/10.1093/nar/gky1055>
- Castro-Magdonel, B. E., Orjuela, M., Camacho, J., García-Chéquer, A. J., Cabrera-Muñoz, L., Sadowinski-Pine, S., ... Ponce-Castañeda, M. V. (2017). miRNome landscape analysis reveals a 30 miRNA core in retinoblastoma. *BMC Cancer*, *17*(1), 458. <https://doi.org/10.1186/s12885-017-3421-3>
- Chakraborty, S., Khare, S., Dorairaj, S. K., Prabhakaran, V. C., Prakash, D. R., & Kumar, A. (2007). Identification of genes associated with tumorigenesis of retinoblastoma by microarray analysis. *Genomics*, *90*(3), 344–353. <https://doi.org/10.1016/J.YGENO.2007.05.002>
- Cieślik, M., & Chinnaiyan, A. M. (2018). Cancer transcriptome profiling at the juncture of clinical translation. *Nature Reviews Genetics*, *19*(2), 93–109. <https://doi.org/10.1038/nrg.2017.96>
- Clark, T. A., Schweitzer, A. C., Chen, T. X., Staples, M. K., Lu, G., Wang, H., ... Blume, J. E. (2007). Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biology*, *8*(4), R64. <https://doi.org/10.1186/gb-2007-8-4-r64>

- Corson, T. W., & Gallie, B. L. (2007, July). One hit, two hits, three hits, more? Genomic changes in the development of retinoblastoma. *Genes Chromosomes and Cancer*. *Genes Chromosomes Cancer*. <https://doi.org/10.1002/gcc.20457>
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., ... Flicek, P. (2018). Ensembl 2019. *Nucleic Acids Research*, *47*, 745–751. <https://doi.org/10.1093/nar/gky1113>
- Dalma-Weiszhausz, D. D., Warrington, J., Tanimoto, E. Y., & Miyada, C. G. (2006). [1] The Affymetrix GeneChip® Platform: An Overview. *Methods in Enzymology*. *Methods Enzymol*. [https://doi.org/10.1016/S0076-6879\(06\)10001-4](https://doi.org/10.1016/S0076-6879(06)10001-4)
- Dimaras, H., & Corson, T. W. (2018). Retinoblastoma, the visible CNS tumor: A review. *Journal of Neuroscience Research*. <https://doi.org/10.1002/jnr.24213>
- Dimaras, H., Kimani, K., Dimba, E. A. O., Gronsdahl, P., White, A., Chan, H. S. L., & Gallie, B. L. (2012). Retinoblastoma. In *The Lancet* (Vol. 379, pp. 1436–1446). Lancet Publishing Group. [https://doi.org/10.1016/S0140-6736\(11\)61137-9](https://doi.org/10.1016/S0140-6736(11)61137-9)
- Elchuri, S. V., Rajasekaran, S., & Miles, W. O. (2018). RNA-Sequencing of primary Retinoblastoma tumors provides new insights and challenges into tumor development. *Frontiers in Genetics*, *9*(MAY). <https://doi.org/10.3389/fgene.2018.00170>
- Engel, B. E., Cress, W. D., & Santiago-Cardona, P. G. (2014, December 18). The retinoblastoma protein: A master tumor suppressor acts as a link between cell cycle and cell adhesion. *Cell Health and Cytoskeleton*. Dove Medical Press Ltd. <https://doi.org/10.2147/CHC.S28079>
- Fabbro, D., Cowan-Jacob, S. W., & Moebitz, H. (2015). Ten things you should know about protein kinases: IUPHAR Review 14. *British Journal of Pharmacology*, *172*(11), 2675–2700. <https://doi.org/10.1111/bph.13096>
- Fischer, A. H., Jacobson, K. A., Rose, J., & Zeller, R. (2008). Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harbor Protocols*, *3*(5). <https://doi.org/10.1101/pdb.prot4986>

