

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

Escuela de Medicina y Ciencias de la Salud



“Identificación de genes con alteraciones en número de copias asociados a  
sobrevida en diversos tipos de cáncer”

Disertación presentada por

Alejandra Guardado Méndez

Sometida a la

Escuela de Medicina y Ciencias de la Salud

como un requisito parcial para obtener el grado académico de

Maestría

en

Ciencias Biomédicas

Monterrey Nuevo León, 26 de julio de 2021

## **Dedicatoria**

A mi familia.

## **Agradecimientos**

Agradezco especialmente al Dr. Víctor Manuel Treviño Alvarado por darme la oportunidad de pertenecer al grupo de bioinformática y sobre todo por brindarme su apoyo, por la paciencia, por compartir sus conocimientos y consejos durante el desarrollo de mi tesis.

Agradezco a los miembros de mi comité de tesis: Dr. Juan Emmanuel Martínez Ledesma, Dr. Maricruz Sepúlveda Villegas y Dr. Raúl Aguirre Gamboa, primero por aceptar formar parte de mi comité y por las sugerencias brindadas para la realización de este proyecto.

Agradezco a mi familia por su amor incondicional, por apoyarme y motivarme en todo momento.

Agradezco a mis compañeros de maestría y del grupo de bioinformática por hacer más grata mi experiencia en la maestría, por todo su apoyo y por tener la disposición siempre de ayudar.

Agradezco al Tecnológico de Monterrey por la beca de colegiatura y al CONACyT por la beca de mantenimiento otorgada, ya que sin ellas la realización de este proyecto no sería posible.

Agradezco a mis amigos que también los considero familia, por estar conmigo en los momentos de estrés, por su apoyo incondicional y por creer siempre en mí.

## Índice

<b>1. Marco teórico .....</b>	<b>10</b>
1.1 Generalidades del cáncer .....	10
1.2 Epidemiología del cáncer.....	11
<b>2. Antecedentes .....</b>	<b>12</b>
2.1 Caracterización molecular del cáncer .....	12
2.2 Biomarcadores en cáncer .....	15
2.3 Alteraciones en número de copias .....	17
2.4 Identificación de biomarcadores pronósticos .....	19
<b>3. Planteamiento del problema .....</b>	<b>21</b>
<b>4. Pregunta de Investigación.....</b>	<b>21</b>
<b>5. Justificación.....</b>	<b>21</b>
<b>6. Hipótesis .....</b>	<b>21</b>
<b>7. Objetivos.....</b>	<b>22</b>
7.1 Objetivos General.....	22
7.2 Objetivos específicos.....	22
<b>8. Metodología.....</b>	<b>23</b>
8.1 Estrategia General.....	23
8.2 Diseño del estudio.....	23
8.3 Fuente de datos .....	23
8.4 Datos de CNA.....	25
8.5 Procesamiento de los datos de CNAs .....	25
8.6 Análisis de sobrevida .....	25
8.7 Datos y Análisis de Expresión Genética.....	26
8.8 Análisis de firmas genómicas .....	26
8.9 Análisis de Fármacos .....	27
8.10 Análisis de Enriquecimiento .....	28
8.11 Validación en literatura.....	29
<b>9. Resultados .....</b>	<b>30</b>
9.1 Obtención de datos y filtrado.....	30
9.2 Análisis de Sobrevida .....	32
9.3 Resumen de regiones de CNAs asociadas a sobrevida.....	34
9.4 Resultados de la validación en literatura .....	36
9.5 Comparación de las pruebas VALORATE y Log-Rank.....	45

9.6 Análisis de enriquecimiento.....	45
9.7 Análisis de firmas genómicas .....	47
9.8 Comparación de pruebas estadísticas .....	50
<b>10. Discusión.....</b>	<b>52</b>
10.1 Mapas de calor del perfil de CNAs en diferentes cánceres. ....	52
10.2 Análisis de sobrevida .....	53
10.3 Regiones conocidas asociadas a sobrevida .....	54
10.4 Regiones nuevas asociadas a sobrevida.....	57
10.5 Análisis de firmas genómicas .....	59
10.6 Análisis de respuesta a fármacos .....	62
10.7. Comparación de pruebas estadísticas para el análisis de sobrevida. ....	63
10.8 Repositorio.....	64
10.9 Perspectivas futuras .....	64
<b>11. Apéndice .....</b>	<b>65</b>
11.1 Apéndice A .....	65
11.2 Apéndice B .....	67
<b>12. Bibliografía.....</b>	<b>68</b>

## Lista de tablas

Tabla 1	Cohortes de pacientes con cáncer analizados y número de muestras. ....	24
Tabla 2	Bases de datos y herramientas utilizadas para el análisis funcional .....	28
Tabla 3	Cálculo de odds ratio por las herramientas del análisis funcional .....	29
Tabla 4	Genes filtrados en las cohortes de cáncer analizados.....	31
Tabla 5	Resultados búsqueda en PubMed para el análisis “All Amplifications” .....	37
Tabla 6	Resultados búsqueda en PubMed para el análisis “All Deletions” .....	39
Tabla 7	Resultados búsqueda en PubMed para el análisis “Deep Amplifications” .....	41
Tabla 8	Resultados búsqueda en PubMed para el análisis “Deep Deletions” .....	43
Tabla 9	Resultados del análisis de fármacos .....	51
Tabla 10	Resumen de regiones asociadas a sobrevida en cáncer.....	65

## Lista de figuras

Figura 1	Gráfico de Incidencia en diversos tipos de cáncer. ....	11
Figura 2	Datos públicos de pacientes con cáncer disponibles en TCGA .....	12
Figura 3	Mutaciones somáticas por mega base en diferentes cánceres. ....	14
Figura 4	Fracción de puntos de corte de CNAs.....	14
Figura 5	Estrategia general para la identificación de genes con alteraciones en número de copias asociados a sobrevida en diversos tipos de cáncer. ....	23
Figura 6	Obtención de datos y filtrado. ....	30
Figura 7	Mapas de calor de los perfiles de CNAs en pacientes con cáncer .....	31
Figura 8	Manhattan plots de valores $-\log_{10}(p)$ asociados con sobrevida en Cánceres. ...	33
Figura 9	Regiones de CNAs más significativas en el análisis “All Amplifications” .....	34
Figura 10	Regiones de CNAs más significativas en el análisis “All Deletions”. ....	35
Figura 11	Regiones de CNAs más significativas en el análisis “Deep Amplifications” ....	35
Figura 12	Regiones de CNAs más significativas en el análisis “Deep deletions”.....	36
Figura 13	Gráfico de dispersión de los valores p calculados por VALORATE y la prueba log rank.....	45
Figura 14	Resultados de los análisis de enriquecimiento. ....	46
Figura 15	<i>Resultados de firmas genómicas de LGG</i> .....	48
Figura 16	Resultados de firmas genómicas en KIPAN.....	49
Figura 17	Correlación de análisis de sobrevida entre pruebas estadísticas.....	50

Figura 18 Cánceres con correlación diferente entre VALORATE y Cox ..... 51

Figura 19 Mapa de calor de los valores  $p$  estimados en el análisis de sobrevida para regiones alteradas por CNAs en diversos cánceres ..... 67

## ***1. Marco teórico***

### *1.1 Generalidades del cáncer*

El cáncer comprende un grupo de enfermedades heterogéneas que se originan por el crecimiento continuo de células anormales, mismas que tienen la capacidad de invadir tejidos de órganos adyacentes y/o diseminarse a diferentes partes del cuerpo en un proceso denominado metástasis, para formar tumores secundarios (Veiseh et al., 2011).

En la mayoría de los casos, el cáncer se presenta de forma esporádica debido a la acumulación de alteraciones genéticas y epigenéticas en células somáticas individuales (Takeshima & Ushijima, 2019). El estudio de genomas de cáncer secuenciados ha revelado que la mayoría de las alteraciones adquiridas son pasajeras o aparentemente neutrales, mientras que solo un grupo reducido de alteraciones, conocidas como "conductoras" son las que conferirán una ventaja selectiva y conducirán a la expansión clonal (Bozic et al., 2010).

Las alteraciones conductoras generalmente afectan la función de genes supresores de tumores, oncogenes y genes reparadores del DNA (Lahtz & Pfeifer, 2011). La desregulación de estos genes afecta procesos normales e inherentes de las células, como los relacionados con el crecimiento, la proliferación y muerte celular (Derbal, 2017). En consecuencia, la célula adquiere características que le confieren ventajas de crecimiento y supervivencia y termina transformándose en una célula cancerosa (Podlaha et al., 2012). Dependiendo del órgano y tipo celular donde las alteraciones se originen, se pueden desarrollar diferentes tipos de cáncer, a la fecha se han descrito más de 200 tipos de cáncer (Song et al., 2015).

En esta enfermedad intervienen además varios factores de riesgo, los más comunes se relacionan con hábitos de comportamiento; consumo de alcohol, tabaco, la inactividad física y el tipo de alimentación, otros se relacionan con factores intrínsecos de los individuos (ej. inflamación, envejecimiento, etc.) y/o con factores hereditarios (predisposición genética). No obstante, se estima que la mitad de los casos y muertes por este padecimiento podrían ser prevenibles si se adoptaran estilos de vida saludable y evitaran los factores de riesgo conocidos (Song et al., 2015).

## 1.2 Epidemiología del cáncer

El cáncer representa uno de los desafíos de salud más importantes a nivel mundial. Solo precedido por enfermedades cardiovasculares, el cáncer es la segunda causa de mortalidad prematura en el mundo. Actualmente, alrededor de 43,8 millones de personas viven con un diagnóstico de cáncer (WHO, 2018).

En 2018, la Agencia Internacional para Investigación del Cáncer registró 18.1 millones de nuevos casos de cáncer y 9.6 millones de defunciones a causa de éste (Cai & Liu, 2021).

A nivel mundial, los cánceres de mama, pulmón, colorrectal, próstata y estómago comprenden casi la mitad de los cánceres diagnosticados. El cáncer de mama ocupa el primer lugar en términos de incidencia, con una contribución del 11,7% al total de nuevos casos de cáncer registrados por año, mientras que el cáncer de pulmón sigue siendo la principal causa de mortalidad seguido de los cánceres de próstata, colorrectal, hígado, estómago y mama. En la figura 1, se muestra en porcentajes la incidencia para diferentes cánceres (Sung et al., 2021).

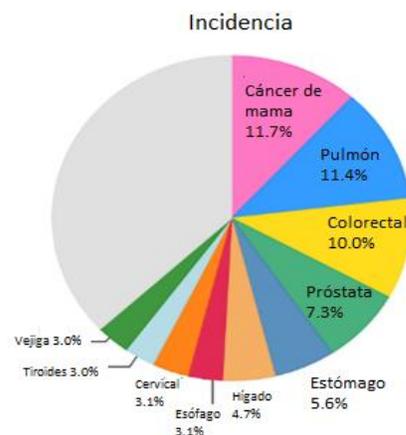


Figura 1 Gráfico de Incidencia en diversos tipos de cáncer (Sung et al., 2021)

En países en vías en desarrollo como México, el número de muertes por cáncer contribuye en más de un 60% al total de muertes registrado anualmente. En México, el

cáncer es la tercera causa de muerte con alrededor de 90, 222 muertes estimadas en 2020 y de acuerdo con proyecciones hechas por GLOBOCAN, el número de muertes será casi el doble para 2040 (WHO, 2021).

## 2. Antecedentes

### 2.1 Caracterización molecular del cáncer

Durante los últimos 15 años, el conocimiento molecular y genético de los tumores se ha expandido gracias a la reducción de los costos y mayor accesibilidad a tecnologías de alto rendimiento (NGS) las cuáles han permitido el análisis del exoma, genoma y transcriptoma de miles de pacientes con diversos tipos de cáncer (Tomczak et al., 2015).

Bases de datos como The Cancer Genome Atlas (TCGA) y The International Cancer Genome Consortium (ICGC) a través de diferentes plataformas de análisis han caracterizado millones de alteraciones somáticas encontradas en muestras tumorales, incluidos datos de mutación, metilación, alteraciones en el número de copias (CNAs) y expresión de ARNm, micro-ARN y ARN largos no codificantes, estos datos se resumen en la figura 2. (Chandran et al., 2016). De estos proyectos, TCGA fue el primer esfuerzo público con la colección más grande de datos multi-ómicos iniciado en 2006 con el análisis de tres de los cánceres más agresivos; glioblastoma, cáncer pulmón, y ovario (Tomczak et al., 2015). El proyecto más tarde se extendió al análisis de 30 tipos de cáncer, generando cerca de 1.2 petabytes de datos finales (Chandran et al., 2016).

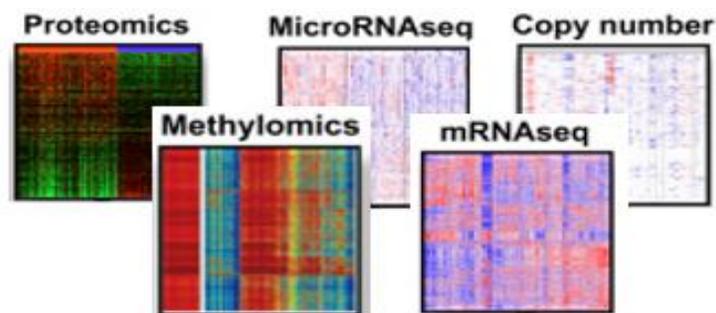


Figura 2 Datos públicos de pacientes con cáncer disponibles en TCGA.

El análisis de los datos generados y otros estudios en cáncer han revelado que un gran número de alteraciones somáticas caracterizan el genoma de tumores, no obstante, solo una fracción de estas alteraciones conducen la transformación de células normales a malignas (Bozic et al., 2010). Inicialmente se identificaron las alteraciones conductoras que, debido a la selección positiva ejercida, eran observadas con frecuencia en múltiples tumores, la mayoría de éstas eran responsables directas del desarrollo de cáncer (Bányai et al., 2021).

Otros estudios basados en NGS mostraron que, para la mayoría de los tipos de cáncer, su genoma contiene un número muy pequeño de genes conductores alterados en un alto porcentaje (>20%) de tumores y un número mucho mayor de genes conductores alterados con muy poca frecuencia que pese a ello eran capaces de tener un efecto fuerte en el fenotipo (Bányai et al., 2021). De modo que muchos genes conductores eran genes raros o putativos pero necesarios para explicar el resto de la carcinogénesis (Waks et al., 2016).

Los datos obtenidos a partir de células de cáncer han generado un mayor entendimiento de la oncogénesis, del desarrollo de metástasis y de la naturaleza de tumores complejos. Estos además han abierto la posibilidad de identificar alteraciones conductoras o biomarcadores que sirvan para estratificar mejor a los pacientes y desarrollar terapias más efectivas con potencial de aumentar el pronóstico y calidad de vida de pacientes con cáncer.

Ahora se sabe que muchas mutaciones y CNA son encontradas en células de cáncer, la figura 3 muestra como la frecuencia de mutaciones por megabase varía en los diferentes cánceres. (Büttner et al., 2019), al igual que la frecuencia de CNAs en la figura 4 (Alaei-Mahabadi et al., 2016). Actualmente se han reconocido cerca de 500 genes causales de cáncer (conductores) incluidos genes alterados por variantes de nucleótido simple (SNV), translocaciones y CNAs (Martínez-Jiménez et al., 2020; Tamborero et al., 2018).

También se han encontrado patrones epigenéticos que caracterizan los genomas de cáncer (Lakshminarasimhan & Liang, 2016), entre ellos; la pérdida global de metilación y la inactivación por hipermetilación de genes supresores de tumor. Además de procesos mutacionales, no muy conocidos, capaces de moldear los genomas de cáncer en eventos catastróficos únicos denominados kataegis, cromoplexia y cromotripsis (Willis et al., 2015)

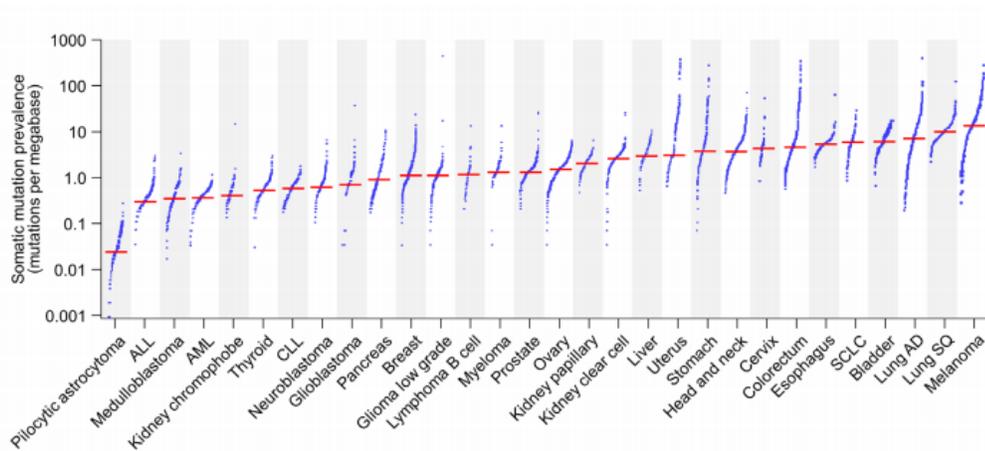


Figura 3 Mutaciones somáticas por mega base en diferentes cánceres. Tomado de Buttner et al.

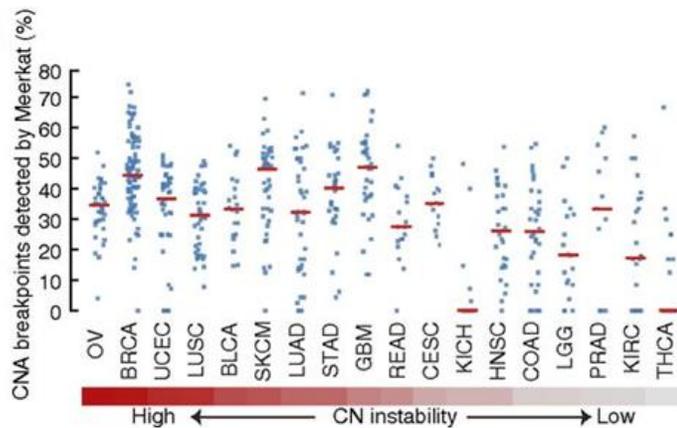


Figura 4 Fracción de puntos de corte de CNAs. Tomado de Alaei-Mahabadi et al.

Conforme se han estudiado los genomas de cáncer se ha vuelto evidente la gran diversidad molecular y fenotípica de los tumores tanto entre pacientes (intertumoral) como entre las células individuales del tumor (intratumoral). Esta heterogeneidad permite que al tumor se adapte a las constantes restricciones del ambiente y evolucione con él tiempo, generando en ciertos casos la presencia de recurrencia, susceptibilidad o resistencia a

tratamientos y por ende predisponiendo a los pacientes a peores pronósticos (Zito Marino et al., 2019).

## 2.2 Biomarcadores en cáncer

Desde hace décadas los biomarcadores han contribuido al progreso de la medicina contra el cáncer ofreciendo diagnósticos y pronósticos cada vez más precisos, así como tratamientos más efectivos. El término “biomarcador” hace referencia a cualquier biomolécula que puede medirse de manera objetiva y sirve como indicador del estado fisiológico o de salud del paciente (Goossens et al., 2015). De acuerdo con su utilidad los biomarcadores suelen clasificarse en diagnósticos (detección temprana), predictivos (respuesta a tratamiento) y pronósticos (agresividad del cáncer) (Chatterjee & Zetter, 2005).

Anteriormente la examinación del cáncer se basaba solo en las características histopatológicas del tumor y clínicas del paciente, ahora es cada vez más frecuente el empleo de biomarcadores moleculares, ya que estos ofrecen una atención más personalizada y suelen hacer mejores predicciones. Estos biomarcadores pueden presentarse a nivel de DNA, RNA, proteína y metabolitos. (Goossens et al., 2015)

Gran parte de los biomarcadores en cáncer que se utilizan actualmente en la clínica son proteínas encontradas en suero de pacientes, por ejemplo, el antígeno específico de próstata (PSA) (Chatterjee & Zetter, 2005), éste es un biomarcador diagnóstico usado desde hace tiempo en la detección temprana de cáncer de próstata. De igual forma la proteína C-reactiva del suero sirve como biomarcador pronóstico de varias neoplasias (Shrotriya et al., 2018). Estos biomarcadores son usualmente más estables que los ácidos nucleicos y fácilmente accesibles (Xi et al., 2017).

Los biomarcadores de RNA comparados con los de proteínas son más sensibles y específicos, además proveen información dinámica del estado fisiológico de las células de cáncer (Xi et al., 2017). También han demostrado ser útiles para clasificar subtipos de cáncer y para predecir pronóstico o respuesta a tratamientos. En cáncer de mama, por ejemplo, mediante microarreglos de cDNA, Perou y Sorlie *et al.* (2000) (Perou et al., 2000), consiguieron clasificar los tumores de estos pacientes en cinco subtipos moleculares:

luminal A, luminal B, HER2, tipo basal y tipo normal, cada uno con diferente patrón de expresión génica y pronóstico.

De los biomarcadores de DNA se han descrito muchos biomarcadores basados en mutaciones puntuales, uno muy conocido es la mutación V600 en el gen *BRAF* la cual se presenta en casi la mitad de los casos de melanomas malignos y predice mayor sobrevida con las terapias dirigidas *BRAF* y *MEK* (Eroglu & Ribas, 2016). Estos biomarcadores sobresalen como biomarcadores predictivos de terapias dirigidas. No obstante, la mayoría de los marcadores de mutaciones solo están presentes en porcentajes bajos (Carr et al., 2016)

Además de mutaciones, en años más recientes creció el interés por los biomarcadores basados en la metilación del ADN y CNAs, ambos son buenos indicadores de pronóstico en cáncer e incluso más prometedores que los biomarcadores de ARN ya que en términos de muestras son alteraciones más estables y robustas que los de expresión génica (Pan et al., 2019; Son et al., 2011). Las metilaciones del DNA se presentan en etapas tempranas al desarrollo del cáncer, por lo que también podrían aplicarse como biomarcadores diagnósticos en pacientes asintomáticos. De estos biomarcadores, se aprobó por la FDA la hipermetilación del promotor de Septina-9 para la detección de etapas tempranas de cáncer colorrectal con alta sensibilidad (Ma et al., 2019).

De los biomarcadores basados en CNAs en varios estudios se ha reconocido su valor pronóstico. Estos se correlacionan con progresión en varias neoplasias y para algunas de ellas la alta carga de CNAs predice cánceres más agresivos y con mayor recurrencia (Hieronymus et al., 2018). En un estudio publicado en 2018, para la identificación de biomarcadores pronóstico en cáncer, usando perfiles de mutaciones y CNAs, se encontró que estos biomarcadores actuando en los mismos genes conductores que mutaciones, proporcionaban más información sobre el pronóstico de los pacientes y mantenían un poder pronóstico más significativo (Smith & Sheltzer, 2018). Estas alteraciones además han mostrado ser capaces de estratificar pacientes con cáncer y asociarlos con diferente evolución clínica.

### *2.3 Alteraciones en número de copias*

Los CNAs representan una de las alteraciones más comunes encontradas en las células de cáncer (Hieronymus et al., 2018). Estas alteraciones se originan como resultado de la inestabilidad genómica y defectos durante la división mitótica.

Los CNA son variaciones estructurales que involucran la pérdida o ganancia de grandes extensiones de DNA. Estos eventos contribuyen al inicio y progresión del cáncer, a menudo a través de cambios inapropiados en la expresión de oncogenes y genes supresores de tumor, los cuáles confieren ventajas proliferativas o metastáticas (Bhattacharya et al., 2020).

Según su tamaño se puede discernir entre CNA amplios y focales. Los CNAs amplios ocurren con mayor frecuencia, éstos abarcan todo el cromosoma o por lo menos más del 25% del brazo cromosómico mientras que los CNAs focales son regiones de menor tamaño, éstos tienen una extensión mayor a 1 kb y hasta varias Mb, su tamaño facilita la identificación de los genes alterados (Zack et al., 2013).

La identificación de CNAs en muestras de tumor se ha realizado comúnmente a partir de datos de microarreglos de hibridación genómica comparativa o arreglos de SNP y en años recientes se ha ampliado al uso de tecnologías de NGS (Taylor et al., 2008). Para poder estimar el número de copias, los datos obtenidos a través de estas tecnologías suelen ser procesados por algoritmos de segmentación, en éstos el genoma es dividido en segmentos individuales formados por secuencias de sondas contiguas que tienen una señal de intensidad o profundidad de lectura similar. Los límites de los segmentos se definen a través de un valor umbral especificado y dependiendo la relación  $\log_2$  de los segmentos con respecto al genoma diploide normal se decide si corresponden a regiones de amplificación o delección (Esteves et al., 2020).

