

BIOMARCADORES DE CÁNCER DE PULMÓN BASADOS EN LA EXPRESIÓN DE GENES, TRANSPONESOS Y SECUENCIAS REPETITIVAS

M. Arroyo Varela¹, R. Larrosa Jiménez², M.A. Cobo Dols³, M.G. Claros Díaz⁴, R. Bautista Moreno⁵.

¹U.G.C. Médico-Quirúrgica de Enfermedades Respiratorias, Hospital Regional Universitario de Málaga.

²Dpto. de Arquitecturas de Computadores, Universidad de Málaga.

³U.G.C. de Oncología, Hospital Regional Universitario de Málaga.

⁴Dpto. de Biología Molecular y Bioquímica, Universidad de Málaga.

⁵Plataforma Andaluza de Bioinformática-SCBI, Universidad de Málaga.

Proyecto financiado con las becas de la Fundación Neumosur 12/2015 y 14/2016.

Palabras clave: ultrasecuenciación, RNA-seq, transposones, elementos móviles, carcinoma epidermoide de pulmón, adenocarcinoma de pulmón.

Keywords: high-throughput sequencing, RNA-seq, transposons, mobile elements, lung cancer, adenocarcinome.

INTRODUCCIÓN

1.1 El cáncer.

Con el término «cáncer» se conoce a un grupo de enfermedades relacionadas en las que se produce una proliferación descontrolada de las células de un organismo. Se trata de una enfermedad genética y compleja, con difícil tratamiento, en la que, a través de una serie de etapas, se transforma progresivamente una célula normal en maligna¹.

Estas células tumorales poseen una serie de características distintivas con respecto a las células normales. Así, *Hanahan y Weinberg*³ establecieron en el año 2000 seis de ellas: el potencial replicativo ilimitado, la autosuficiencia en señales de crecimiento, la insensibilidad a los inhibidores del crecimiento, la evasión de la apoptosis, la angiogénesis mantenida y la capacidad de invadir los tejidos y generar metástasis. Posteriormente, en el año 2011, se revisó su propuesta² y se incorporaron cuatro nuevos distintivos: la capacidad inflamatoria que promueve el tumor, la inestabilidad genómica y mutacional, la capacidad para eludir su destrucción por el sistema inmunitario y, por último, la de crear alteraciones del metabolismo. En su conjunto es lo que se conoce como los *hallmarks* o *distintivos* del cáncer.

Por otro lado y a partir de los datos recogidos en la base de datos GLOBOCAN (*Global Cancer Observatory*)⁴, se ha estimado la incidencia de los distintos tipos de cáncer, encontrándose el cáncer de pulmón y el de mama entre los de mayor incidencia, con un 11,6 % en ambos casos. Sin embargo, si atendemos a la mortalidad, y desde esa misma fuente, se conoce que el cáncer de pulmón es el más mortal, con un 18,4 % del total de muertes, a gran

distancia del cáncer colorrectal, que es el segundo, con un 9,2 %.

A principios del siglo pasado, el cáncer de pulmón era considerado una enfermedad rara. No fue hasta la difusión del hábito tabáquico cuando comenzó a aumentar su incidencia y convertirse en un auténtico problema de salud pública.

El humo del tabaco es un potente carcinógeno, responsable de la gran mayoría de cánceres de pulmón en ambos sexos^{5,6}. Además del tabaco existen otras sustancias carcinogénicas relacionadas con el desarrollo de esta enfermedad, muchas de ellas presentes en exposiciones ocupacionales (amianto, níquel, cromo y arsénico), así como en exposición a radiaciones (gas radón en las minas), o también en exposiciones a la contaminación atmosférica. Por otro lado, los carcinógenos pueden interactuar de manera sinérgica con el tabaco y aumentan el riesgo de desarrollar cáncer de pulmón, como es la sinergia entre el tabaco y la exposición al amianto o al radón⁷. También se ha observado que el estilo de vida puede influir en el desarrollo de esta enfermedad, sobre todo con la falta de ejercicio y el tipo de dieta⁸.