- Friend, S. H., Bernards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M., & Dryja, T. P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature*, 323(6089), 643–646. <https://doi.org/10.1038/323643a0>
- Ganguly, A., Nichols, K. E., Grant, G., Rappaport, E., & Shields, C. (2009). Molecular karyotype of sporadic unilateral retinoblastoma tumors. *Retina*, 29(7), 1002–1012. <https://doi.org/10.1097/IAE.0b013e3181a0be05>
- Ganguly, A., & Shields, C. L. (2010). Differential gene expression profile of retinoblastoma compared to normal retina. *Molecular Vision*, 16, 1292–1303.
- García-Chequer, A. J., Méndez-Tenorio, A., Olguín-Ruiz, G., Sánchez-Vallejo, C., Isa, P., Arias, C. F., ... Ponce-Castañeda, M. V. (2016). Overview of recurrent chromosomal losses in retinoblastoma detected by low coverage next generation sequencing. *Cancer Genetics*, 209(3), 57–69. <https://doi.org/10.1016/j.cancergen.2015.12.001>
- Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, M. S. and B. V. (2015). gplots: Various R Programming Tools for Plotting Data. *R Package Version 3.0.1.1*. Retrieved from <https://cran.r-project.org/package=gplots>
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., ... Hubbard, T. J. (2012). GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Research*, 22(9), 1760–1774. <https://doi.org/10.1101/gr.135350.111>
- Haynes, W. A., Higdon, R., Stanberry, L., Collins, D., & Kolker, E. (2013). Differential Expression Analysis for Pathways. *PLoS Computational Biology*, 9(3). <https://doi.org/10.1371/journal.pcbi.1002967>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>

- Kapatai, G., Brundler, M.-A., Jenkinson, H., Kearns, P., Parulekar, M., Peet, A. C., & Mcconville, C. M. (2013). Gene expression profiling identifies different sub-types of retinoblastoma. *British Journal of Cancer*, *109*. <https://doi.org/10.1038/bjc.2013.283>
- Knudson, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, *68*(4), 820–823. <https://doi.org/10.1073/pnas.68.4.820>
- Kooi, I. E., Mol, B. M., Moll, A. C., van der Valk, P., de Jong, M. C., de Graaf, P., ... Dorsman, J. C. (2015a). Loss of photoreceptoriness and gain of genomic alterations in retinoblastoma reveal tumor progression. *EBioMedicine*, *2*(7), 660–670. <https://doi.org/10.1016/j.ebiom.2015.06.022>
- Kooi, I. E., Mol, B. M., Moll, A. C., van der Valk, P., de Jong, M. C., de Graaf, P., ... Dorsman, J. C. (2015b). Loss of photoreceptoriness and gain of genomic alterations in retinoblastoma reveal tumor progression. *EBioMedicine*, *2*(7), 660–670. <https://doi.org/10.1016/J.EBIOM.2015.06.022>
- Kumar, V., Abbas, A., & Aster, J. (2013). *Patología humana 9ed.*
- Kumar, V., Abbas, A. K., Aster, J. C., Perkins, J. A., Manzarbeitia Arambarri, F., Fernandez Acenero, M. J., ... Cotran, R. S. (2015). *Robbins y Cotran : patologia estructural y funcional / directores Vinay Kumar, Abul K. Abbas, Jon C. Aster ; con ilustraciones de James A. Perkins ; revision cientifica Felix Manzarbeitia Arambarri, Ma. J. Fernandez-Acenero. Patologia estructural y funcional.*
- Lee, W., Bookstein, R., Hong, F., Young, L., Shew, J., & Lee, E. (1987). Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science*, *235*(4794), 1394–1399. <https://doi.org/10.1126/science.3823889>
- Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, *47*(W1), W199–W205. <https://doi.org/10.1093/nar/gkz401>
- Liu, W. –, Mei, R., Di, X., Ryder, T. B., Hubbell, E., Dee, S., ... Smeekens, S. P. (2002).

- Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics*, 18(12), 1593–1599. <https://doi.org/10.1093/bioinformatics/18.12.1593>
- Liu, Y., Zhong, X., Wan, S., Zhang, W., Lin, J., Zhang, P., & Li, Y. (2014). p16INK4a expression in retinoblastoma: a marker of differentiation grade. *Diagnostic Pathology*, 9(1), 180. <https://doi.org/10.1186/s13000-014-0180-1>
- Mairal, A., Pinglier, E., Gilbert, E., Peter, M., Validire, P., Desjardins, L., ... Couturier, J. (2000). Detection of chromosome imbalances in retinoblastoma by parallel karyotype and CGH analyses. *Genes, Chromosomes and Cancer*, 28(4), 370–379. [https://doi.org/10.1002/1098-2264\(200008\)28:4<370::AID-GCC2>3.0.CO;2-8](https://doi.org/10.1002/1098-2264(200008)28:4<370::AID-GCC2>3.0.CO;2-8)
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002, December 6). The protein kinase complement of the human genome. *Science*. <https://doi.org/10.1126/science.1075762>
- McEvoy, J., Flores-Otero, J., Zhang, J., Nemeth, K., Brennan, R., Bradley, C., ... Dyer, M. A. (2011). Coexpression of Normally Incompatible Developmental Pathways in Retinoblastoma Genesis. *Cancer Cell*, 20(2), 260–275. <https://doi.org/10.1016/J.CCR.2011.07.005>
- Mejía-Pedroza, R. A., Espinal-Enríquez, J., & Hernández-Lemus, E. (2018). Pathway-based drug repositioning for breast cancer molecular subtypes. *Frontiers in Pharmacology*, 9(AUG). <https://doi.org/10.3389/fphar.2018.00905>
- Mendoza, P. R., Specht, C. S., Hubbard, G. B., Wells, J. R., Lynn, M. J., Zhang, Q., ... Grossniklaus, H. E. (2015). Histopathologic grading of anaplasia in retinoblastoma. *American Journal of Ophthalmology*, 159(4), 764-776.e3. <https://doi.org/10.1016/j.ajo.2014.12.014>
- Mol, B. M., Massink, M. P. G., van der Hout, A. H., Dommering, C. J., Zaman, J. M. A., Bosscha, M. I., ... Dorsman, J. C. (2014). High resolution SNP array profiling identifies variability in retinoblastoma genome stability. *Genes Chromosomes and Cancer*, 53(1), 1–14. <https://doi.org/10.1002/gcc.22111>

- Ofri, R. (2008). Retina. In *Slatter's Fundamentals of Veterinary Ophthalmology* (pp. 285–317). Elsevier Inc. <https://doi.org/10.1016/B978-072160561-6.50018-6>
- Okoniewski, M. J., & Miller, C. J. (2008, February). Comprehensive analysis of affymetrix exon arrays using bioConductor. *PLoS Computational Biology*. <https://doi.org/10.1371/journal.pcbi.0040006>
- Orjuela, M., Castaneda, V. P., Ridaura, C., Lecona, E., Leal, C., Abramson, D. H., ... Cordon-Cardo, C. (2000). Presence of human papilloma virus in tumor tissue from children with retinoblastoma: An alternative mechanism for tumor development. *Clinical Cancer Research*, 6(10), 4010–4016.
- Pandey, A. N. (2014, October 1). Retinoblastoma: An overview. *Saudi Journal of Ophthalmology*. Elsevier. <https://doi.org/10.1016/j.sjopt.2013.11.001>
- Plomin, R., & Schalkwyk, L. C. (2007). Microarrays. *Developmental Science*, 10(1), 19–23. <https://doi.org/10.1111/j.1467-7687.2007.00558.x>
- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Rodriguez-Galindo, C., Orbach, D. B., & VanderVeen, D. (2015, February 1). Retinoblastoma. *Pediatric Clinics of North America*. W.B. Saunders. <https://doi.org/10.1016/j.pcl.2014.09.014>
- Romero, J. P., Ortiz-Estévez, M., Muniategui, A., Carrancio, S., De Miguel, F. J., Carazo, F., ... Rubio, A. (2018). Comparison of RNA-seq and microarray platforms for splice event detection using a cross-platform algorithm. *BMC Genomics*, 19(1), 703. <https://doi.org/10.1186/s12864-018-5082-2>
- Roskoski, R. (2019, June 1). Properties of FDA-approved small molecule protein kinase

inhibitors. *Pharmacological Research*. Academic Press.
<https://doi.org/10.1016/j.phrs.2019.03.006>