Así mismo se pueden usar diferentes algoritmos que utilizan datos segmentados para distinguir los CNAs que tienen alta probabilidad de contribuir a la patogénesis del cáncer. Uno ampliamente utilizado es el de Identificación genómica de regiones significativas en cáncer (GISTIC 2.0), éste calcula una prueba estadística (G-score) donde la frecuencia de la aberración en múltiples muestras es multiplicada por la amplitud

promedio de ésta. La significancia de cada loci observado se determina usando una corrección por *false Discovery Rate* (FDR). (Mermel et al., 2011)

Estas herramientas bioinformáticas han permitido realizar estudios a gran escala e identificar algunos CNAs que proveen información relevante en relación con cáncer. Los primeros estudios sobre CNAs estuvieron enfocados en entender como estos eventos propiciaban el cáncer, en la identificación de genes conductores y en describir los patrones de distribución de CNAs en diferentes tumores.

Beroukhim R., et al en 2010, en un estudio pan-cáncer que incluía 3,131 perfiles de CNAs, identificaron 158 eventos focales recurrentes, en éstos las amplificaciones y deleciones tenían una frecuencia de incidencia similar. Los eventos más frecuentes fueron la amplificación de *MYC* y la deleción de *CDKN2A/CDKN2B*. La mayoría de los eventos identificados fueron observados en múltiples tumores, no obstante, muchos genes contenidos en ellos no eran conocidos en cáncer. (Beroukhim et al., 2010)

En 2013, Travis I. Zack et al, analizaron el perfil de CNAs de 4,934 especímenes pertenecientes a 11 diferentes cánceres. Ellos observaron que la duplicación del genoma completo ocurría en 37% de los cánceres y que ésta se asociaba con tasas altas de CNAs. Además, mostraron que los eventos ubicados dentro de los cromosomas tendían a ser más cortos que los encontrados en telómeros y que los genes afectados por amplificaciones focales estaban más asociados a ciclo celular, expresión y proliferación. (Zack et al., 2013)

Aunque con frecuencia los CNAs se relacionan con eventos oncogénicos debido a los cambios en la expresión de los genes que afectan, éstos no siempre exhiben un cambio de expresión proporcional con su número de copias, por lo que entender la contribución de estos eventos en la carcinogénesis no siempre es clara. Vazquez-Mena O, et al 2012, en líneas celulares de cáncer cervical, mostraron que de las amplificaciones 3q y 5p, solo 15.6% de los genes afectados se correlacionaban con cambios de expresión, de éstos no todos estaban sobreexpresado, cerca del 19% de los genes estaban silenciados. (Vazquez-Mena et al., 2012) Dentro de las posibles explicaciones se han sugerido los silenciamientos epigenéticos, dominios de cromatina y la regulación por retroalimentación negativa.

Adicionalmente se ha propuesto un modelo denominado “Cancer Gene Island” que explica un mecanismo por el cual los CNAs contribuyen al desarrollo del cáncer en donde cambios sutiles en el número de copias de múltiples genes sensitivos a dosis génica dentro de una misma región son seleccionadas por deleciones focales hemicigotas para maximizar las ventajas proliferativas de los genes a través de una haploinsuficiencia acumulativa. (Solimini et al., 2012)

En años recientes, han crecido las investigaciones de CNAs enfocadas en la identificación de nuevos subtipos moleculares de cáncer, en la predicción de recurrencia, en su papel en la evolución de los tumores y en la generación de resistencia terapéutica.

#### *2.4 Identificación de biomarcadores pronósticos*

Algunos CNAs pueden ayudar a pronosticar el tiempo de vida de los pacientes y tomar mejores decisiones sobre tratamientos, (Smith & Sheltzer, 2018) por lo que su identificación es un tema central en investigación. Algunas de estas alteraciones con utilidad clínica se presentan en una proporción baja a intermedia de tumores. Por ejemplo, duplicaciones en la región que incluye a ERBB2 un oncogén que codifica para el factor de crecimiento epidermal humano se presenta en 10-20% en tumores de cáncer de mama (Marotta et al., 2012), pero solo en 1.9-14.3% en cáncer de pulmón no microcítico (NSCLC) y es predictor pronóstico de alto riesgo (Siena et al., 2018). Otras alteraciones en cambio se presentan en una proporción alta de tumores, por ejemplo, la Co-delección 1p/19q observado en 50-80% de los pacientes con oligodendrogliomas y se asocia a evolución clínica favorable (Li et al., 2014).

Los biomarcadores pronósticos basados en perfiles moleculares se determinan mediante la asociación de características genómicas con la sobrevida de pacientes (Hazra & Gogtay, 2017). Los análisis de supervivencia se utilizan para probar la significancia de estas asociaciones y priorizar aquellas con posible relevancia clínica. Comúnmente se identifican dos grupos de pacientes, los que tienen la característica de interés y los que no. En cada caso se evalúa el tiempo que transcurre desde el inicio del estudio hasta la ocurrencia de un evento, por lo regular la muerte del paciente.

El log-rank es la prueba estadística más comúnmente empleada para comparar la sobrevida entre grupos de pacientes (Bewick et al., 2004). La mayoría de las pruebas log-rank utilizan una aproximación asintótica (distribución  $X^2$ ) la cual requiere que las poblaciones a comparar sean de tamaño similar y el número de eventos sea grande (Vandin F., et al., 2015). En cáncer, muchas alteraciones a analizar se presentan en pocos pacientes, generando grupos de distinto tamaño. Por consiguiente, la prueba log-rank bajo la hipótesis nula no sigue una distribución normal o  $X^2$  y esto puede conducir a inexactitudes en la estimación de los valores  $p$  y por tanto a descubrimientos falsos (Vandin et al., 2015)

La inexactitud de la prueba log-rank se ha documentado en varios estudios, Wang R., et al. (2010) (Wang et al., 2010) utilizaron datos simulados demostraron que el error tipo I del log-rank era más distorsionado cuando consideraban poblaciones pequeñas o muy desbalanceadas.

Asimismo, diversas implementaciones de la prueba log-rank test se han propuesto para realizar estimaciones más precisas de la distribución nula a fin de encontrar asociaciones realmente significativas. Vandin *et al.* (2015) (Vandin et al., 2015) por ejemplo, propusieron ExaLT un algoritmo que utiliza una distribución permutacional exacta apropiada para poblaciones no balanceadas.

Aún con ExaLT, Treviño & Tamez-Pena (2017) (Treviño & Tamez-Pena, 2017; Vandin et al., 2015) demostraron que, para datos genómicos en cáncer, ExaLT no es práctico porque tarda mucho en hacer la aproximación. Los autores desarrollaron VALORATE, un algoritmo diseñado para el análisis de sobrevida en cáncer, idóneo en casos donde las poblaciones a analizar están desbalanceadas.

El uso de métodos precisos y rápidos para la identificación in-silico de posibles biomarcadores es de suma importancia, ya que los genes alterados que son candidatos requieren ser evaluados funcionalmente, esto a menudo conlleva procesos que son costosos y consumen mucho tiempo. Además, por lo anterior, el uso de log-rank asintótico tiende a sugerir genes candidatos que no son lo suficientemente robustos (Goossens et al., 2015).

### ***3. Planteamiento del problema***

Una de las alteraciones más comunes en las células de cáncer son las alteraciones en número de copias (CNAs) éstas contribuyen al inicio y progresión del cáncer. Algunos CNAs pueden ayudar a pronosticar y tomar mejores decisiones sobre tratamientos por lo que su identificación es un tema central en investigación. No obstante, los métodos que se emplean para la identificación de CNAs pronósticos no son exactos ya que las asunciones requeridas por su aplicación no siempre se cumplen en cáncer, esto hace que se detecten alteraciones candidatas poco robustas.

### **4. Pregunta de Investigación**

¿Cuáles genes con alteraciones en número de copias que afectan el tiempo de sobrevida en los distintos tipos de cáncer?

### **5. Justificación**

El cáncer es la segunda causa de muerte en el mundo con un estimado de 9.6 millones de muertes en 2018. Los biomarcadores ayudan a pronosticar y tomar decisiones sobre tratamientos. Los métodos empleados para identificar CNA pronósticos no son exactos porque a menudo se violan las asunciones requeridas en su aplicación. El empleo de métodos más precisos para el análisis de CNA es un área de oportunidad para mejorar el pronóstico y reducir las cifras de mortalidad.

### ***6. Hipótesis***

- El uso de VALORATE en datos de CNA sugerirá diferentes genes con alteraciones en número de copias asociados a sobrevida en diversos tipos de cáncer

## **7. Objetivos**

### 7.1 Objetivos General

- Identificar mediante una prueba más precisa (VALORATE y Cox) diferentes genes con alteraciones en número de copia asociados a sobrevida en diversos tipos de cáncer.

### 7.2 Objetivos específicos

1. Obtener datos de TCGA de 37 subtipos de cáncer
2. Filtrar genes que no contienen información útil
3. Realizar el análisis de sobrevida usando VALORATE
4. Identificar diferencias contra la prueba log-rank
5. Identificar los genes afectados por CNA más significativos
6. Corroborar el tipo de CNA con su expresión diferencial esperada.
7. Determinar genes ya reportados y nuevas asociaciones

## 8. Metodología

### 8.1 Estrategia General

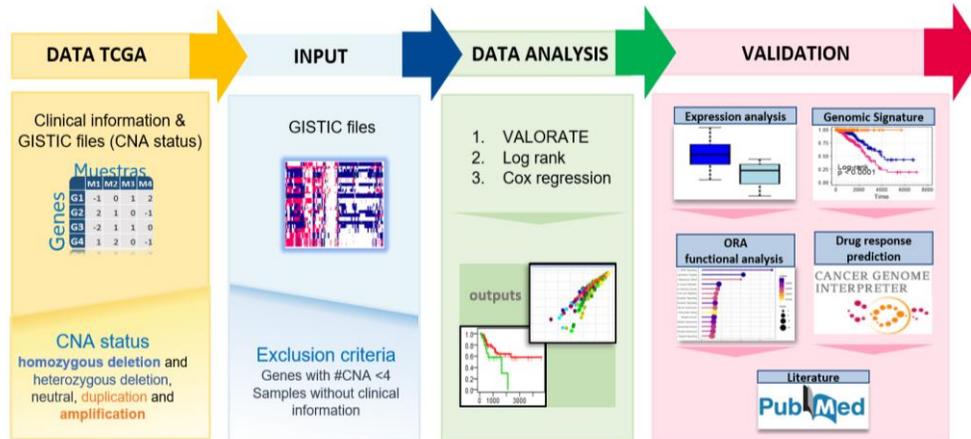


Figura 5 Estrategia general para la identificación de genes con alteraciones en número de copias asociados a supervivencia en diversos tipos de cáncer.

### 8.2 Diseño del estudio

Este es un estudio de investigación observacional comparativo, transversal y retrospectivo.

### 8.3 Fuente de datos

Se obtuvo la información clínica, datos de CNAs y datos de expresión de cohortes de pacientes asociados con 37 diferentes cánceres de la base de datos TCGA. Los datos se descargaron directamente del sitio web Firebrowse (<http://firebrowse.org/>). En total se reunió información de ~ 11, 000 pacientes y 24,776 genes. Los pacientes analizados que, si tenían información de supervivencia declarada, los tipos de cáncer y el % de censurados, se muestran en la Tabla 1.

Tabla 1 Cohortes de pacientes con cáncer analizados y número de muestras.

<b>Código Cáncer</b>	<b>Nombre del Cáncer</b>	<b>Número original de Muestras</b>	<b>Muestras analizadas</b>	<b>Pacientes censurados (%)</b>
<b>ACC</b>	Adrenocortical carcinoma	100	89	67%
<b>BLCA</b>	Bladder Urothelial Carcinoma	409	398	73%
<b>BRCA</b>	Breast invasive carcinoma	1038	1025	90%
<b>CESC</b>	Cervical squamous cell carcinoma and endocervical adenocarcinoma	288	277	79%
<b>CHOL</b>	Cholangiocarcinoma	46	35	54%
<b>COAD</b>	Colon adenocarcinoma	356	345	84%
<b>COADREAD</b>	Colorectal adenocarcinoma	480	409	86%
<b>DLBC</b>	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	57	46	89%
<b>ESCA</b>	Esophageal carcinoma	173	162	65%
<b>GBM</b>	Glioblastoma multiforme	583	572	24%
<b>GBMLGG</b>	Glioma	1091	1080	51%
<b>HNSC</b>	Head and Neck squamous cell carcinoma	528	517	67%
<b>KICH</b>	Kidney Chromophobe	76	65	87%
<b>KIPAN</b>	Pan-kidney cohort (KICH+KIRC+KIRP)	864	853	76%
<b>KIRC</b>	Kidney renal clear cell carcinoma	531	520	69%
<b>KIRP</b>	Kidney renal papillary cell carcinoma	279	268	88%
<b>LAML</b>	Acute Myeloid Leukemia	179	168	38%
<b>LGG</b>	Brain Lower Grade Glioma	519	508	81%
<b>LIHC</b>	Liver hepatocellular carcinoma	349	338	73%
<b>LUAD</b>	Lung adenocarcinoma	489	478	74%
<b>LUSC</b>	Lung squamous cell carcinoma	439	478	67%
<b>MESO</b>	Mesothelioma	97	86	33%
<b>OV</b>	Ovarian serous cystadenocarcinoma	570	559	46%
<b>PAAD</b>	Pancreatic adenocarcinoma	191	180	63%
<b>PCPG</b>	Pheochromocytoma and Paraganglioma	173	162	96%
<b>PRAD</b>	Prostate adenocarcinoma	503	492	96%
<b>READ</b>	Rectum adenocarcinoma	135	124	92%
<b>SARC</b>	Sarcoma	266	254	70%
<b>SKCM</b>	Skin Cutaneous Melanoma	369	358	59%
<b>STAD</b>	Stomach adenocarcinoma	405	394	76%
<b>STES</b>	Stomach and Esophageal carcinoma	567	556	74%
<b>TGCT</b>	Testicular Germ Cell Tumors	144	133	97%
<b>THCA</b>	Thyroid carcinoma	506	495%	97%
<b>THYM</b>	Thymoma	132	121	95%
<b>UCEC</b>	Uterine Corpus Endometrial Carcinoma	546	535	91%

<b>UCS</b>	Uterine Carcinosarcoma	66	56	43%
<b>UVM</b>	Uveal Melanoma	91	80	83%

#### 8.4 Datos de CNA

Para cada cáncer se descargaron los datos de CNAs que fueron determinados GISTIC 2.0 (Plataforma: Affymetrix Genome-wide Human SNP 6.0, datos de nivel 4). El estatus de CNA para los genes individuales en cada muestra se determinó mediante el conjunto de llamadas de números de copias discretas proporcionadas por GISTIC (g); (-1) deleción heterocigota, (-2) deleción homocigota, (0) normal, (1) duplicación y (2) amplificación profunda.

#### 8.5 Procesamiento de los datos de CNAs

De los datos de CNAs se consideraron 4 análisis dependiendo el tipo de variación y magnitud de CNA analizada. En este estudio el análisis “All amp”, se refiere a duplicaciones de material genético de una copia o más por arriba del esperado ( $g \geq +1$ ), el análisis “All del” se refiere a deleciones de material genético de una copia o más por abajo del esperado ( $g \leq -1$ ), el análisis “Deep Amp” ( $g \geq +2$ ) se refiere a la ganancia de dos copias y el análisis “Deep del” ( $g = -2$ ) a la pérdida de ambas copias de una región genómica. Con base a estos análisis se construyeron 4 matrices de gen por muestra con información sobre la presencia y ausencia de la alteración de interés según la variación de CNA analizada.

#### 8.6 Análisis de sobrevida

El análisis de sobrevida se realizó utilizando tres pruebas estadísticas: VALORATE, log rank y el modelo Cox. Estos análisis se realizaron en R utilizando el paquete VALORATE y Survival ambos disponibles en CRAN. Para cada cáncer se consideraron 4 análisis según la variación de CNA analizada. El análisis de sobrevida se realizó para cada gen, descartando aquellos que estaban alterados en menos de 4 pacientes. Como datos de

entrada para el análisis se usaron las matrices de CNAs y la información clínica de los pacientes. De las matrices de CNAs previamente construidas se escogieron a los pacientes con la alteración de interés y con la matriz de GISTIC ( $g=0$ ) se escogieron a los pacientes no alterados.

Después se obtuvieron gráficos de las curvas de sobrevida mediante el método Kaplan-Meier. Se definieron como posibles asociaciones significativas aquellas con un valor  $p < 0.05$ . Además, se obtuvo información sobre las regiones genómicas analizadas utilizando el paquete BioMart.

### *8.7 Datos y Análisis de Expresión Genética*

Para cada cáncer se descargaron datos de RNA-seq (Plataforma: secuenciación de ARN de Illumina HiSeq 2000, nivel 3). Estos fueron normalizados por el método RSEM por lo que solo se transformaron mediante  $\log_{10}(x + 1)$ . Para cada cáncer se realizaron análisis de expresión diferencial considerando los dos niveles de amplificación y delección de CNAs previamente descritos. El análisis se realizó por gen usando la prueba estadística Wilcoxon y dividiendo a los pacientes en dos grupos; pacientes con la alteración de interés y pacientes no alterados.

### *8.8 Análisis de firmas genómicas*

Para determinar un posible marcador compuesto de todas las alteraciones significativas y consistentes con la expresión en cada tipo de cáncer, se construyó una “firma genómica”. Para este análisis se consideraron las regiones encontradas significativas en el análisis de sobrevida con VALORATE. Para cada cáncer se construyeron firmas genómicas según el tipo de variación analizada y firmas que combinaban regiones de amplificación y delección, en estas últimas se evaluaron además diferentes formas de agrupar a los pacientes. Todas las firmas se realizaron considerando solo con los genes asociados a sobrevida y otra en la que además los genes tenían una expresión diferencial significativa.

Para construir las firmas se observó el *Hazard Ratio* (HR) de cada uno de los genes seleccionados, si éste era mayor a 1 se definió como un gen de alto riesgo y menor a 1 como un gen de bajo riesgo. Los genes que no entraban en estas categorías se definieron como no significativos. Si la lista de genes era mayor a 800 entonces solo se escogían los primeros 400 genes de cada grupo de riesgo. De los genes seleccionados, para cada paciente se calculó la suma del HR de sus genes de alto y bajo riesgo. Los pacientes se clasificaron en el grupo de riesgo donde la suma de sus HR era mayor. Para hacer comparables ambas sumas, se dividió 1 entre la suma de HR de los genes de bajo riesgo.

En las firmas genómicas que combinaban regiones de amplificación y delección, se mantuvieron las sumas de HR del grupo de riesgo al cual los pacientes fueron asignados en las firmas individuales. Con esta información primero se construyó una firma genómica en la cual ambas sumas de HR se compararon y se escogió la de mayor valor para poder elegir el nuevo grupo de riesgo al cuál el paciente sería asignado. Después se construyó una segunda firma en la cual los pacientes se agruparon por la combinación de ambos grupos de riesgo.

Por último, se intentó agrupar a los pacientes mediante métodos de agrupamiento jerárquico. Para esto, primero se creó una matriz de gen por muestra con la información sobre la presencia y ausencia de amplificaciones y delecciones en los pacientes. Sobre esta matriz se calculó la distancia euclidiana y se generaron clústeres jerárquicos (método “ward.D2”), se usó el paquete “ComplexHeatmap” en R. Además, se utilizó la función *cutree* para cortar el dendrograma en 6 grupos de pacientes ( $k = 6$ ). Los grupos de pacientes formados por las firmas se describieron con mapas de calor utilizando el paquete “ComplexHeatmap” y el análisis de las firmas mediante curvas de Kaplan-meier utilizando el paquete “Survminer”.

## 8.9 Análisis de Fármacos

Para identificar si una alteración tiene un fármaco potencial ya reportado, se realizó un análisis en bases de datos. El análisis de fármacos se realizó a través una interfaz gráfica implementada como un paquete de Python de la base de datos The Cancer Genome

Interpreter (CGI) (Tamborero et al., 2018). Como datos de entrada se utilizó una lista de genes con su respectiva variación, ésta se comparó contra la evidencia clínica para el cáncer correspondiente. En esta lista se incluyeron los 100 genes más significativos de cada análisis de sobrevida realizado por tipo de variación y que además tenían una expresión diferencial significativa ( $p < .05$ ). Esta base de datos no diferencia la profundidad de la variación, por tanto, todas las alteraciones fueron analizadas como amplificación o deleción. De los resultados se tomaron en cuenta los estudios aprobados por la Food and Drug Administration (FDA) o la National Comprehensive Cancer Network (NCCN) y los estudios avalados por ensayos clínicos, preclínicos o reporte de casos.

### 8.10 Análisis de Enriquecimiento

El análisis de enriquecimiento se realizó considerando los genes más significativos en VALORATE y que tenían una expresión diferencial significativa ( $p < 0.05$ ). En la siguiente tabla se muestran los paquetes en R y bases de datos utilizadas:

Tabla 2 Bases de datos y herramientas utilizadas para el análisis funcional

Paquetes de R	Bases de datos
WebGestaltR	"geneontology_Biological_Process_noRedundant" "geneontology_Molecular_Function_noRedundant" "geneontology_Cellular_Component_noRedundant" "pathway_KEGG", "pathway_Reactome", "pathway_Wikipathway", "disease_Disgenet"
ReactomePa	ReactomePa
rWikiPathways	RWikiPathways
ClusterProfiler	“GO”, “KEGG”, ReactomePa, rWikiPathways
Gprofiler2	“GO”, “REAC”, “WP”
enrichr	"Elsevier_Pathway_Collection" "KEGG_2019_Human" "WikiPathways_2019_Human" "BioPlanet_2019" "InterPro_Domains_2019"

Para realizar el análisis funcional se mantuvieron los parámetros por default de cada una de las herramientas. Algunas de estas herramientas no proporcionaban el *odds ratio* (OR) de forma directa, por lo que en la tabla 3 se muestra el cálculo para cada una de ellas:

Tabla 3 Cálculo de odds ratio por las herramientas del análisis funcional

Paquete	Herramientas	Odds ratio
<b>WebGestaltR</b>	1.-WebGestaltR ()	Se mantuvo el enrichment ratio proporcionado, el cual equivale al overlap/expected. $\text{overlap} (\text{count}/\text{gene set size})$ .
<b>Gprofiler2</b>	2. gost()	$\text{Intersection\_size} (\text{gene count})$ , $\text{query\_size}(\text{geneList})$ $\text{term\_size} (\text{gene set size})$ , $\text{effective\_domain\_size} (\text{gene background size})$
<b>enrichR</b>	3. enrichr()	Si proporcionaba el odds ratio.
<b>ReactomePa</b>	4.enrichPathway ()	Del $\text{geneRatio} = \text{count}/\text{geneList}$ , se obtuvo información del count y geneList
<b>ClusterProfiler</b> <b>/rwikipathways</b>	5. enrichGO() 6. enrichKEGG() 7. enricher()	Del $\text{BgRatio} = \text{gene set size}/\text{background gene list}$ se obtuvo información del gene set size y background size
	Cálculo de riesgo relativo	<b>Odds ratio</b> $(a/b)/(c/d)$ a (gene count), b (geneList - count), c (gene set size - count), d (gene background size - (a + b + c)).

### 8.11 Validación en literatura

Para determinar si las regiones eran conocidas o poco conocidas, para cada cáncer se realizó una búsqueda en PubMed de las 12 regiones más significativas de cada análisis de sobrevida realizado por tipo de variación. La búsqueda se realizó por regiones y genes usando diferentes combinaciones de términos asociados (ej. Gen [TIAB], Gen [TIAB] And CANCER\*[TIAB] OR TUMOR\*[TIAB], etc.). Los genes/regiones con menos de 4 registros se definieron como nuevos o poco conocidos y el resto, como conocidos.

## 9. Resultados

### 9.1 Obtención de datos y filtrado

Para poder realizar los análisis de sobrevida, primero se filtraron los genes que estaban alterados por CNAs en menos de 4 muestras, esto de acuerdo con el mínimo requerido de alteración por las pruebas estadísticas. En la figura 6(A) se muestra un ejemplo del perfil original de CNAs en la cohorte pacientes con glioblastomas (GBM) junto con su distribución de genes con amplificaciones en la Figura 6 (B) y en la figura 7 se muestra un resumen de los perfiles originales de CNAs para otros cánceres.