Sedighi, A., & Li, P. C. H. (2014). Fundamentals of Advanced Omics Technologies: From Genes to Metabolites. *Comprehensive Analytical Chemistry*, 63, 25–46.
<https://doi.org/10.1016/B978-0-444-62651-6.00002-7>

Shetty, O. A., Naresh, K. N., Banavali, S. D., Shet, T., Joshi, R., Qureshi, S., ... Desai, S. B. (2012). Evidence for the presence of high risk human papillomavirus in retinoblastoma tissue from nonfamilial retinoblastoma in developing countries. *Pediatric Blood & Cancer*, 58(2), 185–190. <https://doi.org/10.1002/pbc.23346>

Shi, L., Tong, W., Su, Z., Han, T., Han, J., Puri, R. K., ... Fuscoe, J. C. (2005). Microarray scanner calibration curves: Characteristics and implications. *BMC Bioinformatics*, 6(SUPPL. 2), S11. <https://doi.org/10.1186/1471-2105-6-S2-S11>

Sîrbu, A., Kerr, G., Crane, M., & Ruskin, H. J. (2012). RNA-Seq vs Dual- and Single-Channel Microarray Data: Sensitivity Analysis for Differential Expression and Clustering. *PLoS ONE*, 7(12). <https://doi.org/10.1371/journal.pone.0050986>

Smyth, G. K., & Altman, N. S. (2013). Separate-channel analysis of two-channel microarrays: Recovering inter-spot information. *BMC Bioinformatics*, 14(1), 165.
<https://doi.org/10.1186/1471-2105-14-165>

Sung, C. H., & Chuang, J. Z. (2010, September 20). The cell biology of vision. *Journal of Cell Biology*. Rockefeller University Press. <https://doi.org/10.1083/jcb.201006020>

Tarca, A. L., Romero, R., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology*, 195(2), 373–388.
<https://doi.org/10.1016/j.ajog.2006.07.001>

ThermoFisher. (n.d.). GeneChip™ Scanner 3000 7G. Retrieved March 3, 2020, from <https://www.thermofisher.com/order/catalog/product/00-0210#/00-0210>

ThermoFisher. (2018). *GeneChip™ WT PLUS Reagent Kit Manual Target Preparation for*

GeneChip™ Whole Transcript (WT) Expression Arrays User Guide Document Part Number 703174 Publication Number MAN0018137 Revision A.0.

- Tokar, T., Pastrello, C., & Jurisica, I. (2020). GSOAP: A tool for visualisation of gene set over-representation analysis. *Bioinformatics* (Oxford, England).
<https://doi.org/10.1093/bioinformatics/btaa001>
- Warnes, G., BOLKER, B., Bonebakker, L., GENTLEMAN, R., Liaw, W., LUMLEY, T., ...
Warnes, G. R. (2015, January 1). gplots: Various R Programming Tools for Plotting Data.
- Weijun Luo and Cory Brouwer. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29(14), 1830–1831.
<https://doi.org/10.1093/bioinformatics/btt285>
- Xu, X. L., Fang, Y., Lee, T. C., Forrest, D., Gregory-Evans, C., Almeida, D., ... Cobrinik, D. (2009). Retinoblastoma Has Properties of a Cone Precursor Tumor and Depends Upon Cone-Specific MDM2 Signaling. *Cell*, 137(6), 1018–1031.
<https://doi.org/10.1016/j.cell.2009.03.051>
- Xu, X. L., Singh, H. P., Wang, L., Qi, D. L., Poulos, B. K., Abramson, D. H., ... Cobrinik, D. (2014). Rb suppresses human cone-precursor-derived retinoblastoma tumours. *Nature*, 514(7522), 385–388. <https://doi.org/10.1038/nature13813>
- Yunis, J. J. (1983). The chromosomal basis of human neoplasia. *Science*, 221(4607), 227–236.
<https://doi.org/10.1126/science.6336310>
- Zhang, W., Yu, Y., Hertwig, F., Thierry-Mieg, J., Zhang, W., Thierry-Mieg, D., ... Fischer, M. (2015). Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biology*, 16(1), 133. <https://doi.org/10.1186/s13059-015-0694-1>