Antes de realizar el filtrado de genes, cada perfil de CNAs de las diferentes cohortes de cáncer fueron separados en 4 análisis según el tipo de variación y la magnitud de CNA analizada. A partir de los datos de CNAs individuales se realizó el filtrado de genes. Así, por ejemplo, en la figura 6 (B) se observa que para el análisis “All amplifications” de los datos de GBM, pocos genes tuvieron que ser filtrados debido a su baja frecuencia de alteración en las muestras. En la tabla 4, se puede consultar más información sobre el número de genes filtrados en los diferentes análisis de sobrevida por cohortes de cáncer.

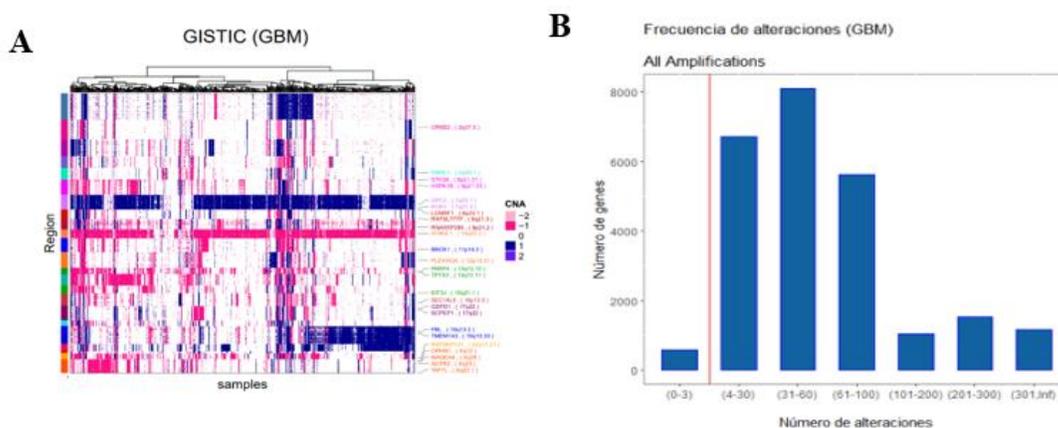


Figura 6 Obtención de datos y filtrado. (A) Perfil de CNAs en GBM. (B) Frecuencia de alteración en genes de pacientes con GBM.



<b>PAAD</b>	106	610	20925	24475
<b>PCPG</b>	5933	4365	23688	24471
<b>PRAD</b>	616	395	19591	14288
<b>READ</b>	2067	2169	23426	24352
<b>SARC</b>	13	0	10992	21642
<b>SKCM</b>	20	20	16553	22476
<b>STAD</b>	0	312	11841	19387
<b>STES</b>	0	41	7281	16869
<b>TGCT</b>	2430	3405	24235	24228
<b>THCA</b>	7630	10912	24217	24286
<b>THYM</b>	12052	11138	24775	24694
<b>UCEC</b>	0	263	8404	22033
<b>UCS</b>	1026	1758	21847	24563
<b>UVM</b>	9451	16063	23559	24500

## 9.2 Análisis de Sobrevida

Se realizó un script general para el análisis de cada tipo de cáncer y se generaron documentos PDF como resultado (3 PDFs por cáncer, 37 conjuntos). Algunos ejemplos de éstos forman parte de las siguientes secciones.

Para evaluar la asociación de sobrevida se construyeron gráficos de tipo Manhattan plots para mostrar los valores  $-\log_{10}(p)$  de los genes por región cromosómica. La figura 8 (A) corresponde al análisis “Deep Amplifications” de pacientes con glioblastoma multiforme, en ésta se marcan algunos de los genes con CNAs más asociados a sobrevida entre ellos; SEC61G, VSTM2A, LANCL2, EGFR, VOPP1, todos en cromosoma 7. La figura 8 (B) muestra los valores  $-\log_{10}(p)$  de los genes, estimados en análisis “All Amplifications” en pacientes con cáncer de mama (BRCA). Algunas de las regiones más significativas se localizan en cromosoma 4, 8 y 22. En cromosoma 4, los genes más significativos incluyen a ARHGEF38, EDNRA y TMEM184C, en el cromosoma 8 se encontró a TRIQK y en cromosoma 21 a los genes AGO2 y BID. La Figura 8(C) es del análisis “All Deletions” en Mesotelioma (MESO), en este análisis las regiones más significativas se encontraron en los cromosomas 8 y 9, en la delección del cromosoma 8 se incluye a los genes CDKN2A, MTAP, CDKN2B y en la del cromosoma 9 se incluye la delección de ARK1C2.

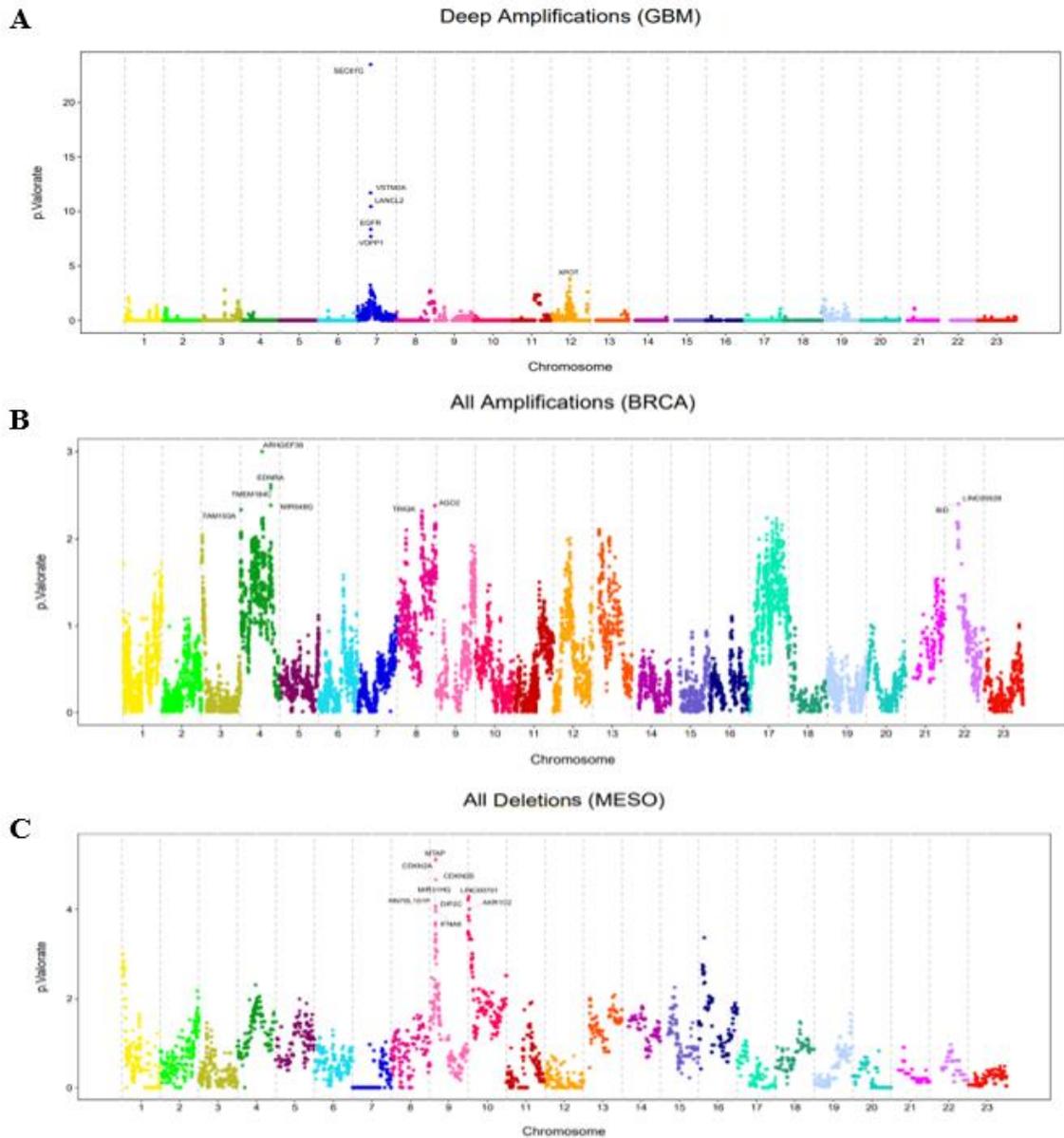


Figura 8      Manhattan plots de los valores  $-\log_{10}(p)$  asociados con sobrevida en diferentes cánceres. (A) Gráfico de Glioblastoma (GBM) con los datos de “Deep Amplifications”. (B) Gráfico de cáncer de mama (BRCA) con los datos de “All Amplifications”. (C) Gráfico de Mesotelioma (MESO) con los datos de “All Deletions”. En negro se marcan los genes más importantes.

### 9.3 Resumen de regiones de CNAs asociadas a sobrevida

En la figura 9 se muestran las regiones de amplificación asociadas a sobrevida en diversos cánceres, considerando los primeros 12 genes significativos ( $p < 0.05$ ) en los análisis de VALORATE. En BRCA se encontraron varias regiones de amplificación asociadas a sobrevida, entre ellas el cromosoma 4, 22 y 8. En KIIHC en cambio se encontraron varios genes en la misma región de amplificación (5p15.33).

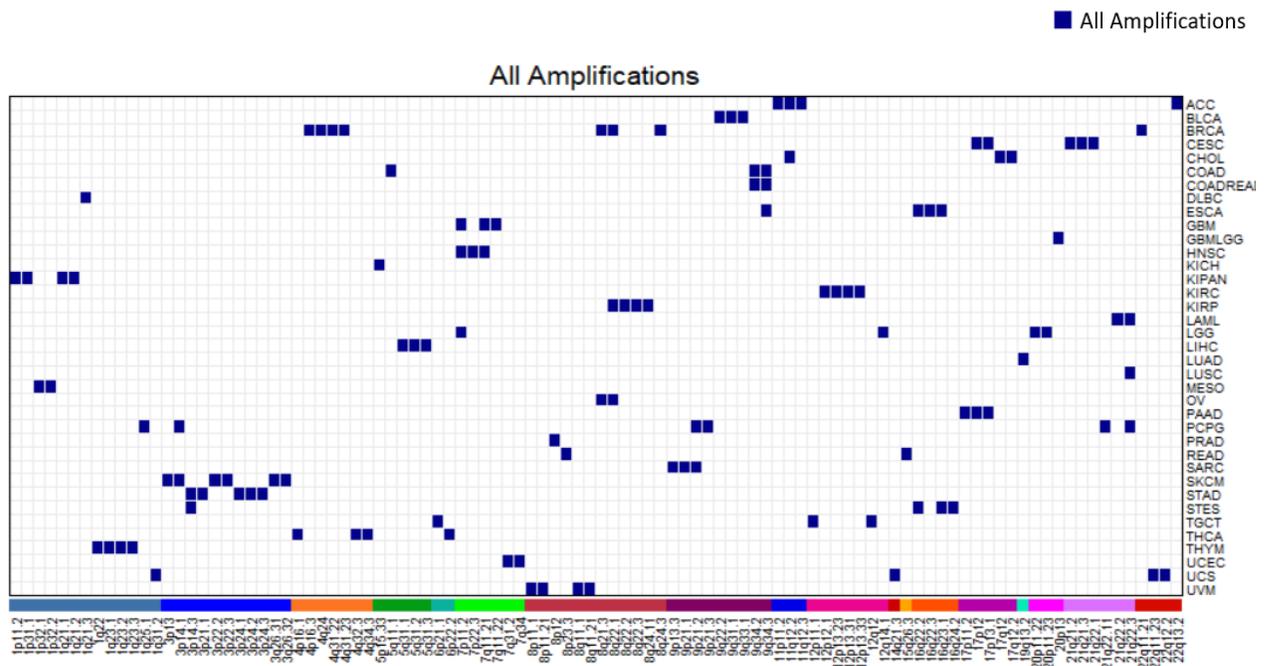


Figura 9 Regiones de CNAs más significativas en el análisis “All Amplifications”

En la figura 10, se muestran algunas de las regiones de delección asociadas a sobrevida en los diferentes cánceres, considerando los 12 primeros genes más significativos por VALORATE. En ACC, por ejemplo, varios genes se encontraron en la región 17p13.1, en cambio UVM se encontraron genes en diferentes regiones de delección.

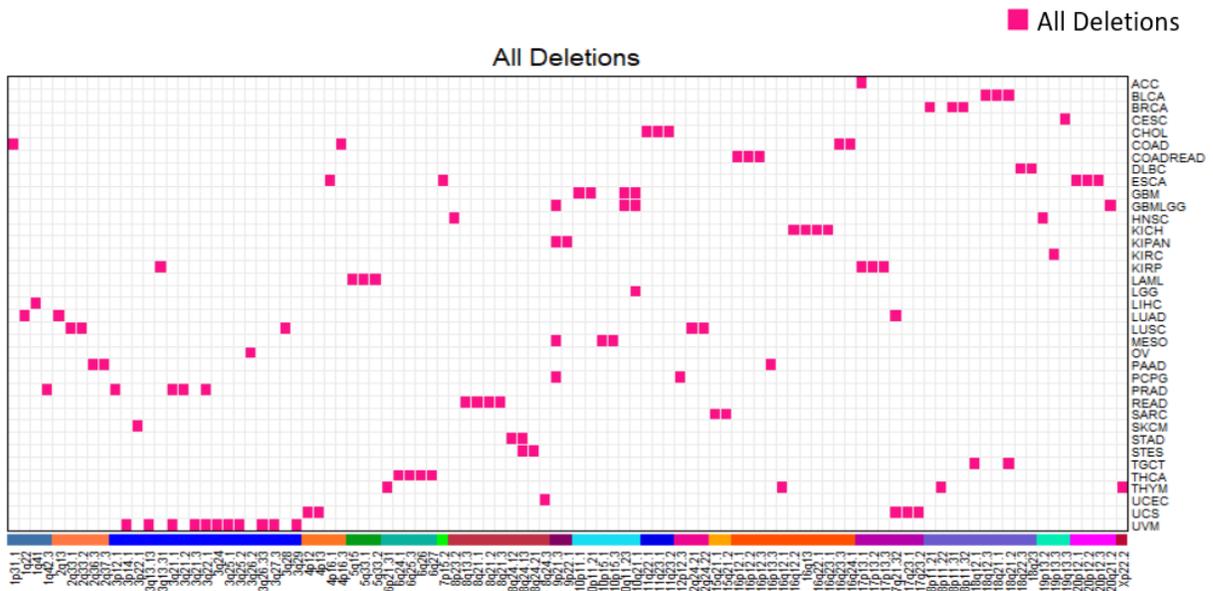


Figura 10 Regiones de CNAs más significativas en el análisis “All Deletions”.

En la figura 11, se muestran algunas de las regiones de amplificación profunda asociadas a sobrevida en los diferentes cánceres. Para algunos cánceres, entre ellos; CHOL, DLBC, KICH y THCA el análisis de VALORATE no se pudo realizar o las regiones identificadas no fueron significativas.

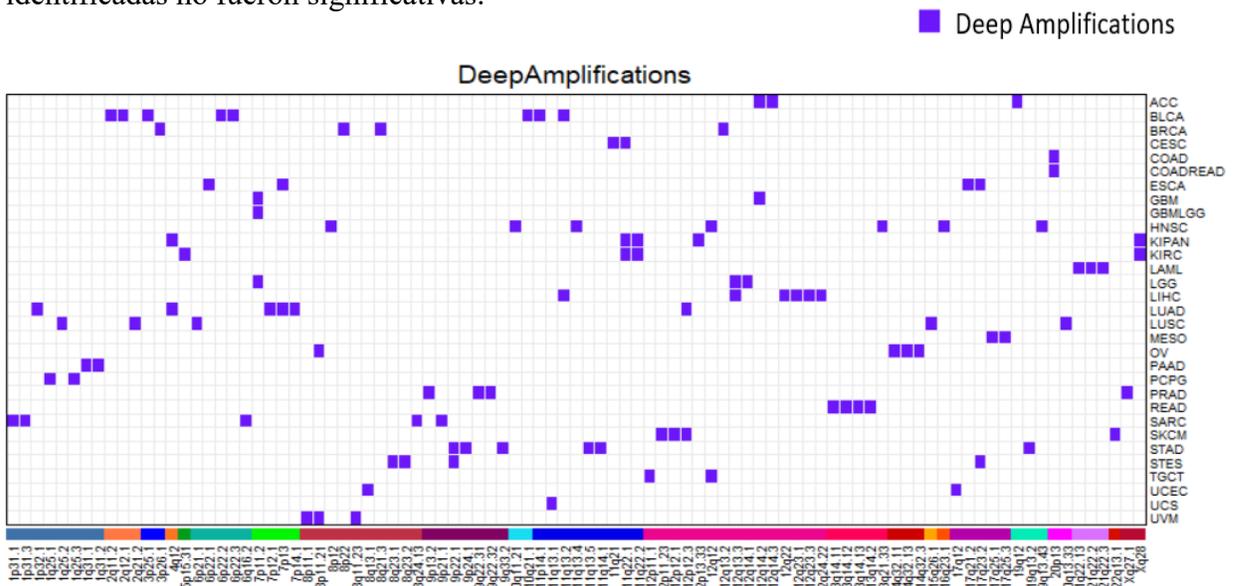


Figura 11 Regiones de CNAs más significativas en el análisis “Deep Amplifications”

En la figura 12, se muestran regiones de delección profunda asociadas a sobrevida en los diferentes cánceres. En este análisis, se encontraron algunas regiones comunes de delección en los diferentes cánceres, principalmente la delección 9p21.3, encontrada en ACC, KIPAN, KIRC, LGG, PAAD y en muchos otros cánceres. En este análisis tampoco se logró identificar regiones asociadas a sobrevida para CHOL, DLBC, KICH y LAML.

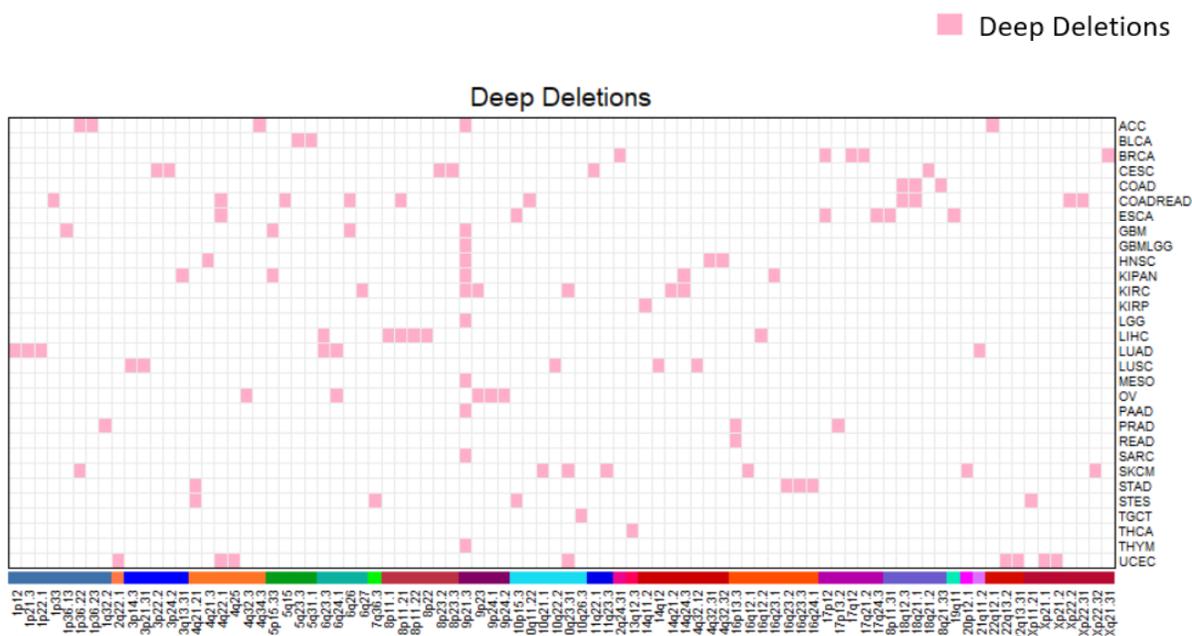


Figura 12 Regiones de CNAs más significativas en el análisis “Deep deletions”.

#### 9.4 Resultados de la validación en literatura

Las siguientes tablas muestran el conocimiento actual que se tiene sobre las regiones de CNAs encontradas significativas en los análisis de VALORATE de acuerdo con los registros encontrados en PubMed. Con letra (C) para indicar genes conocidos en cáncer y (N) para los genes nuevos o poco conocidos. Por ejemplo, para Adrenocortical Carcinoma (ACC) el gen TOB2, también llamado TOB4, fue encontrado como conocido dado que tiene 50 publicaciones reportadas en PubMed. Este gen pertenece a una familia de genes antiproliferativos, (Bai et al., 2017) importantes para el desarrollo de tejidos y como reguladores negativos del ciclo celular. Por el contrario, el gen SNORD67, que codifica

para un RNA pequeño nuclear solo tenía reportado 1 publicación en PubMed relacionada con cáncer (en abstracts).

Tabla 5 Resultados búsqueda en PubMed para el análisis “All Amplifications”  
(primeros 12 genes por cáncer)

“All Amplifications”											
ACC		BLCA		BRCA		CESC		CHOL		COAD	
MIR5582	N	TRIM32	C	ARHGEF38	N	RWDD2B	N	FEN1	C	FAM157B	N
F2	C	PAPPA	C	EDNRA	C	N6AMT1	C	FADS2	N	TUBBP5	N
TOB2	C	DEC1	N	TMEM184C	N	USP16	C	FADS1	C	LINC00094	N
LRP4	C	TNC	N	PRMT10	N	CCT8	C	FADS3	N	WDR5	C
PHF5A	C	TNFSF8	C	BID	C	LTN1	C	MYRF	C	RXRA	C
SNORD67	N	LINC00474	N	LINC00528	N	MAP3K7CL	N	TMEM258	N	RNU6ATAC	N
SNORD67	N	LINC00587	N	MIR548G	N	RNA5SP489	N	MIR611	N	COL5A1	C
CKAP5	C	CYLC2	N	AGO2	C	7SK	C	MIR1908	N	MIR4669	N
UBXN1	C	SHC3	C	FAM193A	N	COX10	C	C17orf78	N	EMB	C
POLR2G	N	SECISBP2	C	TRIQK	N	TRIM16	C	ACACA	C	PARP8	N
RN7SL259P	N	CKS2	C	RN7SKP231	N	TIAM1	C	TADA2A	N	FAM69B	N
PGA3	C	MIR3153	N	SNORA31	N	RCVRN	N	DAGLA	C	SDCCAG3	C
COADREAD		DLBC		ESCA		GBM		GBMLGG		HNSC	
RNA6ATAC	N	SNORA31	C	HCCAT5	N	VSTM2A	N	SMOX	C	MPRS17	N
WDR5	C	SNAPIN	N	ZFH3	C	SEC61G	C	MAVS	C	ZNF713	N
AGPAT2	C	ILF2	C	EGFL7	C	EGFR	C	RNF24	C	SEPT14	C
SNHG7	C	LOR	C	AGPAT2	C	SnoU13	N	UBOX5	C	GBAS	C
EGFL7	C	SPRR2F	C	MIR126	C	SNORA73	C	PANK2	C	RN7SL64P	N
LCN10	N	LCE1E	N	GABARAPL2	C	RNA5SP231	N	AP5S1	C	SNORA73	N
LCN10	N	SPRR2G	N	ADAT1	N	RN7SKP75	N	MIR103A2	C	SEC61G	C
LCN6	C	S100A1	C	CHST6	C	ZNF736	N	HSPA12B	C	ZNF733P	N
MIR126	C	LCE1D	N	TMEM231	N	AUTS2	C	RN7SL839	C	LANCL2	C
OLFM1	C	S100A2	C	CHST5	N	RN7SL371P	N	C20orf27	C	VOPPI	C
MRPS2	N	SPRR3	C	TMEM170A	N	WBSCR17	N	GNRH2	C	RN7SL855P	N
C9orf116	N	S100A13	C	SNORA76	N	ZNF680	N	PTPRA	C	GPR146	N
KICH		KIPAN		KIRC		KIRP		LAML		LGG	
AHRR	C	NEGR1	C	CLEC2D	C	UTP53	C	PRDM15	C	METTL1	C
C5orf55	N	HFE2	C	KLRB1	C	HRSP12	C	ERG	C	AGAP2	C
TRIP13	C	HIST2H3C	N	A2MP1	N	NACAP1	C	WRB	C	TSPAN31	C
CEP72	C	NBPF16	N	SNORA75	N	C8orf47	C	SNORA62	C	CYP27B1	C
PDCD6	C	HIST2H2AC	N	SNORA75	N	C8orf56	C	SnoU13	C	CDK4	C
LRRC14B	N	FCGR1C	C	SSPN	N	MTERFD1	C	HMGNI	C	PAX1	C
SLC6A18	N	NOTCH2NL	C	ITPR2	C	U3	C	SH3BG3	C	MARCH9	C
SLC9A3	C	HIST2H2BA	N	RNA5SP354	N	SNORA72	N	LCA5L	N	METTL21B	C
BRD9	C	GPR89C	N	RNU4ATAC16P	N	SNORD77	N	PSMG1	C	XRN2	C
EXOC3	N	HIST2H2AB	C	NINJ2	C	MIR1273A	C	ETS2	C	RN7SKP140	N
ZDHH11	N	HIST2H2BE	C	KDM5A	C	snoU13	N	BRWD1	N	ZNF716	N
CCDC127	N	HIST2H3D	N	FAM138D	N	SnoU13	C	LINC00114	C	MIR3147	N

<b>LIHC</b>		<b>LUAD</b>		<b>LUSC</b>		<b>MESO</b>		<b>OV</b>		<b>PAAD</b>	
DND1	C	MIA	C	COL6A1	C	C1orf87	N	NECAB1	N	ATPAF2	N
HARS	C	C19orf54	C	FTCD	C	MYSM1	C	LRRC69	N	GID4	N
SNORA74	N	SNRPA	C	COL6A2	C	TACSTD2	C	MIR4661	N	SNORA74	N
TMCO6	N	NUMBL	C	ITGB2	C	HOOK1	C	RN7SL77P	N	SNORA74	N
NDUFA2	N	EGLN2	C	FAM207A	C	NFIA	C	SLC26A7	N	SnoU13	N
WDR55	N	ITPKC	C	PTTG1IP	C	FGGY	N	RNA5SP273	N	WDR16	N
IK	C	RAB4B	C	CSTB	C	JUN	C	TMEM64	C	GAS7	C
MIR3655	N	ADCK4	N	RRP1	C	OMA1	C	SnoU13	N	MAGOH2	N
TMEM173	C	CYP2B7P1	C	PDXK	C	CYP2J2	C	TMEM55A	C	LINC00675	C
SIL1	C	CYP2B6	C	TRAPPC10	C	RN7SL713P	N	RN7SKP231	C	XAF1	C
FSTL4	N	CYP2G1P	N	HSF2BP	C	MIR4711	N	TRIQK	N	C17orf100	N
RPS14	C	CYP2A7	C	RRP1B	C	RN7SL475P	C	OTUD6B	C	SLC13A5	C
<b>PCPG</b>		<b>PRAD</b>		<b>READ</b>		<b>SARC</b>		<b>SKCM</b>		<b>STAD</b>	
TUSC1	C	WRN	C	ERICH1	N	APTX	C	SnoU13	C	ERC2	C
RN7SKP120	N	PURG	C	ZNF596	N	C9orf72	C	MIR4789	C	CACNA2D3	C
TNFSF18	C	DCTN6	C	FBXO25	C	ACO1	C	RNA5SP147	N	ZNF385D	N
TNFSF4	C	TMEM66	N	RPL23AP53	N	RNA5SP281	C	TBL1XR1	N	ESRG	N
PRDX6	C	TEX15	N	OR4F21	N	GVQW1	N	RN7SL418P	N	MIR1305	N
ATP50	C	PPP2CB	C	TDRP	N	NDUFB6	C	PRICKLE2	C	MIR4792	N
CBS	N	TUBBP1	N	TARSL2	N	B4GALT1	C	NAALADL2	C	SLC4A7	C
PKNOX1	N	MBOAT4	N	VIMP	C	SMU1	N	FAM19A1	N	EOMES	C
U2AF1	C	GTF2E2	C	SnoU13	N	PTENP1	C	NLGN1	C	THRB	C
PRICKLE2	N	SMIM18	C	SNRPA1	C	CHMP5	N	KCNMB2	N	RN7SL859P	N
ITSN1	N	UBXN18	C	TM2D3	C	NFX1	C	STAC	C	NEK10	C
LINC00649	C	GSR	C	FAM138E	N	BAG1	C	DCLK3	N	HMGB1P5	N
<b>STES</b>		<b>TGCT</b>		<b>UVM</b>		<b>THYM</b>		<b>UCEC</b>		<b>UCS</b>	
ERC2	C	CPNE8	C	POTEA	N	FCRL2	N	ZC3HAV1	C	GLRX2	C
ADAT1	N	ALG10B	C	FNTA	C	DUSP23	N	CFTR	C	HORMAD2	C
TERF2IP	C	RNA5SP358	C	HGSNAT	C	FCRLB	N	CTTNBP2	C	SUSD2	C
KARS	C	RNA5SP359	C	SGK196	N	NR1I3	C	ASZ1	N	C14orf182	N
KLHD4	N	ALG10	C	U3	C	SPAT1	C	TTC26	N	RN7SL3	N
JPH3	C	SYT10	C	RN7SKP41	N	NTRK1	C	PARP12	C	GGT5	C
SLC7A5	C	SLC35B2	C	RN7SKP32	N	TMEM79	C	PRSS58	N	KLHDC2	N
TMEM170A	N	TMEM63B	C	UBE2V2	C	GBA	C	PRSS37	C	ARF6	C
CHST6	C	CDC5L	C	MCM4	C	SLAMF1	C	OR9A4	N	RN7SL2	N
CALB2	C	SLC29A1	C	SNAI2	C	KIAA0907	C	PRSS1	C	RNA7SKP193	N
CHST5	C	SPATS1	C	RN7SKP294	N	FDPS	C	EPHB6	C	C14orf183	N
TMEM231	N	C6orf223	C	SPIDR	N	PKLR	C	PRSS3P2	N	POM121L9P	N
<b>THCA</b>		<b>LRRC16A</b>		<b>VEGFC</b>		<b>CLNK</b>		<b>SPOCK3</b>			
	N										
	C										
	N										
	C										

En la Tabla 6 se muestran los resultados de la búsqueda en PubMed para el análisis “All Deletions”. En este se encontró, por ejemplo, para Cáncer cervical (CESC) el gen

DIRAS1 como conocido dado que tiene 18 publicaciones reportadas en PubMed. Este gen codifica para una GTPasa relacionada con la vía RAS (Sutton et al., 2018), Aunque no se ha descrito tanto para este cáncer, su subexpresión se le ha relacionado con peor pronóstico en cáncer de ovario, adenocarcinoma esofágico y cáncer colorrectal (Sutton et al., 2018; Zheng et al., 2017; Zhu et al., 2013). Por el contrario, el gen ZNF555 no tenía publicaciones en PubMed relacionadas con cáncer, este gen codifica para una proteína de tipo dedos de zinc cuya función todavía no es clara.

Tabla 6 Resultados búsqueda en PubMed para el análisis “All Deletions”

“All Deletions”											
ACC		BLCA		BRCA		CESC		CHOL		COAD	
GAS7	C	CTIF	N	POTEC	C	DIRAS1	C	CNTN5	N	RN7SL538P	N
ELP5	N	LIPG	C	ANKRD20A5P	N	ZNF556	N	C11orf57	N	LRRC7	N
GABARAP	C	SETBP1	C	CYP4F35P	N	SLC39A3	N	C11orf93	C	TTL7	C
PHF23	C	MIR4319	N	CXADRP3	C	SGTA	C	C11orf34	N	RN7SKP247	N
CLDN7	C	MRO	C	ZNF519	N	ZNF554	N	C11orf92	C	SNORA48	N
DVL2	C	SMAD4	C	SEH1L	N	ZNF555	N	snoR66	N	CDH13	C
YBX2	C	ME2	C	DLGAP1	C	THOP1	C	snoU13	N	TLDC1	N
SLC2A4	C	ELAC1	N	CEP192	C	GADD45B	C	FDX1	C	KIAA0513	C
ACADVL	C	RN7SL695P	N	ADCYAP1	C	GNG7	C	SDHD	C	KLHL36	N
ASGR2	N	SIGLEC15	C	TYMS	C	TMPRSS9	N	FDXACB1	N	ATP2C2	C
DLG4	C	RNA5SP455	N	CETN1	C	TIMM13	N	DLAT	C	COX4I1	C
ASGR1	C	RPL17	C	YES1	C	LMNB2	C	TTC12	C	COTL1	C
COADREAD		DLBC		ESCA		GBM		GBMLGG		HNSC	
CACNG3	N	ZADH2	C	HNRNPA2B1	C	SNORA40	N	ELAVL2	C	CSMD1	C
TMEM159	N	CTDP1	C	NFE2L3	C	ZNF33A	N	MIR4473	N	ACTL9	N
ZP2	C	PARD6G	C	ESF1	N	ZNF37A	C	MIR4474	N	MYO1F	C
SNORA1	N	FAM69C	N	SEL1L2	N	MIR605	N	MIR548F1	N	ADAMTS10	C
ACSM3	C	CNDP1	C	NDUFAF5	N	PRKG1	C	CSTF2T	N	MARCH2	C
DCUN1D3	N	NFATC1	C	TBC1D14	N	ANKRD30A	C	MBL2	C	OR2Z1	N
LYRM1	C	MBP	C	RN7SL86P	N	RN7SL314P	N	C10orf40	N	HNRNPM	C
ACSM1	C	TSHZ1	C	RNU105B	N	NAMPTL	N	ZWINT	C	PRAM1	N
GP2	C	SALL3	C	FLRT3	C	DKK1	C	PGBD3	N	ZNF414	N
ACSM2B	N	KCNG2	N	SPTLC3	C	CSTF2T	N	MIR1296	N	RETN	C
ACSM5	N	CNDP2	C	ISM1	C	FAM21A	N	MTRNR2L	N	STXBP2	C
AMOD	C	PQLC1	N	JAG1	C	LINC00843	N	BICC1	C	TRAPC5	N
KICH		KIPAN		KIRC		KIRP		LAML		LGG	
FAM192A	N	C9orf53	N	CCDC94	N	MIR4447	N	FAM81B	N	MIR548F1	N
DUS2L	N	CDKN2A	C	C19orf10	N	LSAMP	C	POU5F2	N	BICC1	C
CYB5B	C	CDKN2B	C	HDGFRP2	C	INCA1	C	MIR227	N	IPMK	C
EXOCSC6	N	MTAP	C	RN7SL84P	N	SLC25A11	C	TNIP1	C	ZWINT	C
TERF2IP	C	FAM154A	N	RN7SL528P	N	PHF23	C	G3BP1	C	TFAM	C
BBS2	C	DMRTA1	C	RN7SL626P	N	TNFSF12	C	FAT2	C	UBE2D1	C
RANBP10	C	HAUS6	C	RN7SL121P	N	NDEL1	C	NDST1	C	CISD1	C

FUK	C	SCARNA8	N	TNFAIP8L1	N	SOX15	C	RPS14	C	MTRNR2L5	N
SIAH1	C	RRAGA	C	UHRF1	C	KDM6B	C	FAM114A2	N	MIR3924	N
DDX28	N	SNORA30	N	TICAM1	C	UBE2G1	C	SLC36A1	C	RN7SKP196	N
WWP2	C	PTPLAD2	C	STAP2	C	PAFAH1B1	C	GM2A	C	LINC00844	C
WDR59	N	SNORD39	N	UBXN6	N	TAX1BP3	C	MFAP3	N	RNA5SP318	N

<b>LIHC</b>		<b>LUAD</b>		<b>LUSC</b>		<b>MESO</b>		<b>OV</b>		<b>PAAD</b>	
PTPN14	C	SNORA26	N	SNORD56	N	C9orf53	N	SEC62	C	DNAJA3	C
ESRRG	C	SLC25A44	N	ALS2CR11	C	MTAP	C	LRRIQ4	N	CORO7	N
CENPF	C	PMF1	C	ABI2	C	CDKN2A	C	LRRC31	N	SPHKAP	N
KCNK2	C	SEMA4A	C	ICOS	C	CDKN2B	C	SMAD7	N	THAP4	N
SNORD116	C	LMNA	C	CYP20A1	N	LINC00701	C	SKIL	C	BOK	C
RNA5SP76	N	EFCAB13	N	CTLA4	C	DIP2C	C	SLC7A14	C	PDCD1	C
KCTD3	N	SEPT10	C	CD28	C	MIR31HG	C	PHC3	C	ATG4B	C
USH2A	C	LIMS3L	N	RAPH1	N	RN7SL151P	N	GPR160	C	D2HGDH	C
RN7SN464P	N	LIMS3	N	SNORA1	C	AKR1C2	C	MYNN	C	GAL3ST2	C
RRP15	N	LIMS3	N	snoU13	N	IFNA6	N	TERC	C	ING5	C
TGFB2	C	LINC00116	C	HRK	C	IFNA8	C	LRRC34	N	NEU4	C
C1orf143	N	ITGB3	C	WDR12	C	IFNA2	C	ACTRT3	N	DTYMK	C

<b>PCPG</b>		<b>PRAD</b>		<b>READ</b>		<b>SARC</b>		<b>SKCM</b>		<b>STAD</b>	
RERG	C	MIR5688	N	WWP1	C	ATP8B4	N	SEC22C	N	SMAD12	C
RERGL	C	LINC00971	N	TRPA1	C	RN7SL307P	N	SS18L2	N	DERL1	C
PIK3C2G	C	MIR54811	N	CPNE3	C	USP8	C	NKTR	C	RNY4P5	N
GRIN2B	C	FAM86JP	N	RALYL	C	RNA5SP395	N	VIPR1	C	HAS2	C
MGST1	C	C3orf36	N	STAU2	C	SHC4	C	FAM198A	N	MTBP	N
SNORD39	N	GGPS1	C	CNGB3	C	GALK2	C	CCDC13	N	MRPL13	N
MIR491	C	MIRA4753	N	CNBD1	N	COPS2	C	HHATL	C	SNTB1	N
SNORA30	N	LYST	C	SLC7A13	N	SECISBP2L	N	CYP8B1	C	TBC1D31	N
C9orf53	N	GNG4	C	U8	N	RN7SL577P	N	HGD1A	C	DSCC1	C
KLHL9	C	SNORA14B	N	snoU13	C	MIR4716	N	ZBTB47	N	RN7SL396P	N
IFNW1	N	TOMM20	C	RMDN1	N	RNA5SP394	N	RN7SL567P	N	RNA5SP277	N
IFNA10	N	RBM34	N	MMP16	C	GABPB1	C	KLHL40	N	snoU13	N

<b>STES</b>		<b>TGCT</b>		<b>THCA</b>		<b>THYM</b>		<b>UCEC</b>		<b>UCS</b>	
NDUFB9	C	NOL4	C	T	N	FRMPD4	N	MROH1	N	RN7SKP199	N
FAM84B	C	MIR4529	N	SDIM1	N	IP6K3	C	BOP1	C	GABRG1	C
KIAA0196	C	RAB27B	C	LINC00473	C	MLN	C	ARHGAP39	N	CA4	C
ZNF572	N	POLI	C	SNORD45	N	ITPR3	C	SCXB	N	SKAP1	C
SQLE	C	MBD2	C	GPR126	N	LEMD2	N	HSF1	C	HEATR6	C
MIR4662B	N	DCC	C	MAP3K4	C	LINC00336	N	KIFC2	N	MIR4737	N
TMEM65	N	CCDC68	C	LPA	C	MIR3934	N	ZNF7	C	RPS6KB1	C
RN7SL329P	N	C18orf54	N	PLG	C	MNF1	C	CPSF1	N	RN7SKP82	N
RNF139	C	SNORA37	N	AGPAT4	C	SBP1	C	ZNF517	C	RNFT1	N
TATDN1	C	STARD6	N	SNORD28	N	GGNBP1	N	TONSL	C	SCARNA20	N
TRMT12	C	DYNAP	N	snoU13	C	PIEZO2	C	SCXA	N	BRIP1	C
CCDC26	C	RNA5SP459	N	SNORA20	C	PAPD5	C	ANK2	N	INTS2	N

<b>UVM</b>	
C3orf79	N
LINC00884	N
APOD	C
CPNE4	N
CASR	C

SnoU13	N
PRICKLE2	C
DNAJB8	C
U3	C
SERP1	C
DCUN1D1	C
HHLA2	N

En la Tabla 7 se muestran resultados de la búsqueda en PubMed para el análisis “Deep Amplifications”. En este se encontró, por ejemplo, para Cáncer de Vejiga (BLCA) el gen IL1R1 (Lappano et al., 2020) que codifica para un receptor de interleucina 1, tipo 1, fue encontrado como conocido dado que tiene 67 publicaciones reportadas en PubMed. Este se ha relacionado con desarrollo de cáncer de mama a través de la exacerbación de respuestas inflamatorias. (Lappano et al., 2020) Por el contrario, el gen LRRC16A solo tenía reportadas 1 publicación en PubMed asociada con el término “cáncer” (en abstracts). Este gen codifica para la proteína 16A con repeticiones ricas en leucina, esta proteína es importante en la regulación de la polimerización de actina y recientemente se ha descrito sobre su participación en desordenes metabólicos. (Vadakedath & Kandi, 2018)

Tabla 7 Resultados búsqueda en PubMed para el análisis “Deep Amplifications”

“Deep Amplifications”											
ACC	BLCA	BRCA	CEC	STES	COAD						
GNS	C	MIR605	N	RNA5SP255	N	CEP57	C	TLK2	C	LZTS3	C
WIF1	C	IL1RL2	N	C8orf48	N	RNA5SP345	N	KCNV1	N	ITPA	C
SnoU13	N	IL1R1	C	DLC1	C	FAM76B	N	NUDCD1	C	MRPS26	N
TBC1D30	N	LSM3	C	OR6C3	N	MTMR2	C	ENY2	C	C20orf141	N
RNU6ATAC42P	N	PELI3	C	OR6C75	N	MAL2	C	MRC2	C	FASTKD5	N
C19orf12	C	TMEM43	C	OR6C6	N	MIR1260B	N	PKHD1L1	C	AVP	C
CCNE1	C	MRPL11	C	OR6C1	N	CCDC82	N	EIF3E	C	C20orf194	N
URI1	C	RN7SKP158	N	ITPR1	C	RNA5SP346	N	RSPO2	C	CPXM1	C
PLEKHF1	N	XPC	C	OR6C74	N	JRKL	N	DENND4C	N	SLC4A11	C
C12orf66	N	CHCHD4	C	OR10A7	N	RNA5SP347	N	SYBU	N	PCED1A	N
TMEM5	N	RNY5P5	N	RN7SKP231	C	KDM4E	C	RNA5SP275	N	RN7SL555P	N
XPOT	C	LRRC16A	N	SnoU13	N	SRSF8	N	SNORD112	N	GNRH2	C
COADREAD	UCEC	ESCA	GBM	GBMLGG	HNSC						
DDRGK1	C	TRIM55	N	MRC2	C	SEC61G	C	EGFR	C	MIR3664	N
CPXM1	C	ARMC1	N	ZNF385C	N	VSTM2A	C	SNORA73	N	WVOX	C
MRPS26	N	PDE7A	C	NKIRAS2	C	LANCL2	C	SEC61G	C	NADSYN1	C
ITPA	C	DNAJC5B	N	MARCH10	N	EGFR	C	LANCL2	C	DHCR7	C

FASTKD5	N	MTFR1	C	TRIM15	C	VOPP1	C	VOPP1	C	CXCL12	C
C20orf141	N	VCPIP1	C	HCG18	C	snoU13	N	VSTM2A	N	TMEM117	N
RN7SL555P	N	C8orf44	C	RPP21	C	SNORA73	N	snoU13	N	ADAMTS20	C
C20orf194	N	MRM1	N	HCG17	N	snoU13	N	FKBPL9	N	ZNF835	N
GNRH2	C	CCL4L1	C	TRIM26	C	SNORD83	N	SEPT14	C	UNC93B6	N
LZTS3	C	CCL4L2	N	TRIM39	C	XPOT	C	RN7SL64P	N	RN7SL761P	N
SLC4A11	C	DHRS11	N	TLK2	C	GNS	C	ZNF713	N	MBOAT4	N
UBOX5	N	TBC1D3C	C	NUDCD3	N	TBK1	C	GBAS		TUBBP1	N
<hr/>											
<b>UVM</b>		<b>KIPAN</b>		<b>KIRC</b>		<b>UCS</b>		<b>LAML</b>		<b>LGG</b>	
FNTA	C	TSPAN9	C	KIAA1377	N	ESRRA	C	ABCG1	C	SNORA73	N
U3	C	PDGFRA	C	MMP7	C	TEX40	N	SNORA51	N	CYPP7B1	C
HGSNAT	C	FIP1L1	C	C11orf70	N	GPR137	C	MIR4760	N	MARCH9	N
POTEA	N	MMP7	C	MMP27	C	FKBP2	N	C21orf88	N	CDK4	C
SGK195	N	snoU13	N	TMEM123	N	PLCB3	C	KCNJ15	C	METTL1	C
RN7SKP41	N	snoU13	N	snoU13	N	BAD	C	IGSF5	N	AGAP2	C
PCMTD1	C	TMEM123	N	snoU13	N	PPP1R14B	C	DSCAM	C	TSPAN31	C
MCM4	C	MMP27	N	MTMRR	C	KCNK4	N	DSCR10	N	EGFR	C
RN7SKP32	N	KIAA1377	N	LCA10	N	STIP1	C	PCP4	C	TSMF	N
RNU105C	N	SNORD36	C	C5orf49	N	DNAJC4	N	B3GALT5	N	OS9	C
RN7SK294	N	TMEM	N	TKTL1	C	TRPT1	N	PSMG1	C	METTL21B	N
LINC00293	N	TMEM187	N	RN7SL687P	N	VEGFB	N	ERG	C	SEC61G	C
<hr/>											
<b>LIHC</b>		<b>LUAD</b>		<b>LUSC</b>		<b>MESO</b>		<b>OV</b>		<b>PAAD</b>	
NT5DC3	N	RN7SL496P	N	NCKAP5	N	BAIAP2	C	CLMN	C	RN7SKP156	N
KSR2	C	CDK13	C	RN7SKP154	N	FDXR	C	MIR3173	N	RNA5SP73	N
SYT12	N	FGGY	N	SLCO3A1	C	TMC6	C	DICER1	C	FAM5C	C
STAB2	C	SPDYE1	N	MTG2	N	METTL23	N	RN7SL506P	N	RGS18	N
RHOD	C	LINC00265	C	ENPPA	C	CYTH1	N	TTC7B	N	MIR1278	N
B4GALNT1	C	KIAA1211	C	HRH3	C	TMEM105	C	SFRP1	C	B3GALT2	N
SLC17A8	N	POLR2J4	C	CLICC5	C	CCDC40	C	GPR66	C	snoU109	N
snoU13	N	ART4	C	RUNX2	C	SMIM5	N	LINC00617	N	U3	C
KIF5A	C	URGCP	C	SS18L1	C	HID1	N	SCARNA13	N	RGS13	C
CRADD	C	H2AFJ	C	ENPP5	N	EVPL	C	GLRX5	C	GLRX2	C
DCTN2	C	UBE2D4	N	ABL2	C	FBF1	N	ATG2B	C	RGS1	C
ARHGEF25	N	RN7SKP218	N	OSBPL2	C	GALK1	C	C14orf132	N	CDC73	C
<hr/>											
<b>PCPG</b>		<b>PRAD</b>		<b>READ</b>		<b>SARC</b>		<b>SKCM</b>		<b>STAD</b>	
TNFSF18	N	DCAF10	N	HTR2A	C	PGM1	C	RN7SKP262	N	DENND4C	N
PRDX6	N	RN7SKP171	N	NUFIP1	C	HAS2	C	RNA5SP354	N	C11orf30	C
TNFSF4	N	SLC25A51	N	KIAA1704	N	ITGB3BP	C	KRAS	C	PTPRD	C
SnoU13	N	TRMT10B	N	COG3	N	LEPR	C	RN7SL459P	N	PRKRIR	N
RGL1	C	FRMPD1	N	LRCH1	N	PRDM13	C	RN7SL68P	N	LYPD4	N
TSEN15	N	SHB	C	ZC3H13	C	AK4	C	NPTXR	N	DMRTC2	N
COLGALT2	N	EXOSC3	N	SIAH3	C	TNNI3K	N	LRMP	C	RPS19	C
-		MCF2	N	TSC22D1	C	LEPROT	C	PLCZ1	C	TRAF1	C
-		PHF2	C	ESD	C	DNAJC6	C	MIR920	N	CNTRL	C
-		FAM120AOS	N	SPERT	N	LRRC7	C	CAPZA3	N	C5	C
-		FAM120A	N	KCTD4	N	RN7SL854P	C	FAM227A	N	WNT11	C
-		HIATL1	C	SERP2	C	MIR873	C	DMC1	C	DLG2	C
<hr/>											
<b>TGCT</b>		<b>CHOL</b>									
ALG10B	N	<b>KIRP</b>	-								
RNA5SP359	N	<b>THCA</b>	-								

RNA5SP358	N	<b>THYM</b>	-
SYT10	N	<b>KICH</b>	-
ALG10	C	<b>DLBC</b>	-
CPNE8	C		
-	-		

En la Tabla 8 se muestran resultados de la búsqueda en PubMed para el análisis “All Deletions”. En este se encontró, por ejemplo, para Cáncer de mama (BRCA) el gen TGIF2LX o factor de crecimiento inducido por el factor homeobox 2, ligado a X, fue encontrado como conocido dado que tiene 50 publicaciones reportadas en PubMed. Este gen es regulador negativo de señalización celular y se ha encontrado que la sobreexpresión de este gen inhibe la proliferación de células de cáncer de colón (Akbari et al., 2017). Por el contrario, el gen TMEM132B, solo tenía reportado 3 publicaciones en PubMed relacionadas con el término “cáncer”. De estas publicaciones solo una realmente describe su participación en cáncer, la modificación del gen por N<sup>6</sup>-Metiladenosina se encontró asociada a pronóstico en pacientes con adenocarcinoma de recto (Huang et al., 2020).

Tabla 8 Resultados búsqueda en PubMed para el análisis “Deep Deletions”

“Deep Deletions”											
ACC		BLCA		BRCA		CESC		COADREAD		COAD	
CASZ1	C	HINT1	C	SHISA6	N	CNTN5	C	PARK2	C	SMAD2	C
TTC28	C	LYRM7	N	TMEM132B	N	KBTBD11	N	SHROOM2	C	C18orf32	N
LINC00290	N	SEPT8	N	ZNF18	N	MYOM2	C	SMAD2	C	MIR4744	C
RERE	C	IL13	C	TGIF2LX	C	ARHGEF10	C	PRKG1	C	SNORD58C	N
SNORA77	N	C5orf56	N	NLE1	N	CLN8	C	GPR143	C	DYM	C
CHEK2	C	IL4	C	KRT39	N	MIR596	N	TBL1X	C	SNORD58A	N
CDKN2A	C	IRF1	C	DNAH9	C	RN7SL695P	N	AGBL4	N	SNORD58B	N
SLC2A7	N	IL5	C	KRT40	N	RARB	C	KIAA0825	N	MIR1539	N
UTS2	C	KIF3A	C	FNDC8	N	ITGA9	C	MIR548AO	N	RPL17	C
CDKN2B	C	MIR3936	N	Vault	C	RN7SL872P	N	SFRP1	C	SYT4	C
ENO1	C	SLC22A5	C	KRT223P	N	RNA5SP251	N	RN7SKP248	N	SERPINB2	C
RN7SL451P	N	RAD50	C	SnoU13	N	CSMD1	C	SLC14A2	C	RAN5SP455	N
KIPAN		KIRC		ESCA		GBM		GBMLGG		HNSC	
CDKN2B	C	C9orf53	N	CCSER1	N	C9orf53	N	CDKN2A	C	CDC42BPB	C
CDKN2A	C	CDKN2B	C	MAP2K4	C	CDKN2A	C	CDKN2B	C	AMN	C
C9orf53	N	CDKN2A	C	RPL2P122	N	RN7SK151P	N	C9orf53	N	TRAF3	C
LSAMP	C	PTEN	C	DIP2C	C	MTAP	C	MTAP	C	NOTCH1	C
MTAP	C	MTAP	C	RN7SL754P	N	CDKN2B	C	RN7SL151P	N	MIR4309	N
MIR4457	N	LINC00871	N	RNA5SP298	N	NKD2	C	DMRTA1	C	RN7SL546P	N

CLPTM1L	C	DMRTA1	C	ZMYND11	C	MIR31HG	C	MIR31HG	C	ANKRD9	N
DLST	C	PTPRD	C	RNA5SP297	N	IFNE	C	SNORD39	N	TECPR2	N
WVOX	C	RPS6KA2	C	LINC00662	C	PLG	C	MIR31	C	RCOR1	C
AREL1	N	RN7SL5P	N	TUBB8	N	IFNA1	C	IFNE	C	RN7SKP96	N
FCF1	N	PROX2	N	MYL12A	C	IGSF21	N	IFNA1	C	C9orf53	N
ABCD4	C	VRTN	N	LINC00511	C	MIR31	C	IFNA8	C	CDKN2A	C
<hr/>											
<b>LIHC</b>		<b>LUAD</b>		<b>LUSC</b>		<b>KIRP</b>		<b>MESO</b>		<b>LGG</b>	
FNTA	C	ABCD3	C	SMAD8	N	ZNF219	C	CDKN2B	C	MTAP	C
U3	C	ARHGAP29	C	STXBP6	C	HNRNPC	C	MTAP	C	C9orf53	N
SGK196	N	F3	C	WNT5A	C	RNASE3	C	C9orf53	N	CDKN2B	C
HGSNAT	C	HMGCS2	C	FBLN5	C	TMEM253	N	CDKN2A	C	CDKN2A	C
POTEA	N	REG4	C	C10orf11	C	SLC39A2	C	RN7SL151P	N	RN7SL151P	N
RN7SKP41	N	ADAM30	N	CCD36	C	ARHGEF40	C	RMRPP5	N	RN7SKP120	N
ALDH8A1	C	MIR3688	N	USP4	N	RNASE13	N	MIR31HG	C	RMRPP5	N
TACC1	C	ECT2L	C	RHOA	C	RNASE7	C	IFNA13	C	IZUMO3	N
PFMT	C	MIR548A2	N	C3orf62	N	RPGRIP1	C	RN7SKP120	N	ELAVL2	C
MTMR7	C	SAMSN1	C	MIR4271	N	NDRG2	C	IFNA6	N	IFNA1	C
FTO	C	IL20RA	C	GPX1	C	TPPP2	C	DMRTA1	C	IFNE	C
ASAH1	C	COX5BP2	N	ADK	C	SUPT16H	C	KLHL9	C	SNORD39	N
<hr/>											
<b>PRAD</b>		<b>STES</b>		<b>TGCT</b>		<b>SARC</b>		<b>OV</b>		<b>STAD</b>	
UNKL	N	RNA5SP298	N	C19orf91	C	MTAP	N	SNORD27	N	KIAA0513	C
UBE21	C	RNA5SP297	N	INPP5A	C	C9orf53	N	ECT2L	N	ZDHHC7	N
GNPTG	C	TUBB8	N	GPR123	N	CDKN2B	N	C9orf123	N	CMIP	C
TSR3	N	ZMYND11	C	TTC40	N	CDKN2A	N	TMEM192	N	ZNF469	C
BAIAP3	C	MIR595	N	SPRN	C	MIR31HG	N	SPOCK3	N	PLCG2	C
SERTAD4	N	RASGEF1B	C	CYP2E1	C	RN7SL151P	N	PTPRD	C	7SK	C
SYT14	C	WDR60	C	VENTX	C	KLHL9	N	RN7SL5P	N	SDR42E1	N
DIEXF	N	LINC00689	C	SYCE1	C	IFNA8	N	GLIS3	C	MIR4720	N
HHAT	C	UBQLN2	C	ECHS1	C	IFNA2	N	REPS1	N	CDH13	C
ZZEF1	N	VIPR2	C	MIR202	C	IFNA6	N	CCD28A	C	RASGEF1B	C
UBE2G1	C	MIR5707	N	FRG2B	C	IFNA13	N	KLHL2	N	HSD17B2	C
snoU13	N	-	-	-	-	IFNA1	N	ABRACL	C	HSDL1	N
<hr/>											
<b>UCEC</b>		<b>SKCM</b>		<b>PAAD</b>		<b>THYM</b>		<b>READ</b>		<b>THCA</b>	
GRID2	C	NLGN4X	C	C9orf53	N	CDKN2B	C	RBFOX1	N	UBL3	C
snoU13	N	BICC1	C	CDKN2A	C	C9orf53	N	-	-	-	-
RNA5SP501	N	MACROD2	C	CDKN2B	C	MTAP	C	-	-	-	-
LRP1B	C	SNORD74	C	MTAP	C	CDKN2A	C	-	-	-	-
ARFGAP3	N	ABCC12	C	-	-	-	-	-	-	-	-
CH25H	C	TMPRSS4	C	-	-	-	-	-	-	-	-
LIPA	C	CTNNB1P1	C	-	-	-	-	-	-	-	-
RIBC2	N	-	-	-	-	-	-	-	-	-	-
SMC1B	C	-	-	-	-	-	-	-	-	-	-
DMD	C	-	-	-	-	-	-	-	-	-	-
ANK2	C	-	-	-	-	-	-	-	-	-	-
<hr/>											
<b>UVM</b>		<b>PCPG</b>		<b>KICH</b>		<b>CHOL</b>		<b>UCS</b>		<b>LAML</b>	
-	-	-	-	-	-	-	-	-	-	-	-
<b>DLBC</b>		<b>UCS</b>		<b>LAML</b>							
-	-	-	-	-	-	-	-	-	-	-	-

### 9.5 Comparación de las pruebas VALORATE y Log-Rank

La figura 13 (A) muestra la correlación de los valores  $p$  estimados para el análisis “All del” en cáncer de próstata, entre VALORATE y la prueba log rank. En rosa se marcan algunos de los genes más significativos del análisis de VALORATE y en azul genes donde las estimaciones de los valores  $p$  por ambas pruebas difiere más, cada gen es representado por un punto y su color corresponde a la frecuencia muestras con el gen alterado. La figura 13 (B) muestra el gráfico de correlación del análisis “Deep amplifications” en mesotelioma.

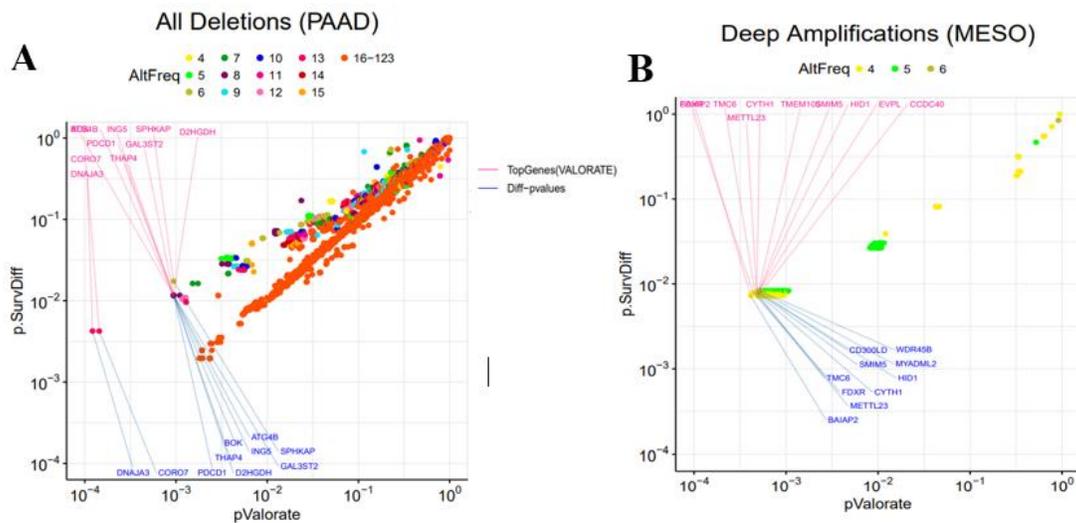


Figura 13 Gráfico de correlación de los valores  $p$  calculados por VALORATE y la prueba log rank. (A) Resultados del análisis “All deletions” en pacientes con adenocarcinoma de páncreas (PAAD). (B) Resultados del análisis “Deep Amplifications” la cohorte de pacientes con mesotelioma (MESO).

### 9.6 Análisis de enriquecimiento

Del análisis de enriquecimiento se obtuvo información sobre los procesos biológicos y/o vías de señalización en los que se encontraban involucrados los genes asociados a sobrevida. En la figura 14 (A) se muestran, por ejemplos los resultados del

análisis “All Deletions” del cáncer PAAD. En este los genes significativos en el análisis de sobrevida estaban asociados con las vías de señalización JNK y MAPK.

En la figura 14 (B) se muestran los resultados del análisis “Deep Deletions” del cáncer BRCA. En este los genes significativos en el análisis de sobrevida estaban asociados con las vías de señalización de leptina, IL-(5,9), PDGFR-b, BDNF y con componentes de estructura anatómica

En la figura 14 (C) se muestran los resultados del análisis “All Amplifications” del cáncer MESO. En este los genes significativos en el análisis de sobrevida estaban asociados con varios procesos de empalme y del complejo U2 del espliceosoma. En rosa se marcan los términos enriquecidos que mantuvieron un valor p ajustado significativo.

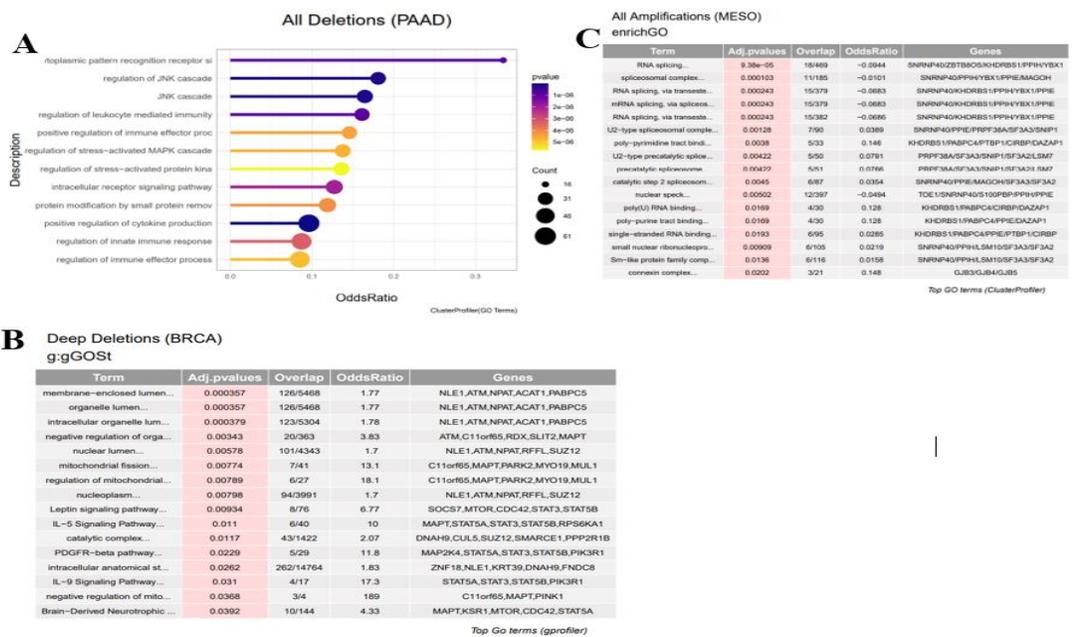


Figura 14 Resultados de los análisis de enriquecimiento. (A) Gráfico lollipop mostrando los resultados del análisis funcional con los datos “All Deletions” en cáncer de páncreas. (B) Resultados del análisis funcional con los datos “Deep Deletions” de pacientes con cáncer de mama. En rosa se indican valores p ajustados <.05. (C) Resultados del análisis funcional realizado con los datos “All Amplifications” en pacientes con BRCA.

### *9.7 Análisis de firmas genómicas*

Las firmas genómicas se utilizaron para evaluar si, en conjunto, los genes asociados a sobrevida podrían servir para estratificar a los pacientes. En la figura 15 (A) se muestra un mapa de calor que describe los grupos de pacientes con gliomas de bajo grado, formados a partir de las firmas de “All Amplifications” correspondientes a este cáncer. Varios genes incluidos en esta firma se encuentran dentro de las regiones de amplificación 7p15.3 y 20q13.33. En este mapa de calor además se observan pacientes asociados a alto y bajo riesgo. En la figura 15 (B) se muestran las curvas de Kaplan-meier de los pacientes con LGG basado en la firma “All amp”. En el gráfico de la figura 15 (B) se observa una clara estratificación de los pacientes en diferentes grupos de riesgo. La figura 15 (C) muestra las curvas de Kaplan-meier del análisis “All del”. Por último, la figura 15 (D) muestra las curvas de Kaplan-meier usando los datos de amplificación y delección.



del”. Por último, la figura 16 (D) muestra las curvas de Kaplan-meier usando los datos de amplificación y deleción.

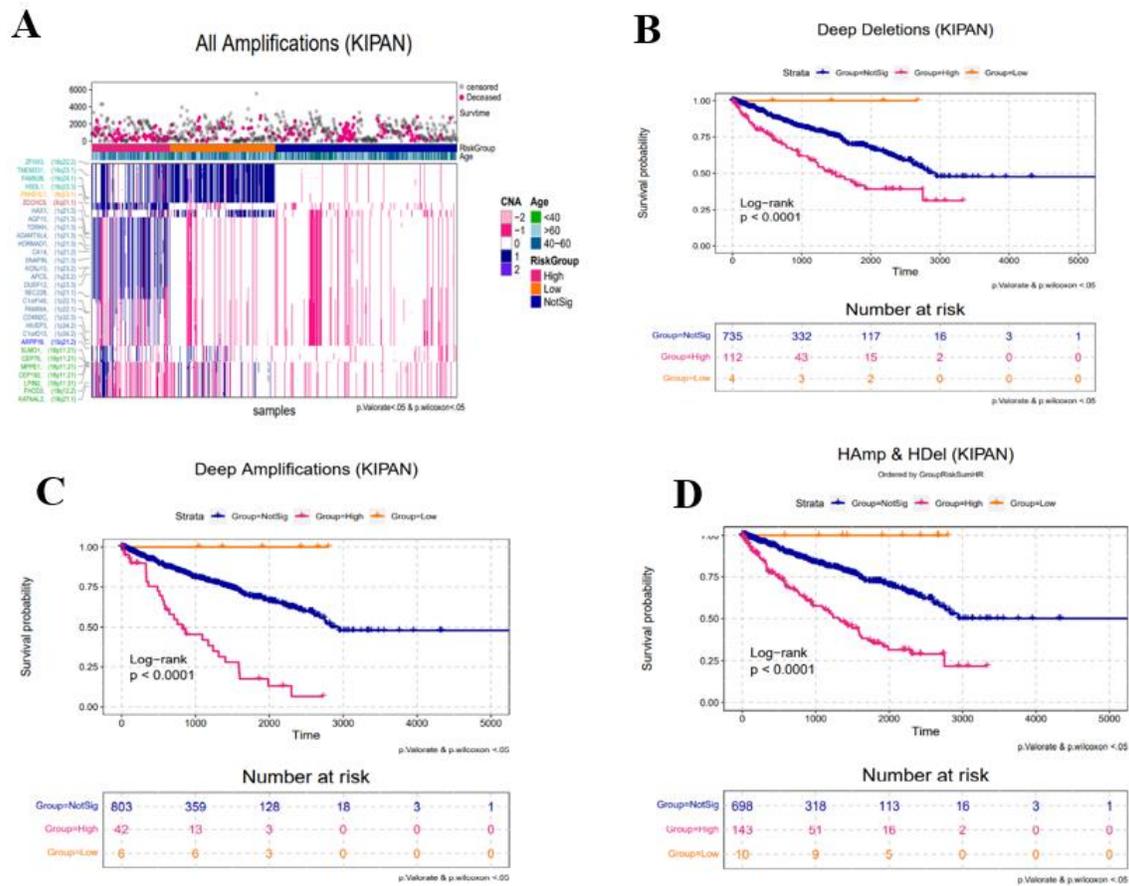


Figura 16 Resultados de firmas genómicas en KIPAN. (A) Mapas de calor con la firma genómica construida para pacientes con cáncer de riñón (KIPAN) con los resultados del análisis “All amplifications”. (B-D) Kaplan-meier de la estratificación de KIPAN con las firmas genómicas construidas de los análisis, (A) “Deep deletions”, (B) “Deep Amplifications” y (D) de la combinación de los análisis Deep amplifications" y “Deep deletions”.

## 9.8 Comparación de pruebas estadísticas

De los análisis de supervivencia realizados con diferentes pruebas estadísticas, se encontró mayor correlación entre VALORATE y el modelo Cox en 30 tipos de cáncer, por ejemplo, en COAD, HNSC y OV como se muestra en Figura 17 (A) mientras que la correlación menos conservada fue entre VALORATE y Log-Rank (Figura 17 (B)). No obstante, para HNSC y UVM se observó una correlación diferente (Figura 18).

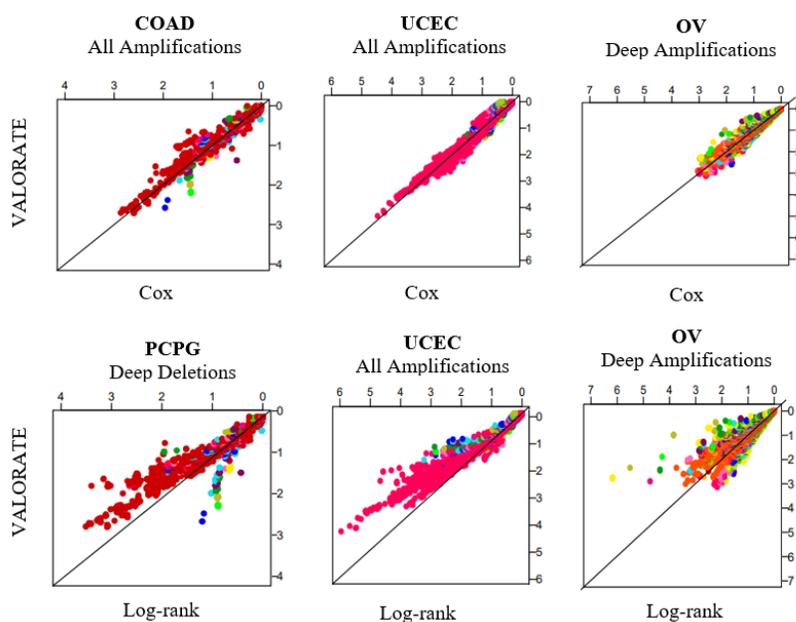


Figura 17 Correlación de análisis de supervivencia entre pruebas estadísticas. (A) Mayor Correlación entre Cox y VALORATE. (B) Menor Correlación entre VALORATE y log Rank.

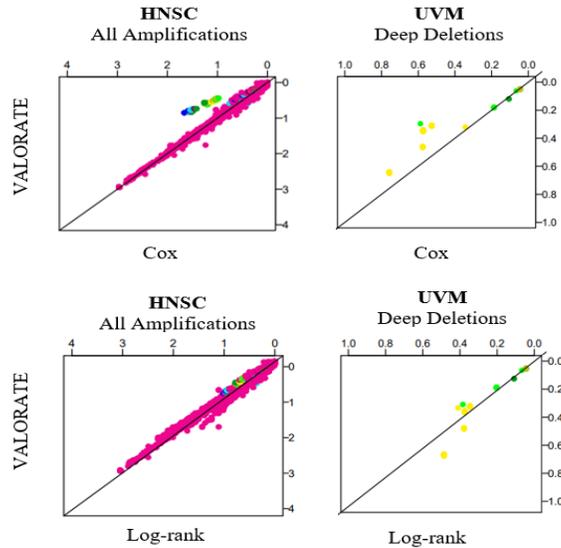


Figura 18 Cancers con correlación diferente entre VALORATE y Cox. (A) Menor Correlación entre Cox y Valorate. (B) Mayor Correlación entre Valorate y log Rank.

### 9.9 Resultados del análisis de fármacos

Del análisis de fármacos se obtuvo información sobre la evidencia clínica en cáncer de pequeñas moléculas inhibidoras y de agentes terapéuticos estudiados contra los biomarcadores seleccionados, que en este caso fueron los genes y su respectivo tipo de variación. En la tabla 15 se muestran resultados con evidencia clínica de agentes terapéuticos con potencial respuesta en pacientes con estos biomarcadores.

Tabla 9 Resultados del análisis de fármacos

Biomarcador evaluado	Cáncer	Fármaco Evaluado	Evidencia
<b>CDK4 (amp)</b>	ACC, GBM, LGG, LIHC	Inhibidor CDK 4/6 LEE011.	Ensayos tempranos
<b>CDKN2B (del)</b>	GBM, KIRC, LGG, MESO, PAAD, SARC	CDK 4/6	Pre-clínicos
<b>CDKN2A (del)</b>	GBM, HNSC, KIRC, LGG, MESO, PAAD, SARC	CDK 4/6 Inhibidor llorasertib	Pre-clínicos Ensayos tempranos
<b>CCNE1 (amp)</b>	ESCA, KICH,	CDK2	Pre-clínico
<b>RB1 (del)</b>	BLCA	Cisplatín	Pre-clínico en BLCA
<b>PBRM1 (del)</b>	BRCA, LUSC	Inhibidores EZH2	Pre-clínico
<b>BAP1 (del)</b>	BRCA, LUSC	Inhibidores PARP	Pre-clínico
<b>PTEN (del)</b>	GBM, KIPAN, KIRC, LGG	Inhibidor PI3K Sirolimus	Preclínicos

<b>SUZ12 (del)</b>	BRCA, AML	Inhibidores BET	Pre-clínico
<b>STK11 (del)</b>	CESC, UCEC	Inhibidores mTOR	Pre-clínico
<b>MYC (amp)</b>	GBM	Inhibidor BET Inhibidor CDK7 Inhibidor FACT	Preclínicos
<b>EGFR (amp)</b>	HNSC, STES	Inhibidor PI3K	Case-report (para HNSC)
<b>CCND1 (amp)</b>	HNSC	Inhibidor EGFR	Pre-clínico
<b>AKT2 (amp)</b>	LUAD	Inhibidor CDK4/6 MK2206	Pre-clínico
<b>CDKN1B (del)</b>	READ	Inhibidor CDK2/4	Pre-clínico
<b>SETD2 (del)</b>	SARC	WEE1	Pre-clínico

## 10. Discusión

Algunas alteraciones en número de copias serían muy útiles si se incorporaran a la práctica clínica ya que pueden ayudar a pronosticar y tomar mejores decisiones sobre tratamientos. No obstante, los métodos que se emplean para la identificación de CNAs pronósticos no son exactos, ya que las asunciones requeridas por su aplicación no siempre se cumplen en cáncer, esto hace que se detecten alteraciones candidatas poco robustas.

Dado que no se han analizado antes CNAs con métodos más exactos, este proyecto de investigación tuvo como objetivo usar VALORATE para identificar genes con alteraciones en número de copia asociados a sobrevida en diversos tipos de cáncer. Asimismo, se creó un proceso de análisis general para identificar CNAs de alto y bajo riesgo, para validar la expresión de los genes involucrados, para realizar un análisis funcional básico de los genes y para encontrar evidencia clínica de respuesta a fármacos.

El proyecto se realizó usando datos públicos de cohortes de pacientes asociados con 37 diferentes cánceres los cuales estaban disponibles en la base de datos TCGA.

### 10.1 Mapas de calor del perfil de CNAs en diferentes cánceres.

Antes de realizar los análisis de sobrevida se crearon mapas de calor para visualizar los perfiles de CNAs de las diferentes cohortes de pacientes con cáncer. Esto sirvió para corroborar los perfiles de CNAs previamente descritos para distintos cánceres.

De los mapas de calor estudiados, los pacientes con cáncer de OV mostraron el perfil de CNA más heterogéneo y un número alto de CNAs (Zack et al., 2013). Otros cánceres con alta carga de CNAs fueron LUSC, LUAD, ESCA, SARC, BLCA, BRCA y CESC.

En PAAD, PRAD y UCEC en cambio se observó claramente un grupo reducido de pacientes desprovistos de CNAs y el resto con un perfil de CNAs heterogéneo o cargado por un tipo de variación. En estudios previos estos grupos se han asociado con diferente evolución clínica, pero en todos ellos la alta carga de CNA se relaciona con cánceres más agresivos (Gotoh et al., 2019; Network, 2015).

En los cánceres GBM y GBMLGG se encontró la amplificación del cromosoma 7 y deleción del cromosoma 10, ambas son alteraciones comunes en estos pacientes. (C. H. Chen et al., 2021; Lopez-Gines et al., 2005) Por último, en los cánceres THYM, THCA y en particular LAML similar a otros estudios se observó que gran parte de los pacientes exhibían un perfil carente de CNAs (Franch-Expósito et al., 2020).

## *10.2 Análisis de sobrevida*

Para realizar el análisis de sobrevida, de cada cáncer se descargó la información clínica de los pacientes y los datos de CNAs determinados por GISTIC (g) directamente de firebrowse. El estatus de CNA para los genes individuales en cada muestra correspondía a un número de copias discreto; (-1) deleción heterocigota, (-2) deleción homocigota, (0) normal, (1) duplicación y (2) amplificación profunda. En cada cáncer se consideraron 4 análisis dependiendo el tipo de variación y la magnitud.

Para entender los resultados a continuación, en este estudio el análisis “All amplifications”, se refiere a duplicaciones de material genético de una copia o más por arriba del esperado ( $g \geq +1$ ), el análisis “All deletions” ( $g \leq -1$ ) se refiere a deleciones de material genético de una copia o más por abajo del esperado ( $g \leq -1$ ), el análisis “Deep Amplifications” ( $g = +2$ ) se refiere a la ganancia de dos copias y el análisis “Deep deletions” ( $g = -2$ ) a la pérdida de ambas copias de una región genómica. En cada análisis se consideraron 3 pruebas estadísticas; VALORATE, log rank y el modelo Cox.

El análisis de sobrevida con VALORATE permitió encontrar regiones genómicas afectadas por CNAs poco exploradas y conocidas en cáncer. También permitió generar hipótesis y corroborar información para algunos biomarcadores recientemente sugeridos.

De este análisis se observó que las regiones más significativas eran representadas por deleciones y gran parte de ellas fueron encontradas en los cánceres de GBM, GBMLGG, LGG, KIRC, PAAD, MESO y UCEC. En todos los cromosomas se encontraron regiones asociadas a sobrevida, pero con frecuencia en los cromosomas 1, 3, 8, 9 y 12. La alteración más recurrente entre los diferentes cánceres fue la deleción 9p21.3, una región muy conocida donde se ubican los genes *MTAP*, *CDKN2A*, *CDKN2B* entre otros, casi siempre con una disminución en la expresión y asociada a un mayor riesgo.

### *10.3 Regiones conocidas asociadas a sobrevida*

De las regiones fuertemente asociadas a sobrevida, la mayoría eran conocidas en cáncer, aun así, estas contenían datos interesantes y sirvieron para validar el análisis. De las regiones conocidas más significativas ( $p < .0001$ ), en LGG, por ejemplo, se encontró la deleción 10q21.1 como indicador de alto riesgo en 94 pacientes. Esta región es más conocida en gliomas de alto grado, varios genes contenidos en ella estaban sub expresados (*IPMK*, *TFAM*, *UBE2D1* y *CISD1*) mientras que dos (*BICC1* y *ZWINT*) estaban sobre expresados. *ZWINT* codifica para una proteína del cinetocoro que asegura la correcta segregación de los cromosomas, este gen está sobreexpresado y ligado a mal pronóstico en varios tumores incluido gliomas (Waks et al., 2016). *BICC1* codifica para una proteína de unión al ARN, conocida por reprimir la traducción de ARNm específicos durante la embriogénesis. (Park et al., 2016) Su sobreexpresión se asocia con un pronóstico bajo en cáncer de mama, gástrico y en ciertos estudios en gliomas (Amjad et al., 2020; Zhao et al., 2020). El papel de *BICC1* y la regulación de *ZWINT* para este cáncer no se ha esclarecido por completo. *CISD1* es un transportador de hierro, en gliomas su sobreexpresión es indicador de pronóstico favorable, lo que hace sentido a los resultados de este análisis. (Liu et al., 2020). *UBE2D1* es una enzima conjugadora de ubiquitina, poco explorada en gliomas (Lei et al., 2018). De los demás genes, contrario a este análisis solo la regulación positiva se ha asociado con un pronóstico desfavorable en gliomas.

Otras regiones conocidas con asociación pronóstica fuerte se encontraron en PAAD en el análisis de “All Amp”. Las regiones 17p11.2, 17p12 y 17p13.1 estaban amplificadas en cerca de 9 pacientes. En la región 17p13.1 se encontraron sub expresados los genes XAF1 y SLC39A5. El gen XAF1 es antagonista de la familia de inhibidores de apoptosis, ésta sub expresado y correlacionado con mal pronóstico en varios cánceres incluido páncreas, casi siempre por hipermetilación del promotor. (Jeong et al., 2018) SLC39A5 codifica para un transportador de solutos unido a membrana, su sobreexpresión favorece la progresión de cáncer de mama y esófago(Chandler et al., 2016; Jin et al., 2015). En PAAD, SLC39A5 tiene una alta expresión en células  $\beta$  pancreáticas y es regulador importante de zinc el cuál sirve para la síntesis de insulina. En ratones diabéticos el gen se encuentra sub expresado. (Wang et al., 2019)

En BRCA se encontraron muy asociadas a sobrevida las deleciones 18p11.21 (CXADRP3, ZNF519, SEH1L, CEP192), 18p11.31 (DLGAP1) y 18p11.32 (YES1, TYMS, ADCYAP1), en alrededor de 287 paciente, todas asociadas con alto riesgo. YES1 es muy conocido en BRCA, codifica para una tirosina quinasa, tanto su amplificación como deleción se han descrito para este cáncer. (Bilal et al., 2010) CEP192 es una proteína esencial para la maduración del centrosoma durante la mitosis, su relación con cáncer no ha sido establecida.(Gomez-Ferreria et al., 2007). ZNF519 y DLGAP1 en cambio, estaban fuertemente sub expresados, éstos casi no se han explorado en BRCA.

En CESC se encontró la deleción 19p13.3 asociada a un mayor riesgo en 87 pacientes. Esta región incluyó varios genes significativamente sub expresados ( $p < .0001$ ); SLC39A3, SGTA, ZNF554, ZNF555, THOP1, TIMM13, LMNB2, muchos de ellos han sido poco explorados para este cáncer. SLC39A3 codifica para un transportador metálico unido a membrana, solo un estudio ha relacionado su sobreexpresión con estados de displasia cervical avanzada. (Kneller et al., 2007) SGTA por su parte codifica una co-chaperona que posee la repetición tetratricopéptido. (Zhu et al., 2014) La sobreexpresión de SGTA se ha relacionado con progresión en ciertas neoplasias. En línea celular de cáncer cervical se encontró que el *knockout* de SGTA reducía la proliferación celular. (Zhu et al., 2014) A diferencia de estos genes, la sub-expresión de ZNF554 y THOP1, es la que se relacionó con un pronóstico desfavorable, en este caso en gliomas y cáncer de pulmón de

células no escamosas, respectivamente, indicando un posible rol como gen supresor de tumor.(Balogh et al., 2020; Qi et al., 2014)

En KIRC también se halló la delección 19p13.3, como indicador de muy alto riesgo en 16 pacientes. Varios genes en esta región incluidos; CCDC94, TNFAIP8L1, UHRF1, TICAM1, STAP2 y UBXN6, estaban sub expresados a excepción de UHRF1. Casi ninguno de estos genes es conocido en KIRC. TNFAIP8L1 (TIPE1), codifica para la proteína 8 inducida por TNF- $\alpha$  y ésta relacionada con el desarrollo de varias neoplasias (Padmavathi et al., 2018). TICAM1 (TRIF) es una proteína adaptadora que contiene el dominio TIR a través del cual interactúan receptores tipo Toll.(Oshiumi et al., 2003) La sobreexpresión de TICAM1 suprime la poliposis en ratones APC<sup>Min/+</sup> mientras que el *knockout* reduce la sobrevida de estos(Ono et al., 2017). STAP-2 regula vías de señalización intracelular, su sobreexpresión es la que se ha relacionado con crecimiento de varias neoplasias.(Kitai et al., 2017) UHRF1, es un conocido regulador epigenético, ésta sobreexpresado en varios cánceres incluido cáncer de riñón, donde se ha sugerido como potencial blanco terapéutico. (Ma et al., 2015)

En MESO se hallaron las delecciones 9p21.3, que incluye varios genes conocidos y las regiones 10p15.3 y 10p15.1 estas últimas estaban presentes en cerca de 28 pacientes, todas asociadas con alto riesgo. La región 10p15.1 incluía al ARNnci LINC00701 para el cual no había datos de expresión y al gen DIP2C el cual estaba un poco subexpresado ( $p < .05$ ). La pérdida de DIP2C en cáncer de pulmón, mama y colorrectal conlleva a cambios en la metilación de ADN y favorece la transición epitelio mesenquimal. (Li et al., 2017)

De las regiones conocidas, algunas un poco menos significativas  $p < .005$  fue interesante encontrar genes con expresión diferencial muy marcada. En BRCA, por ejemplo, se observó la amplificación de 4q31.23 en 106 paciente, esta región asociada a alto riesgo incluye a los genes TMEM184C, y PRMT10 fuertemente sobre expresados ( $p < .00001$ ). Ambos genes han sido poco estudiados. PRMT10 codifica para una N-methyltransferasa cuyo blanco no se conoce y tampoco se ha reportado en cáncer de mama, solo en cáncer de próstata Harada N. et. al mostraron que PRMT10 favorecía la proliferación celular dependiente del receptor de andrógenos (Harada et al., 2015; Michalak & Visvader, 2016).

En BLCA fueron identificadas varias regiones de amplificación profunda. En particular la región 3p25.1 observada en cerca de 33 pacientes, ésta contenía varios genes significativamente sobreexpresados ( $p < .00001$ ), entre ellos; TMEM43, LSM3, XPC y CHCHD4. Ninguno de los genes es conocido en BLCA, a excepción de XPC del cual algunos polimorfismos se han relacionado con un mayor riesgo (Sankhwar et al., 2016). La sobre expresión de estos genes induce metástasis en varios tipos de cáncer, sin embargo, en este análisis la amplificación de 3p25.1 se asoció con una mejor sobrevida. Lyng H, et al. mostraron que la ganancia de LSM3 se correlacionaba con mayor expresión del gen y esta con grupos de cáncer cervical de rápida proliferación, fosforilación oxidativa (OXPHOS) y metástasis. (Lyng et al., 2006) CHCHD4 regula la OXPHOS y promueve la proliferación de células cancerosas en condiciones de hipoxia.(Thomas et al., 2019) La sobreexpresión de TMEM43 se ha relacionado con la progresión de tumores de cerebro (Jiang et al., 2017).

En UCEC se encontró la amplificación profunda de alto riesgo, 8q13.1 en 14 pacientes. En dicha región se hallaron algunos genes sobre expresados incluidos ARMC1, MTFR1 y VCPIP1, casi ninguno de ellos relacionados con cáncer. MTFR1 es un regulador de la fisión de mitocondrias (Wang et al., 2015). ARMC1 es una subunidad del complejo MICOS/MIB el cual regula el mantenimiento y organización de crestas mitocondriales.(Huynen et al., 2016) La ausencia de ARMC1 en mitocondrias causa que estas aparezcan fragmentadas y con movilidad reducida.(Wagner et al., 2019) VCPIP1 (VCIP135) es una desubiquitinasa involucrada en el ensamblaje del aparato de Golgi durante la mitosis.(Kuznetsova et al., 2007; Uchiyama et al., 2002)

#### *10.4 Regiones nuevas asociadas a sobrevida*

En cuanto a las regiones nuevas asociadas a sobrevida, se encontraron varias regiones de amplificación, una de las más significativas ( $p < .0001$ ) se encontró en SARC con la amplificación profunda de 1p31.3 observada en 12 pacientes. Esta es una región de alto riesgo con varios genes sobre expresados, entre ellos; PGM1, ITGB3BP, LEPR y LEPROT. PGM1 es un marcador de sarcomas histiocíticos (Zhao et al., 2015). Se ha encontrado que la expresión de PGM1 se requiere para la acumulación de glucógeno en varias líneas celulares de cáncer como alternativa para proliferar en condiciones de

privación de glucosa y estados de hipoxia (Bae et al., 2014). LEPR por su parte es un promotor de proliferación celular y metástasis en varias neoplasias, entre ellas; mama, colorrectal, próstata y carcinoma hepatocelular (Huang et al., 2021; Lin et al., 2018). En condrosarcomas se ha sugerido como blanco terapéutico de metástasis (Huang et al., 2021; Lin et al., 2018).

También se encontraron regiones nuevas en ACC en el análisis de amplificaciones, en éste se observaron las regiones 11p11.2 y 22q13.2 en cerca de 8 pacientes, ambas indicadoras de alto riesgo. En esta última región se halló el gen PHF5A significativamente sobre expresado. Este gen es un regulador de empalme alternativo de pre-RNA, juega un rol oncogénico en varios tipos de cáncer, en los cuáles su sobreexpresión igual es indicador de mal pronóstico. (Yang et al., 2018)

En otros cánceres se encontraron regiones nuevas asociadas a sobrevida no tan significativas pero interesantes. En PCPG, por ejemplo, se encontró la amplificación de 1q25.1 asociada a sobrevida ( $p < .01$ ) en cerca de 26 pacientes. Esta región de bajo riesgo incluye al gen PRDX6 el cual estaba fuertemente sobre expresado ( $p < .00001$ ). La amplificación profunda de esta región también se asoció con sobrevida, aunque en sólo 6 pacientes. PRDX6 se encuentra sobreexpresado en varios cánceres incluido el cervical, ovario, y mama (Hu et al., 2020).

En SKCM se encontraron significativas ( $p < .005$ ) las regiones de amplificación 3q26.31 y 3q26.32, ambas indicadoras de mejor pronóstico en 71 pacientes. Estas regiones incluían varios genes poco explorados (Snou13, MIR4789, RNA5SP147, TBL1XR1, NAALADL2), la mayoría de éstos sin datos de expresión. Cabe mencionar que estas regiones se han reconocido como sitios frágiles frecuentemente alterados en cáncer. La región 3q26.31 contiene al gen NAALADL2, la pérdida de este gen se ha reportado en varios tumores, aunque también su ganancia recientemente se encontró asociada con cánceres de próstata más agresivos. (Simpson et al., 2021) En esta región además se han descrito varios oncogenes (Simpson et al., 2021).

En THYM se encontraron asociadas a sobrevida ( $p < .005$ ) las amplificaciones 1q23.1 y 1q23.2, éstas fueron observadas en 24 pacientes, ambas asociadas a muy alto riesgo. Casi todos los genes contenidos en estas regiones estaban sobre expresados

(FCRL2, DUSP23, FCRLB, NR1L3) con excepción de SPTA1 y SLAMF1 los cuales tenían una regulación negativa.

### *10.5 Análisis de firmas genómicas*

Posterior al análisis de sobrevida se construyeron firmas genómicas para determinar si los genes afectados por CNAs en conjunto lograban estratificar a los pacientes en diferentes grupos de riesgo. Para cada cáncer se construyeron firmas genómicas considerando las regiones encontradas significativas según el tipo de variación analizada y firmas que combinaban regiones de amplificación y delección para ver si éstas mejoraban aún más la estratificación, en estas se evaluaron además diferentes formas de agrupar a los pacientes.

Del análisis de firmas genómicas se encontró que, para algunos cánceres, entre ellos; BLCA, GBMLGG, HNSC, KIPAN, LGG, PRAD, LUSC, LIHC, GBM y SARC era posible estratificar de forma significativa a los pacientes en grupos de bajo y alto riesgo. Estos resultados son interesantes ya que, aunque se han descrito CNAs pronósticos en diferentes cánceres, para la mayoría de éstos aún no se describen firmas basadas en números de copias que puedan estratificar correctamente a los pacientes (Wang et al., 2021).

Para otros cánceres en cambio, las firmas genómicas no siempre distinguían los grupos de alto y bajo riesgo, a veces solo identificaban al grupo de alto riesgo, entre ellos; ACC, KICH, KIRC, KIRP, LAML, MESO y PAAD. De estos cánceres casi no hay estudios sobre firmas pronósticas basadas en CNAs.

En otros cánceres, era mejor o solo se podía estratificar cuando se consideraba solo los CNAs profundos, entre ellos BRCA, OV, STES, CESC, COADREAD, ESCA y STAD. Algunos de estos cánceres cuentan con perfiles genómicos complejos, como en el caso de cáncer de ovario y mama, (Macintyre et al., 2018), para estos casos las firmas genómicas basadas en regiones de amplificación y delecciones profundas mostraron mayor capacidad para estratificar a los pacientes.

Por último, se encontró que para algunos cánceres (UVM, UCS, TGCT, READ, DLBC, THYM, THCA) no era posible estratificar a los pacientes en grupos de riesgo cuando se analizaban con las firmas de CNAs profundos. Muchos de estos cánceres cuentan con un número reducido de muestras y para algunos de estos cánceres los CNAs profundos son raros.

### *10.6 Análisis de enriquecimiento*

Además del análisis de firmas genómicas, se realizó un análisis funcional para identificar en que procesos biológicos y vías de señalización estaban sobre representados los genes asociados a sobrevida y con expresión diferencial significativa ( $p < .05$ ). Para realizar el análisis se utilizaron varias herramientas y algunas bases de datos comúnmente utilizadas, entre ellas; KEGG, REACTOME, WikiPathways y Gene Ontology.

En general, se encontró gran concordancia entre las herramientas y bases de datos utilizadas en cuanto a los procesos biológicos y/o vías de señalización identificadas, salvo algunas excepciones. Los resultados del análisis funcional en un cáncer variaban según el tipo de alteración analizada, en la mayoría de los casos las deleciones proporcionaban más información sobre vías de señalización que las amplificaciones. También se distinguieron vías de señalización y procesos biológicos característicos según el tipo de cáncer. A continuación, se mencionan algunos de los resultados del análisis funcional considerando los términos sobrerrepresentados con valor  $p$  ajustado  $< 0.05$ .

En PAAD en el análisis de “All del” se encontraron varios genes enriquecidos en regulación de la vía JNK, respuesta inmune innata, regulación de leucocitos, actividad de MAPK, modificación de proteínas, señalización por interferones y señalización por factor de necrosis tumoral (TNF). Todas estas vías se han reportado para este cáncer. La vía MAPK esta frecuentemente alterada en células de cáncer pancreático junto con su efector JNK el cual efectúa tanto actividades supresivas de tumor como pro-oncogénicas (Furukawa, 2015). Estudios previos validan el resultado de nuestro análisis en páncreas generando confianza en los demás resultados.

En BRCA, en particular en el análisis “Deep del” se encontraron varios genes enriquecidos en las vías de señalización de leptina, PDGFR-beta, IL-(5,9), BDNF y receptor de andrógenos. La sobreexpresión de AR es frecuente en los diferentes subtipos de cáncer de mama (Kensler et al., 2019), las IL- (5,9) contribuyen al desarrollo de metástasis en estos pacientes y de PDGFR-b todavía se ésta explorando su papel en los diferentes subtipos (Jitariu et al., 2018), pero su sobreexpresión es un factor pronóstico desfavorable. Estos resultados concuerdan con lo reportado en literatura indicando que los genes delecionados en nuestro análisis son relevantes en vías de señalización desreguladas en BRCA.

En MESO en el análisis “All amp” se identificaron varios procesos biológicos relacionados con empalme de ARN, actividad catalítica del espliceosoma tipo U2 y metabolismo de ARN. La desregulación del empalme de mRNA es un evento conocido en los tipos de mesotelioma peritoneal maligno y en años recientes se ha sugerido como una estrategia terapéutica para el manejo de estos pacientes (Sciarrillo et al., 2019). Aunque las alteraciones en número de copia son también frecuentes en estos pacientes su relación con eventos de empalme es poco explorada. Nuestro análisis sugiere que los genes amplificados participan en los procesos de empalme por lo que podrían ser de utilidad en el manejo de estos pacientes.

En UCEC, el análisis funcional de “All amp”, se caracterizó por procesos biológicos relacionados con actividad helicasa de ADN, transición fase G2/M, actividad RNA polimerasa II, varios procesos metabólicos, sistemas de reparación de DNA de doble cadena y reparación por escisión de nucleótidos (TC-NER y GG-NER). En estudios previos la formación de cánceres endometriales se ha asociado con la exposición de estrógenos, los cuáles causan aductos y daños oxidativos en las bases del ADN (Krupa et al., 2011; Sobczuk et al., 2012). Estos daños suelen ser eliminados por sistemas de reparación entre ellos; el de escisión de bases (BER) y el sistema de reparación por escisión de nucleótidos (NER)(Sobczuk et al., 2012). La desregulación de estos sistemas es un factor importante en la carcinogénesis del cáncer endometrial. Nuestros resultados indican que los genes amplificados están asociados con vías de señalización relevantes de cáncer endometrial y

algunos procesos que en otros cánceres indican un fenotipo más agresivo (G2/M)(Oshi, Newman, et al., 2020; Oshi, Takahashi, et al., 2020).

### *10.6 Análisis de respuesta a fármacos*

Por último, se realizó un análisis para encontrar evidencia de vulnerabilidad de los CNAs significativos contra algún agente terapéutico ya sea para el tratamiento del cáncer para el cual la alteración fue identificada o para el tratamiento de otros cánceres. El análisis de fármacos se realizó a través de la base de datos The Cancer Genome Interpreter (CGI) usando como datos de entrada los genes con su respectiva variación, que estaban asociados a sobrevida y tenían una expresión diferencial significativa ( $p < .05$ ). Esta base de datos no diferencia la magnitud de la variación, por tanto, todas las alteraciones fueron analizadas como amplificaciones o deleciones.

De este análisis solo 16 genes tenían una respuesta potencial a diferentes clases de fármacos. Las vulnerabilidades más frecuentes se encontraron para las deleciones de CDKN2A y CDKN2B las cuales estaban presentes en varios cánceres, entre ellos GBM, GBMLGG, HNSC, KIPAN, KIRC, LGG, MESO, PAAD y SARC. De los resultados del análisis se encontró un estudio preclínico donde sugiere que estos pacientes podrían responder a inhibidores CDK4/6. (Stone et al., 2012). También, se encontró un ensayo temprano con lloerasertib (inhibidor AURKA-VEGF) para pacientes con la deleción CDKN2A, este ensayo de fase II todavía está en proceso de evaluación (NCT02478320).

La mayoría de los resultados que se obtuvieron del análisis de fármacos ya han sido ampliamente discutidos en literatura, pero en alguno de ellos analizando otro tipo de alteraciones. De forma general este análisis sirvió sobre todo para conocer el panorama o la evidencia clínica de algunos fármacos contra los biomarcadores seleccionados.

En BLCA la deleción profunda de RB1 observada en 9.3% (37/398) de los pacientes, se evaluó en un ensayo temprano en pacientes con cáncer de vejiga con invasión muscular. El ensayo demostró que los pacientes que tenían alteraciones en genes asociados a reparación del DNA entre ellos ATM, RB1 y FANCC predecían un beneficio clínico después de quimioterapia con cisplatino. (PMID 26238431).

En CESC se encontró que la delección STK11 observada en 35% (97/277) de los pacientes podría tal vez beneficiarlos de los tratamientos con inhibidores mTOR. STK11/LKB1 es un gen supresor de tumor que codifica una serina/treonina quinasa que regula positivamente la vía AMPK involucrada en el control de crecimiento y metabolismo celular. La activación de AMPK resulta en un decremento de la actividad de mTOR. En un estudio pre-clínico con modelos murinos del Síndrome de Peutz-Jeghers, caracterizados por deficiencia de STK11/LKB1 presentaban la vía mTOR hiperactivada. Tumores y tejidos derivados de estos ratones respondían a inhibidores mTOR (PMID 19541609).

En AML, la delección de SUZ12, aunque observada en solo 4.76% de los pacientes, de acuerdo con evidencia pre-clínica podría beneficiar a estos pacientes el empleo de inhibidores BET, debido a que la inactivación del SUZ12 desencadena un cambio epigenético que sensibiliza a las células de cáncer a los inhibidores de bromodominio (PMID 25119042). Revisando fuera de la base de datos se encontraron más estudios de pacientes con neoplasias mieloide que eran tratados con estos inhibidores, aunque los resultados de estos estudios son prometedores, todavía falta definir bien los regímenes de tratamiento a los cuales estos agentes puedan ser más efectivos y menos tóxicos. (N. C. Chen et al., 2021)

#### *10.7. Comparación de pruebas estadísticas para el análisis de sobrevida.*

Para los análisis de sobrevida se usaron las herramientas VALORATE, el modelo Cox y la prueba log rank. En la mayoría de los análisis se encontró mayor correlación entre VALORATE y el modelo Cox.

Algunas excepciones se encontraron en los cánceres GBM (“All Amp” y “All del”), PCPG (“Deep Del”), READ (“Deep Amp”), TGCT (“All Del y All Amp”) y THCA (“Deep Amp”), en estos la correlación fue mayor entre la prueba log rank y el modelo Cox. En cambio, en los cánceres ESCA (“All Amp”) y UVM (“Deep Del”), la correlación fue mayor entre VALORATE y la prueba log Rank.

## 10.8 Repositorio

Todos los resultados de este proyecto y actualizaciones se pueden consultar en el siguiente sitio: [https://tecmx-my.sharepoint.com/:f:/g/personal/a00829642\\_itesm\\_mx/EpCqPwna7t9Bm-swF\\_8ZOeABmg3cNbgs43FRd\\_gn4yfykg?e=dpogzx](https://tecmx-my.sharepoint.com/:f:/g/personal/a00829642_itesm_mx/EpCqPwna7t9Bm-swF_8ZOeABmg3cNbgs43FRd_gn4yfykg?e=dpogzx)

## 10.9 Perspectivas futuras

Con los resultados de este proyecto se espera aportar mayor conocimiento sobre los CNAs y sus genes diana, implicados en la progresión de diversos cánceres, así como nuevos biomarcadores que mejoren la estratificación y, por ende, el pronóstico y calidad de vida de los pacientes.

A pesar de que en los años actuales se han identificado muchos biomarcadores de CNAs con valor pronóstico mediante análisis *in-silico*, pocos logran superar los posteriores procesos de validación debido a la falta de robustez en estos, por lo que en este trabajo favorecemos el empleo de métodos más exactos y esperamos que nuestros resultados puedan servir para corroborar o complementar estudios previos.

En el futuro será necesario hacer análisis más integrativos para entender mejor los mecanismos a través de los cuales actúan los CNAs, ya que muchas veces los cambios de expresión en los genes blanco no son proporcionales al número de copias, dificultando la interpretación de éstos.

Este proyecto también reconoce algunas limitaciones ligadas a la base de datos The Cancer Genome Atlas, comunes a otros estudios y que podrían considerarse en estudios futuros, entre ellas; que las muestras de cáncer, aunque son inclusivas de distintas poblaciones, no son representativas para algunas de ellas, y que la población control seguían siendo pacientes con cáncer.

## 11. Apéndice

### 11.1 Apéndice A

La tabla 10 muestra un resumen de las regiones asociadas a sobrevida en diversos cánceres, considerando los primeros 12 genes significativos ( $p < 0.05$ ) en los cuatro análisis de VALORATE.

Tabla 10 Resumen de regiones asociadas a sobrevida en cáncer

	<b>All Amplifications</b>	<b>All Deletions</b>	<b>Deep Amplifications</b>	<b>Deep Deletions</b>
<b>ACC</b>	11p11.2, 22q13.2, 11q12.3, 11q12.2	17p13.1	12q14.3, 19q12, 12q14.2	1p36.22, 22q12.1, 4q34.3, 1p36.23, 22q12.1, 9p21.3
<b>BLCA</b>	9q33.1, 9q32, 9q31.1, 9q22.2	18q21.1, 18q12.3, 18q21.2	10q21.1, 2q12.1, 2q11.2, 3p25.1, 11q13.2, 11p14.1, 3p25.1, 6p22.3, 6p22.2	5q23.3, 5q31.1
<b>BRCA</b>	4q24, 4q31.22, 4q31.23, 22q11.21, 8q24.3, 4p16.3, 8q22.1, 8q21.3, 4q24	18p11.21, 18p11.31, 18p11.32	8p22, 12q13.2, 3p26.1, 8q21.3	17p12, 12q24.31, 17p12, Xq21.31, 17q12, 17q21.2
<b>CESC</b>	21q21.3, 21q21.2, 17p12, 21q22.11, 17p13.1	19p13.3	11q21, 11q22.1	11q22.1, 8p23.3, 18q21.2, 3p24.2, 3p22.2, 8p23.2
<b>CHOL</b>	11q12.2, 11q12.2, 17q12	11q22.1, 11q23.1, 11q23.2	NS	NS
<b>COAD</b>	9q34.3, 9q34.2, 5q11.1	1p31.1, 4p16.3, 16q23.3, 16q24.1	20p13	18q21.1, 18q12.3, 18q21.33
<b>COADREAD</b>	9q34.2, 9q34.3,	16p12.1, 16p12.3, 16p12.2,	20p13	6q26, Xp22.2, 18q21.1, 10q11.23, Xp22.2, Xp22.31, 1p33, 5q15, 8p11.21, 4q22.1, 18q12.3
<b>DLBC</b>	1q21.3	18q22.3, 18q23,	NS	NS
<b>ESCA</b>	16q22.3, 16q22.2, 9q34.3, 16q23.1,	7p15.2, 20p12.1, 4p16.1, 20p12.3, 20p12.2	17q23.2, 17q21.2, 6p22.1, 7p13	4q22.1, 17p12, 10p15.3, 19q11, 10p15.3, 18p11.31, 17q24.3
<b>GBM</b>	7p11.2, 7q11.22, 7q11.21	10p11.21, 10p11.1, 10q21.1, 10q11.23	7p11.2, 12q14.2, 12q14.3	9p21.3, 5p15.33, 6q26, 1p36.13
<b>GBMLGG</b>	20p13	9p21.3, 10q21.1, 10q21.2, 10q11.23, 10q21.3	7p11.2,	9p21.3,
<b>HNSC</b>	7p11.2, 7q11.21, 7p22.3	8p23.2, 19p13.2,	11q13.4, 16q23.1, 10q11.21, 12q12, 19q13.43, 13q21.33,	14q32.32, 9q34.3, 14q32.31, 4q21.3, 9p21.3

			8p12	
<b>KICH</b>	5p15.33	16q13, 16q22.1, 16q23.1, 16q12.2,	NS	NS
<b>KIPAN</b>	1p31.1, 1q21.1, 1q21.2	9p21.3, 9p22.1	12p13.33, 4q12, 11q22.2, 11q22.1, Xq28	9p21.3, 3q13.31, 5p15.33, 14q24.3, 16q23.1, 14q24.3,
<b>KIRC</b>	12p13.31, 12p12.1, 12p11.23, 12p13.33	19p13.3,	11q22.1, 11q22.2, 5p15.31, Xq28	9p21.3, 10q23.31, 14q21.2, 9p24.1, 6q27, 9p23, 14q24.3
<b>KIRP</b>	8q24.11, 8q22.2, 8q22.3, 8q22.1	3q13.31, 17p13.2, 17p13.1, 17p13.3	NS	14q11.2,
<b>LAML</b>	21q22.3, 21q22.2,	5p15, 5q33.1, 5q33.2	21q22.3, 21q22.2, 21q22.13	NS
<b>LGG</b>	12q14.1, 20p11.22, 20p11.23, 7p11.2	10q21.1,	7p11.2, 12q14.1, 12q13.3	9p21.3,
<b>LIHC</b>	5q31.3, 5q31.2, 5q31.1, 5q33.1	1q41,	12q23.3, 12q24.22, 11q13.2, 12q23.1, 12q13.3, 12q22	8p11.21, 8p11.1, 6q23.3, 8p11.22, 17p11.2, 8p22, 16q12.2,
<b>LUAD</b>	19q13.2,	1q22, 17q21.32, 2q13,	7p14.1, 1p32.1, 7p13, 4q12, 7p12.3, 7p12.1	1p21.3, 1p22.1, 1p12, 6q24.1, 6q23.3, 21q11.2
<b>LUSC</b>	21q22.3	12q24.21, 2q33.1, 2q33.2, 3q28, 12q24.22, 2q33.2	2q21.2, 15q26.1, 20q13.33, 6p21.1, 1q25.2	10q22.2, 14q12, 3p14.3, 14q32.12, 10q22.2, 3p21.31
<b>MESO</b>	1p32.1, 1p31.3, 1p32.2	9p21.3, 10p15.3, 10p15.1	17q25.3, 17q25.1,	9p21.3,
<b>OV</b>	8q21.3, 8q22.1	3q26.2	14q32.13, 14q32.11, 8p11.21, 14q32.2	9p24.1, 6q24.1, 4q32.3, 9p23, 9p24.2,
<b>PAAD</b>	17p11.2, 17p12, 17p13.1,	16p13.3, 2q36.3, 2q37.3,	1q31.1, 1q31.2,	9p21.3
<b>PCPG</b>	9p21.2, 9p21.3, 1q25.1, 21q22.11, 21q22.3, 3p14.1	12p12.3, 12p13.1, 9p21.3,	1q25.1, 1q25.3	NS
<b>PRAD</b>	8p12	3p12.1, 3q21.2, 3q22.1, 1q42.3	9p13.2, Xq27.1, 9q22.31, 9q22.32	16p13.3, 1q32.2, 17p13.2
<b>READ</b>	8p23.3, 15q26.3,	8q21.3, 8q13.3, 8q21.2, 8q21.11,	13q14.2, 13q14.12, 13q14.13, 13q14.11,	16p13.3
<b>SARC</b>	9p21.1, 9p21.2, 9p13.3	15q21.2, 15q21.1,	1p31.3, 8q24.13, 6q16.2, 1p31.1, 9p21.1	9p21.3
<b>SKCM</b>	3q26.31, 3q26.32, 3p13, 3p14.1, 3p22.3, 3p22.2	3p22.1	12p12.1, 12p11.23, 12p12.3, 22q13.1,	Xp22.32, 10q21.1, 20p12.1, 10q23.31, 16q12.1, 11q23.3, 1p36.22
<b>STAD</b>	3p14.3, 3p21.1, 3p24.3, 3p24.2, 3p24.1,	8q24.12, 8q24.13,	9p22.1, 11q13.5, 9p24.1, 11q14.1,	16q24.1, 16q23.2, 16q24.2, 16q23.3, 4q21.21
<b>STES</b>	3p14.3, 16q23.1, 16q24.2, 16q22.2	8q24.13, 8q24.21, 8q24.13,	17q23.2, 8q23.2, 8q23.1, 17q23.2, 9p22.1,	10p15.3, 7q36.3, 4q21.21, Xp11.21
<b>TGCT</b>	12q12, 12p11.1, 6p21.1	18q12.1, 18q21.2,	12q12, 12p11.1,	10q26.3,
<b>THCA</b>	6p22.2, 4q34.3, 4p16.1, 4q32.3	6q27, 6q24.1, 6q26, 6q25.3,	NS	13q12.3

<b>THYM</b>	1q23.1, 1q23.2, 1q23.3, 1q22,	Xp22.2, 6p21.31, 18p11.22, 16q12.1	NS	9p21.3
<b>UCEC</b>	7q34, 7q31.2	8q24.3	8q13.1, 17q12	4q22.1, Xp21.1, 2q22.1, 22q13.2, 10q23.31, 22q13.31, Xp21.2, 4q25
<b>UCS</b>	1q31.2, 22q12.2, 22q11.23, 14q21.3,	4p12, 17q23.1, 17q21.32, 4p13, 17q23.2	11q13.1,	NS
<b>UVM</b>	8p11.1, 8p11.21,	3p25.2, 3q29, 3q22.1, 3q21.1, 3q27.3, 3p14.1, 3q21.3, 3q24, 3q25.1, 3q26.33, 3q13.3	8p11.21, 8p11.1, 8q11.23, 8q11.1, 8q11.21	NS

## 11.2 Apéndice B

La figura 19 muestra un mapa de calor de los valores  $p$  estimados para 15,000 locis de CNAs, en los análisis de sobrevida con VALORATE.

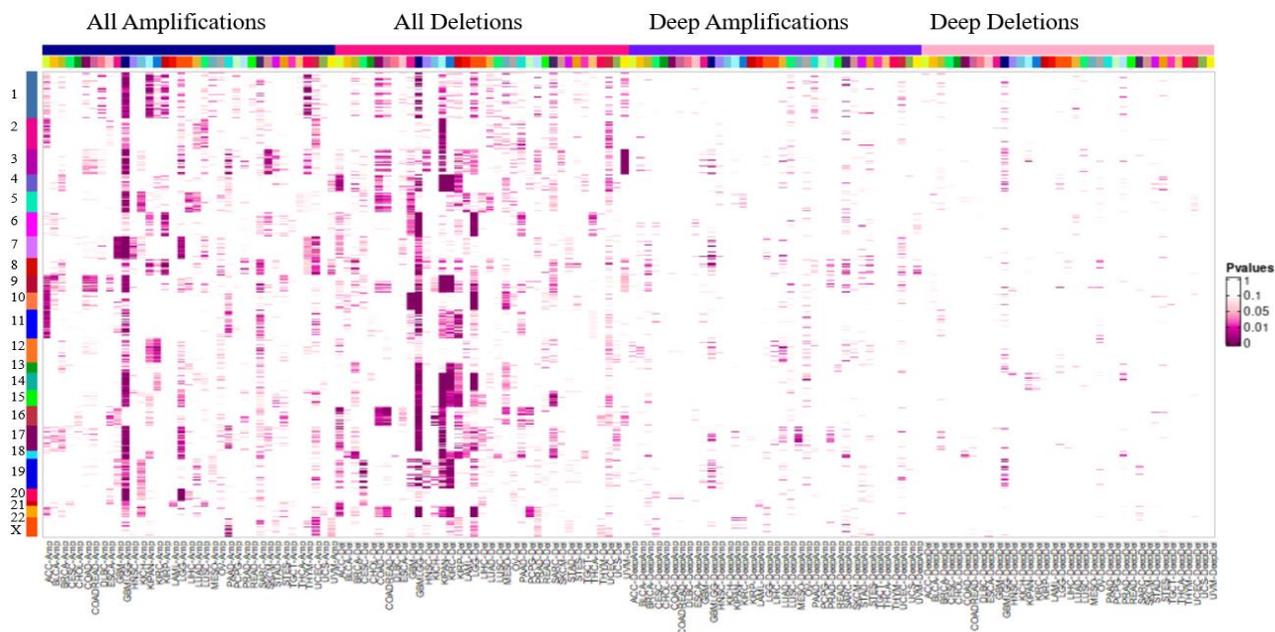


Figura 19 Mapa de calor de los valores  $p$  estimados en el análisis de sobrevida para regiones alteradas por CNAs en diversos cánceres

## 12. Bibliografía

Akbari, A., Agah, S., Heidari, M., Mobini, G. R., Faghiloo, E., Sarveazad, A., & Mirzaei, A. (2017). Homeodomain Protein Transforming Growth Factor Beta-Induced Factor 2 Like, X-Linked Function in Colon Adenocarcinoma Cells. *Asian Pac J Cancer Prev*(18), 2101-2108. <https://doi.org/10.22034/APJCP.2017.18.8.2101>

Amjad, E., Asnaashari, S., Sokouti, B., & Dastmalchi, S. (2020). Systems biology comprehensive analysis on breast cancer for identification of key gene modules and genes associated with TNM-based clinical stages. *Sci Rep*, 10(1), 10816. <https://doi.org/10.1038/s41598-020-67643-w>

Bae, E., Kim, H. E., Koh, E., & Kim, K. S. (2014). Phosphoglucomutase1 is necessary for sustained cell growth under repetitive glucose depletion. *FEBS Lett*, 588(17), 3074-3080. <https://doi.org/10.1016/j.febslet.2014.06.034>

Bai, Y., Qiao, L., Xie, N., Shi, Y., Liu, N., & Wang, J. (2017). Expression and prognosis analyses of the Tob/BTG antiproliferative (APRO) protein family in human cancers. *PLoS One*, 12(9), e0184902. <https://doi.org/10.1371/journal.pone.0184902>

Balogh, A., Reiniger, L., Hetey, S., Kiraly, P., Toth, E., Karaszi, K., . . . Than, N. G. (2020). Decreased Expression of. *Int J Mol Sci*, 21(16). <https://doi.org/10.3390/ijms21165762>

Beroukhi, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., . . . Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283), 899-905. <https://doi.org/10.1038/nature08822>

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: survival analysis. *Crit Care*, 8(5), 389-394. <https://doi.org/10.1186/cc2955>

Bhattacharya, A., Bense, R. D., Urzúa-Traslaviña, C. G., de Vries, E. G. E., van Vugt, M. A. T. M., & Fehrmann, R. S. N. (2020). Transcriptional effects of copy number alterations in a large set of human cancers. *Nat Commun*, 11(1), 715. <https://doi.org/10.1038/s41467-020-14605-5>

Bilal, E., Alexe, G., Yao, M., Cong, L., Kulkarni, A., Gijjala, V., . . . Bhanot, G. (2010). Identification of the YES1 Kinase as a Therapeutic Target in Basal-Like Breast Cancers. *Genes Cancer*, 1(10), 1063-1073. <https://doi.org/10.1177/1947601910395583>

Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., . . . Nowak, M. A. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A*, 107(43), 18545-18550. <https://doi.org/10.1073/pnas.1010978107>

Bányai, L., Trexler, M., Kerekes, K., Csuka, O., & Patthy, L. (2021). Use of signals of positive and negative selection to distinguish cancer genes and passenger genes. *Elife*, 10. <https://doi.org/10.7554/eLife.59629>

Büttner, R., Longshore, J. W., López-Ríos, F., Merkelbach-Bruse, S., Normanno, N., Rouleau, E., & Penault-Llorca, F. (2019). Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open*, 4(1), e000442. <https://doi.org/10.1136/esmoopen-2018-000442>

Cai, Z., & Liu, Q. (2021). Understanding the Global Cancer Statistics 2018: implications for cancer control. *Sci China Life Sci*, 64(6), 1017-1020. <https://doi.org/10.1007/s11427-019-9816-1>

Carr, T. H., McEwen, R., Dougherty, B., Johnson, J. H., Dry, J. R., Lai, Z., . . . Barrett, J. C. (2016). Defining actionable mutations for oncology therapeutic development. *Nat Rev Cancer*, 16(5), 319-329. <https://doi.org/10.1038/nrc.2016.35>

Chandler, P., Kochupurakkal, B. S., Alam, S., Richardson, A. L., Soybel, D. I., & Kelleher, S. L. (2016). Subtype-specific accumulation of intracellular zinc pools is associated with the malignant phenotype in breast cancer. *Mol Cancer*, 15, 2. <https://doi.org/10.1186/s12943-015-0486-y>

Chandran, U. R., Medvedeva, O. P., Barmada, M. M., Blood, P. D., Chakka, A., Luthra, S., . . . Jacobson, R. S. (2016). TCGA Expedition: A Data Acquisition and Management System for TCGA Data. *PLoS One*, 11(10), e0165395. <https://doi.org/10.1371/journal.pone.0165395>

Chatterjee, S. K., & Zetter, B. R. (2005). Cancer biomarkers: knowing the present and predicting the future. *Future Oncol*, 1(1), 37-50. <https://doi.org/10.1517/14796694.1.1.37>

Chen, C. H., Lin, Y. J., Lin, Y. Y., Lin, C. H., Feng, L. Y., Chang, I. Y., . . . Huang, C. Y. (2021). Glioblastoma Primary Cells Retain the Most Copy Number Alterations That Predict Poor Survival in Glioma Patients. *Front Oncol*, 11, 621432. <https://doi.org/10.3389/fonc.2021.621432>

Chen, N. C., Borthakur, G., & Pemmaraju, N. (2021). Bromodomain and extra-terminal (BET) inhibitors in treating myeloid neoplasms. *Leuk Lymphoma*, 62(3), 528-537. <https://doi.org/10.1080/10428194.2020.1842399>

Derbal, Y. (2017). Perspective on the dynamics of cancer. *Theor Biol Med Model*, 14(1), 18. <https://doi.org/10.1186/s12976-017-0066-5>

Eroglu, Z., & Ribas, A. (2016). Combination therapy with BRAF and MEK inhibitors for melanoma: latest evidence and place in therapy. *Ther Adv Med Oncol*, 8(1), 48-56. <https://doi.org/10.1177/1758834015616934>

Esteves, L., Caramelo, F., Ribeiro, I. P., Carreira, I. M., & de Melo, J. B. (2020). Probability distribution of copy number alterations along the genome: an algorithm to distinguish different tumour profiles. *Sci Rep*, 10(1), 14868. <https://doi.org/10.1038/s41598-020-71859-1>

Franch-Expósito, S., Bassaganyas, L., Vila-Casadesús, M., Hernández-Illán, E., Esteban-Fabro, R., Díaz-Gay, M., . . . Camps, J. (2020). CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. *Elife*, 9. <https://doi.org/10.7554/eLife.50267>

Furukawa, T. (2015). Impacts of activation of the mitogen-activated protein kinase pathway in pancreatic cancer. *Front Oncol*, 5, 23. <https://doi.org/10.3389/fonc.2015.00023>

Gomez-Ferreria, M. A., Rath, U., Buster, D. W., Chanda, S. K., Caldwell, J. S., Rines, D. R., & Sharp, D. J. (2007). Human Cep192 is required for mitotic centrosome and spindle assembly. *Curr Biol*, 17(22), 1960-1966. <https://doi.org/10.1016/j.cub.2007.10.019>

Goossens, N., Nakagawa, S., Sun, X., & Hoshida, Y. (2015). Cancer biomarker discovery and validation. *Transl Cancer Res*, 4(3), 256-269. <https://doi.org/10.3978/j.issn.2218-676X.2015.06.04>

Gotoh, O., Sugiyama, Y., Takazawa, Y., Kato, K., Tanaka, N., Omatsu, K., . . . Mori, S. (2019). Clinically relevant molecular subtypes and genomic alteration-independent differentiation in gynecologic carcinosarcoma. *Nat Commun*, 10(1), 4965. <https://doi.org/10.1038/s41467-019-12985-x>

Harada, N., Takagi, T., Nakano, Y., Yamaji, R., & Inui, H. (2015). Protein arginine methyltransferase 10 is required for androgen-dependent proliferation of LNCaP prostate cancer cells. *Biosci Biotechnol Biochem*, 79(9), 1430-1437. <https://doi.org/10.1080/09168451.2015.1025035>

Hazra, A., & Gogtay, N. (2017). Biostatistics Series Module 9: Survival Analysis. *Indian J Dermatol*, 62(3), 251-257. [https://doi.org/10.4103/ijd.IJD\\_201\\_17](https://doi.org/10.4103/ijd.IJD_201_17)

Hieronymus, H., Murali, R., Tin, A., Yadav, K., Abida, W., Moller, H., . . . Sawyers, C. L. (2018). Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *Elife*, 7. <https://doi.org/10.7554/eLife.37294>

Hu, X., Lu, E., Pan, C., Xu, Y., & Zhu, X. (2020). Overexpression and biological function of PRDX6 in human cervical cancer. *J Cancer*, 11(9), 2390-2400. <https://doi.org/10.7150/jca.39892>

Huang, H., Zhang, J., Ling, F., Huang, Y., Yang, M., Zhang, Y., . . . Tang, J. (2021). Leptin Receptor (LEPR) promotes proliferation, migration, and invasion and inhibits apoptosis in hepatocellular carcinoma by regulating ANXA7. *Cancer Cell Int*, 21(1), 4. <https://doi.org/10.1186/s12935-020-01641-w>

Huang, W., Li, G., Wang, Z., Zhou, L., Yin, X., Yang, T., . . . Yu, H. (2020). A Ten-N. *Front Oncol*, 10, 567931. <https://doi.org/10.3389/fonc.2020.567931>

Huynen, M. A., Mühlmeister, M., Gotthardt, K., Guerrero-Castillo, S., & Brandt, U. (2016). Evolution and structural organization of the mitochondrial contact site (MICOS) complex and the mitochondrial intermembrane space bridging (MIB) complex. *Biochim Biophys Acta*, 1863(1), 91-101. <https://doi.org/10.1016/j.bbamcr.2015.10.009>

Jeong, S. I., Kim, J. W., Ko, K. P., Ryu, B. K., Lee, M. G., Kim, H. J., & Chi, S. G. (2018). XAF1 forms a positive feedback loop with IRF-1 to drive apoptotic stress response and suppress tumorigenesis. *Cell Death Dis*, 9(8), 806. <https://doi.org/10.1038/s41419-018-0867-4>

Jiang, C., Zhu, Y., Zhou, Z., Gumin, J., Bengtsson, L., Wu, W., . . . Lin, X. (2017). TMEM43/LUMA is a key signaling component mediating EGFR-induced NF-κB activation and tumor progression. *Oncogene*, 36(20), 2813-2823. <https://doi.org/10.1038/onc.2016.430>

Jin, J., Li, Z., Liu, J., Wu, Y., Gao, X., & He, Y. (2015). Knockdown of zinc transporter ZIP5 (SLC39A5) expression significantly inhibits human esophageal cancer progression. *Oncol Rep*, 34(3), 1431-1439. <https://doi.org/10.3892/or.2015.4097>

Jitariu, A. A., Raica, M., Cîmpean, A. M., & Suci, S. C. (2018). The role of PDGF-B/PDGFR-BETA axis in the normal development and carcinogenesis of the breast. *Crit Rev Oncol Hematol*, 131, 46-52. <https://doi.org/10.1016/j.critrevonc.2018.08.002>

Kensler, K. H., Regan, M. M., Heng, Y. J., Baker, G. M., Pyle, M. E., Schnitt, S. J., . . . Tamimi, R. M. (2019). Prognostic and predictive value of androgen receptor expression in postmenopausal women with estrogen receptor-positive breast cancer: results from the Breast International Group Trial 1-98. *Breast Cancer Res*, 21(1), 30. <https://doi.org/10.1186/s13058-019-1118-z>

Kitai, Y., Iwakami, M., Saitoh, K., Togi, S., Isayama, S., Sekine, Y., . . . Matsuda, T. (2017). STAP-2 protein promotes prostate cancer growth by enhancing epidermal growth factor receptor stabilization. *J Biol Chem*, 292(47), 19392-19399. <https://doi.org/10.1074/jbc.M117.802884>

Kneller, J. M., Ehlen, T., Maticic, J. P., Miller, D., Van Niekerk, D., Lam, W. L., . . . Jones, S. J. (2007). Using LongSAGE to Detect Biomarkers of Cervical Cancer Potentially Amenable to Optical Contrast Agent Labelling. *Biomark Insights*, 2, 447-461.

Krupa, R., Sobczuk, A., Popławski, T., Wozniak, K., & Blasiak, J. (2011). DNA damage and repair in endometrial cancer in correlation with the hOGG1 and RAD51 genes polymorphism. *Mol Biol Rep*, 38(2), 1163-1170. <https://doi.org/10.1007/s11033-010-0214-z>

Kuznetsova, E. B., Kekeeva, T. V., Larin, S. S., Zemliakova, V. V., Babenko, O. V., Nemtsova, M. V., . . . Strel'nikov, V. V. (2007). [Novel methylation and expression markers associated with breast cancer]. *Mol Biol (Mosk)*, 41(4), 624-633.

Lahtz, C., & Pfeifer, G. P. (2011). Epigenetic changes of DNA repair genes in cancer. *J Mol Cell Biol*, 3(1), 51-58. <https://doi.org/10.1093/jmcb/mjq053>

Lakshminarasimhan, R., & Liang, G. (2016). The Role of DNA Methylation in Cancer. *Adv Exp Med Biol*, 945, 151-172. [https://doi.org/10.1007/978-3-319-43624-1\\_7](https://doi.org/10.1007/978-3-319-43624-1_7)

Lappano, R., Talia, M., Cirillo, F., Rigracciolo, D. C., Scordamaglia, D., Guzzi, R., . . . Maggiolini, M. (2020). The IL1 $\beta$ -IL1R signaling is involved in the stimulatory effects triggered by hypoxia in breast cancer cells and cancer-associated fibroblasts (CAFs). *J Exp Clin Cancer Res*, 39(1), 153. <https://doi.org/10.1186/s13046-020-01667-y>

Lei, L., Bandola-Simon, J., & Roche, P. A. (2018). Ubiquitin-conjugating enzyme E2 D1 (Ube2D1) mediates lysine-independent ubiquitination of the E3 ubiquitin ligase March-I. *J Biol Chem*, 293(11), 3904-3912. <https://doi.org/10.1074/jbc.RA117.001322>

Li, J., Miao, N., Liu, M., Cui, W., Liu, X., Li, X., . . . Biekemituofu, H. (2014). Clinical significance of chromosome 1p/19q loss of heterozygosity and Sox17 expression in oligodendrogliomas. *Int J Clin Exp Pathol*, 7(12), 8609-8615.

Li, J., Ping, J. L., Ma, B., Chen, Y. R., & Li, L. Q. (2017). DIP2C expression in breast cancer and its clinical significance. *Pathol Res Pract*, 213(11), 1394-1399. <https://doi.org/10.1016/j.prp.2017.09.007>

Lin, T. C., Huang, K. W., Liu, C. W., Chang, Y. C., Lin, W. M., Yang, T. Y., & Hsiao, M. (2018). Leptin signaling axis specifically associates with clinical prognosis and is multifunctional in regulating cancer progression. *Oncotarget*, 9(24), 17210-17219. <https://doi.org/10.18632/oncotarget.24966>

Liu, Y., Xu, Z., Jin, T., Xu, K., Liu, M., & Xu, H. (2020). Ferroptosis in Low-Grade Glioma: A New Marker for Diagnosis and Prognosis. *Med Sci Monit*, 26, e921947. <https://doi.org/10.12659/MSM.921947>

Lopez-Gines, C., Cerda-Nicolas, M., Gil-Benso, R., Pellin, A., Lopez-Guerrero, J. A., Callaghan, R., . . . Barbera, J. (2005). Association of chromosome 7, chromosome 10 and EGFR gene amplification in glioblastoma multiforme. *Clin Neuropathol*, *24*(5), 209-218.

Lyng, H., Brøvig, R. S., Svendsrud, D. H., Holm, R., Kaalhus, O., Knutstad, K., . . . Stokke, T. (2006). Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. *BMC Genomics*, *7*, 268. <https://doi.org/10.1186/1471-2164-7-268>

Ma, J., Peng, J., Mo, R., Ma, S., Wang, J., Zang, L., . . . Fan, J. (2015). Ubiquitin E3 ligase UHRF1 regulates p53 ubiquitination and p53-dependent cell apoptosis in clear cell Renal Cell Carcinoma. *Biochem Biophys Res Commun*, *464*(1), 147-153. <https://doi.org/10.1016/j.bbrc.2015.06.104>

Ma, Z., Williams, M., Cheng, Y. Y., & Leung, W. K. (2019). Roles of Methylated DNA Biomarkers in Patients with Colorectal Cancer. *Dis Markers*, *2019*, 2673543. <https://doi.org/10.1155/2019/2673543>

Macintyre, G., Goranova, T. E., De Silva, D., Ennis, D., Piskorz, A. M., Eldridge, M., . . . Brenton, J. D. (2018). Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet*, *50*(9), 1262-1270. <https://doi.org/10.1038/s41588-018-0179-8>

Marotta, M., Chen, X., Inoshita, A., Stephens, R., Budd, G. T., Crowe, J. P., . . . Tanaka, H. (2012). A common copy-number breakpoint of ERBB2 amplification in breast cancer colocalizes with a complex block of segmental duplications. *Breast Cancer Res*, *14*(6), R150. <https://doi.org/10.1186/bcr3362>

Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., . . . Lopez-Bigas, N. (2020). A compendium of mutational cancer driver genes. *Nat Rev Cancer*, *20*(10), 555-572. <https://doi.org/10.1038/s41568-020-0290-x>

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhi, R., & Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*, *12*(4), R41. <https://doi.org/10.1186/gb-2011-12-4-r41>

Michalak, E. M., & Visvader, J. E. (2016). Dysregulation of histone methyltransferases in breast cancer - Opportunities for new targeted therapies? *Mol Oncol*, *10*(10), 1497-1515. <https://doi.org/10.1016/j.molonc.2016.09.003>

Network, C. G. A. R. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, *163*(4), 1011-1025. <https://doi.org/10.1016/j.cell.2015.10.025>

Ono, J., Shime, H., Takaki, H., Takashima, K., Funami, K., Yoshida, S., . . . Seya, T. (2017). The TLR3/TICAM-1 signal constitutively controls spontaneous polyposis through suppression of c-Myc in Apc. *J Biomed Sci*, *24*(1), 79. <https://doi.org/10.1186/s12929-017-0387-z>

Oshi, M., Newman, S., Tokumaru, Y., Yan, L., Matsuyama, R., Endo, I., . . . Takabe, K. (2020). High G2M Pathway Score Pancreatic Cancer is Associated with Worse Survival, Particularly after Margin-Positive (R1 or R2) Resection. *Cancers (Basel)*, *12*(10). <https://doi.org/10.3390/cancers12102871>

Oshi, M., Takahashi, H., Tokumaru, Y., Yan, L., Rashid, O. M., Matsuyama, R., . . . Takabe, K. (2020). G2M Cell Cycle Pathway Score as a Prognostic Biomarker of

Metastasis in Estrogen Receptor (ER)-Positive Breast Cancer. *Int J Mol Sci*, 21(8). <https://doi.org/10.3390/ijms21082921>

Oshiumi, H., Matsumoto, M., Funami, K., Akazawa, T., & Seya, T. (2003). TICAM-1, an adaptor molecule that participates in Toll-like receptor 3-mediated interferon-beta induction. *Nat Immunol*, 4(2), 161-167. <https://doi.org/10.1038/ni886>

Padmavathi, G., Banik, K., Monisha, J., Bordoloi, D., Shabnam, B., Arfuso, F., . . . Kunnumakkara, A. B. (2018). Novel tumor necrosis factor- $\alpha$  induced protein eight (TNFAIP8/TIPE) family: Functions and downstream targets involved in cancer progression. *Cancer Lett*, 432, 260-271. <https://doi.org/10.1016/j.canlet.2018.06.017>

Pan, X., Hu, X., Zhang, Y. H., Chen, L., Zhu, L., Wan, S., . . . Cai, Y. D. (2019). Identification of the copy number variant biomarkers for breast cancer subtypes. *Mol Genet Genomics*, 294(1), 95-110. <https://doi.org/10.1007/s00438-018-1488-4>

Park, S., Blaser, S., Marchal, M. A., Houston, D. W., & Sheets, M. D. (2016). A gradient of maternal Bicaudal-C controls vertebrate embryogenesis via translational repression of mRNAs encoding cell fate regulators. *Development*, 143(5), 864-871. <https://doi.org/10.1242/dev.131359>

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., . . . Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797), 747-752. <https://doi.org/10.1038/35021093>

Podlaha, O., Riester, M., De, S., & Michor, F. (2012). Evolution of the cancer genome. *Trends Genet*, 28(4), 155-163. <https://doi.org/10.1016/j.tig.2012.01.003>

Qi, L., Li, S. H., Si, L. B., Lu, M., & Tian, H. (2014). Expression of THOP1 and its relationship to prognosis in non-small cell lung cancer. *PLoS One*, 9(9), e106665. <https://doi.org/10.1371/journal.pone.0106665>

Sankhwar, M., Sankhwar, S. N., Bansal, S. K., Gupta, G., & Rajender, S. (2016). Polymorphisms in the XPC gene affect urinary bladder cancer risk: a case-control study, meta-analyses and trial sequential analyses. *Sci Rep*, 6, 27018. <https://doi.org/10.1038/srep27018>

Sciarrillo, R., Wojtuszkiewicz, A., El Hassouni, B., Funel, N., Gandellini, P., Lagerweij, T., . . . Giovannetti, E. (2019). Splicing modulation as novel therapeutic strategy against diffuse malignant peritoneal mesothelioma. *EBioMedicine*, 39, 215-225. <https://doi.org/10.1016/j.ebiom.2018.12.025>

Shrotriya, S., Walsh, D., Nowacki, A. S., Lorton, C., Aktas, A., Hullihen, B., . . . Estfan, B. (2018). Serum C-reactive protein is an important and powerful prognostic biomarker in most adult solid tumors. *PLoS One*, 13(8), e0202555. <https://doi.org/10.1371/journal.pone.0202555>

Siena, S., Sartore-Bianchi, A., Marsoni, S., Hurwitz, H. I., McCall, S. J., Penault-Llorca, F., . . . Trusolino, L. (2018). Targeting the human epidermal growth factor receptor 2 (HER2) oncogene in colorectal cancer. *Ann Oncol*, 29(5), 1108-1119. <https://doi.org/10.1093/annonc/mdy100>

Simpson, B. S., Pye, H., & Whitaker, H. C. (2021). The oncological relevance of fragile sites in cancer. *Commun Biol*, 4(1), 567. <https://doi.org/10.1038/s42003-021-02020-5>

- Smith, J. C., & Sheltzer, J. M. (2018). Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *Elife*, 7. <https://doi.org/10.7554/eLife.39217>
- Sobczuk, A., Poplawski, T., & Blasiak, J. (2012). Polymorphisms of DNA repair genes in endometrial cancer. *Pathol Oncol Res*, 18(4), 1015-1020. <https://doi.org/10.1007/s12253-012-9537-5>
- Solimini, N. L., Xu, Q., Mermel, C. H., Liang, A. C., Schlabach, M. R., Luo, J., . . . Elledge, S. J. (2012). Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science*, 337(6090), 104-109. <https://doi.org/10.1126/science.1219580>
- Son, J. W., Jeong, K. J., Jean, W. S., Park, S. Y., Jheon, S., Cho, H. M., . . . Kang, J. (2011). Genome-wide combination profiling of DNA copy number and methylation for deciphering biomarkers in non-small cell lung cancer patients. *Cancer Lett*, 311(1), 29-37. <https://doi.org/10.1016/j.canlet.2011.06.021>
- Song, Q., Merajver, S. D., & Li, J. Z. (2015). Cancer classification in the genomic era: five contemporary problems. *Hum Genomics*, 9, 27. <https://doi.org/10.1186/s40246-015-0049-8>
- Stone, A., Sutherland, R. L., & Musgrove, E. A. (2012). Inhibitors of cell cycle kinases: recent advances and future prospects as cancer therapeutics. *Crit Rev Oncog*, 17(2), 175-198. <https://doi.org/10.1615/critrevoncog.v17.i2.40>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, 71(3), 209-249. <https://doi.org/10.3322/caac.21660>
- Sutton, M. N., Yang, H., Huang, G. Y., Fu, C., Pontikos, M., Wang, Y., . . . Bast, R. C. (2018). RAS-related GTPases DIRAS1 and DIRAS2 induce autophagic cancer cell death and are required for autophagy in murine ovarian cancer cells. *Autophagy*, 14(4), 637-653. <https://doi.org/10.1080/15548627.2018.1427022>
- Takeshima, H., & Ushijima, T. (2019). Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *NPJ Precis Oncol*, 3, 7. <https://doi.org/10.1038/s41698-019-0079-0>
- Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M. P., Vivancos, A., Rovira, A., . . . Lopez-Bigas, N. (2018). Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*, 10(1), 25. <https://doi.org/10.1186/s13073-018-0531-8>
- Taylor, B. S., Barretina, J., Socci, N. D., Decarolis, P., Ladanyi, M., Meyerson, M., . . . Sander, C. (2008). Functional copy-number alterations in cancer. *PLoS One*, 3(9), e3179. <https://doi.org/10.1371/journal.pone.0003179>
- Thomas, L. W., Stephen, J. M., Esposito, C., Hoer, S., Antrobus, R., Ahmed, A., . . . Ashcroft, M. (2019). CHCHD4 confers metabolic vulnerabilities to tumour cells through its control of the mitochondrial respiratory chain. *Cancer Metab*, 7, 2. <https://doi.org/10.1186/s40170-019-0194-y>
- Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*, 19(1A), A68-77. <https://doi.org/10.5114/wo.2014.47136>

Treviño, V., & Tamez-Pena, J. (2017). VALORATE: fast and accurate log-rank test in balanced and unbalanced comparisons of survival curves and cancer genomics. *Bioinformatics*, 33(12), 1900-1901. <https://doi.org/10.1093/bioinformatics/btx080>

Uchiyama, K., Jokitalo, E., Kano, F., Murata, M., Zhang, X., Canas, B., . . . Kondo, H. (2002). VCIP135, a novel essential factor for p97/p47-mediated membrane fusion, is required for Golgi and ER assembly in vivo. *J Cell Biol*, 159(5), 855-866. <https://doi.org/10.1083/jcb.200208112>

Vadakedath, S., & Kandi, V. (2018). Probable Potential Role of Urate Transporter Genes in the Development of Metabolic Disorders. *Cureus*, 10(3), e2382. <https://doi.org/10.7759/cureus.2382>

Vandin, F., Papoutsaki, A., Raphael, B. J., & Upfal, E. (2015). Accurate computation of survival statistics in genome-wide studies. *PLoS Comput Biol*, 11(5), e1004071. <https://doi.org/10.1371/journal.pcbi.1004071>

Vazquez-Mena, O., Medina-Martinez, I., Juárez-Torres, E., Barrón, V., Espinosa, A., Villegas-Sepulveda, N., . . . Berumen, J. (2012). Amplified genes may be overexpressed, unchanged, or downregulated in cervical cancer cell lines. *PLoS One*, 7(3), e32667. <https://doi.org/10.1371/journal.pone.0032667>

Veiseh, O., Kievit, F. M., Ellenbogen, R. G., & Zhang, M. (2011). Cancer cell invasion: treatment and monitoring opportunities in nanomedicine. *Adv Drug Deliv Rev*, 63(8), 582-596. <https://doi.org/10.1016/j.addr.2011.01.010>

Wagner, F., Kunz, T. C., Chowdhury, S. R., Thiede, B., Fraunholz, M., Eger, D., & Kozjak-Pavlovic, V. (2019). Armadillo repeat-containing protein 1 is a dual localization protein associated with mitochondrial intermembrane space bridging complex. *PLoS One*, 14(10), e0218303. <https://doi.org/10.1371/journal.pone.0218303>

Waks, Z., Weissbrod, O., Carmeli, B., Norel, R., Utro, F., & Goldschmidt, Y. (2016). Driver gene classification reveals a substantial overrepresentation of tumor suppressors among very large chromatin-regulating proteins. *Sci Rep*, 6, 38988. <https://doi.org/10.1038/srep38988>

Wang, K., Zhang, D. L., Long, B., An, T., Zhang, J., Zhou, L. Y., . . . Li, P. F. (2015). NFAT4-dependent miR-324-5p regulates mitochondrial morphology and cardiomyocyte cell death by targeting Mtf1. *Cell Death Dis*, 6, e2007. <https://doi.org/10.1038/cddis.2015.348>

Wang, R., Lagakos, S. W., & Gray, R. J. (2010). Testing and interval estimation for two-sample survival comparisons with small sample sizes and unequal censoring. *Biostatistics*, 11(4), 676-692. <https://doi.org/10.1093/biostatistics/kxq021>

Wang, S., Li, H., Song, M., Tao, Z., Wu, T., He, Z., . . . Liu, X. S. (2021). Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. *PLoS Genet*, 17(5), e1009557. <https://doi.org/10.1371/journal.pgen.1009557>

Wang, X., Gao, H., Wu, W., Xie, E., Yu, Y., He, X., . . . Wang, F. (2019). The zinc transporter Slc39a5 controls glucose sensing and insulin secretion in pancreatic  $\beta$ -cells via Sirt1- and Pgc-1 $\alpha$ -mediated regulation of Glut2. *Protein Cell*, 10(6), 436-449. <https://doi.org/10.1007/s13238-018-0580-1>

- Willis, N. A., Rass, E., & Scully, R. (2015). Deciphering the Code of the Cancer Genome: Mechanisms of Chromosome Rearrangement. *Trends Cancer*, 1(4), 217-230. <https://doi.org/10.1016/j.trecan.2015.10.007>
- Xi, X., Li, T., Huang, Y., Sun, J., Zhu, Y., Yang, Y., & Lu, Z. J. (2017). RNA Biomarkers: Frontier of Precision Medicine for Cancer. *Noncoding RNA*, 3(1). <https://doi.org/10.3390/ncrna3010009>
- Yang, Y., Zhu, J., Zhang, T., Liu, J., Li, Y., Zhu, Y., . . . Wu, Q. (2018). PHD-finger domain protein 5A functions as a novel oncoprotein in lung adenocarcinoma. *J Exp Clin Cancer Res*, 37(1), 65. <https://doi.org/10.1186/s13046-018-0736-0>
- Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., . . . Beroukhim, R. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat Genet*, 45(10), 1134-1140. <https://doi.org/10.1038/ng.2760>
- Zhao, J., Niu, X., Wang, Z., Lu, H., Lin, X., & Lu, Q. (2015). Histiocytic sarcoma combined with acute monocytic leukemia: a case report. *Diagn Pathol*, 10, 110. <https://doi.org/10.1186/s13000-015-0350-9>
- Zhao, R., Peng, C., Song, C., Zhao, Q., Rong, J., Wang, H., . . . Xie, Y. (2020). BICC1 as a novel prognostic biomarker in gastric cancer correlating with immune infiltrates. *Int Immunopharmacol*, 87, 106828. <https://doi.org/10.1016/j.intimp.2020.106828>
- Zheng, R., Gao, D., He, T., Zhang, M., Zhang, X., Linghu, E., . . . Guo, M. (2017). Methylation of. *Clin Epigenetics*, 9, 50. <https://doi.org/10.1186/s13148-017-0348-0>
- Zhu, T., Ji, Z., Xu, C., Peng, Z., Gu, L., Zhang, R., & Liu, Y. (2014). Expression and prognostic role of SGTA in human breast carcinoma correlates with tumor cell proliferation. *J Mol Histol*, 45(6), 665-677. <https://doi.org/10.1007/s10735-014-9586-z>
- Zhu, Y. H., Fu, L., Chen, L., Qin, Y. R., Liu, H., Xie, F., . . . Guan, X. Y. (2013). Downregulation of the novel tumor suppressor DIRAS1 predicts poor prognosis in esophageal squamous cell carcinoma. *Cancer Res*, 73(7), 2298-2309. <https://doi.org/10.1158/0008-5472.CAN-12-2663>
- Zito Marino, F., Bianco, R., Accardo, M., Ronchi, A., Cozzolino, I., Morgillo, F., . . . Franco, R. (2019). Molecular heterogeneity in lung cancer: from mechanisms of origin to clinical implications. *Int J Med Sci*, 16(7), 981-989. <https://doi.org/10.7150/ijms.34739>