El diagnóstico del cáncer de pulmón suele ser tardío en la mayoría de las ocasiones, dada la escasa sintomatología que origina el tumor a nivel pulmonar. Es por ello que se hace necesario poder disponer de técnicas diagnósticas más certeras y sensibles. En este sentido, la aparición de las llamadas tecnologías Ómicas (*Genómica, Transcriptómica, Proteómica y Metabolómica*) han abierto un amplio campo de investigación, con posibilidades para mejorar tanto el diagnóstico como el tratamiento del cáncer, así como el de otras muchas enfermedades. Es más, a partir de los datos

Recibido: 09.12.2019 Aceptado: 15.04.2020

Dra. Macarena Arroyo Varela
macarrojo@gmail.com

derivados de los estudios del genoma se han podido determinar que ciertas mutaciones tienen una serie de efectos adversos o que incluso pueden ser susceptibles de dianas terapéuticas. Son los denominados marcadores tumorales. Estos conocimientos han llevado a la comunidad científica al desarrollo de las llamadas terapias personalizadas, las cuales han cambiado drásticamente la práctica clínica y han mejorado los resultados obtenidos en la supervivencia de los pacientes.

La mayoría de los marcadores tumorales utilizados a día de hoy pasan por la identificación de mutaciones puntuales, amplificaciones y translocaciones génicas¹². Sin embargo, en muchos pacientes y en muchos tipos de cáncer, estas alteraciones no son suficientes para determinar los tipos histológicos o seleccionar tratamientos dirigidos.

Así, en cáncer de pulmón, a diferencia de otros tumores epiteliales, como el cáncer de mama, donde el uso clínico de marcadores tumorales está muy estudiado y extendido (los marcadores tumorales *BRCA1* y *BRCA2* como indicadores de predisposición^{9,10}; y los marcadores tumorales de los genes receptores estrogénicos o el *HER2/neu* como valor pronóstico y terapéutico¹⁰), no existe ningún marcador tumoral asociado a la detección precoz del cáncer de pulmón con aplicación clínica¹¹⁻¹³.

Por otro lado, muchas de las variaciones provocadas por el desarrollo del cáncer también se producen a nivel de expresión de secuencias genómicas, los ARNs, sin que medie cambios en su secuencia. Esta actividad transcripcional suele depender de estados fisiológicos o patológicos de la célula, lo que demuestra que su estudio en profundidad puede ser una fuente de nuevos conocimientos para entender el desarrollo del cáncer, así como para la identificación de nuevos biomarcadores. Los trabajos más recientes empiezan a indicar que la RNA-seq, el estudio de todos los ARNs que se expresan en una célula, es la técnica que ofrece más posibilidades a la hora de abordar el desarrollo de biomarcadores, por encima del uso de los ya conocidos paneles de genes que se utilizan al estudiar el genoma¹⁴.

1.2 El genoma.

La secuenciación total del genoma humano ha sido uno de los hitos de la historia reciente de la comunidad científica. Como conclusión de este proyecto se publicó en el año 2001 la secuenciación del 90 % del genoma humano¹⁵. Este hecho ha permitido a la comunidad científica dilucidar algunas de las variaciones genómicas relacionadas de forma directa con multitud de enfermedades.

A finales del 2007, con la llegada de las tecnologías

NGS (secuenciación de nueva generación, del inglés *next generation sequencing*) y el consiguiente abaratamiento de costes, se empezaron a generar secuencias de forma masiva, lo que abrió el abanico de los estudios genómicos. Todo esto ha supuesto un salto cuantitativo que ofrece un acercamiento rápido y eficiente a los mecanismos moleculares del cáncer¹⁶.

Gracias a esos avances hoy conocemos que el genoma humano está compuesto por un 2% de secuencias que codifican proteínas, un porcentaje, aún no determinado, de secuencias genómicas reguladoras y de secuencias intergénicas, y un 45% por secuencias repetitivas¹⁵, unas con patrones de repetición simple, y otras más dispersas, con capacidad de movimiento (transposones) Inicialmente se pensaba que todas estas secuencias repetitivas no tenían ninguna función en el organismo que las albergaba, por lo que algunos autores las denominaron «ADN basura»¹⁷. Otros propusieron que existía sólo para mantenerse a sí mismos y lo llamaron «ADN egoísta»¹⁸. Mucho más recientemente se ha propuesto incluso llamarlo «parásitos genómicos»¹⁹. Pero a medida que se ha ido teniendo más información sobre la influencia de las secuencias repetitivas en el entorno genómico, se les va encontrando una función. Así, se ha demostrado que están implicadas en el origen de algunas enfermedades²⁰, entre ellas el cáncer^{21,22}.

1.3 Clasificación de los transposones.

Los transposones se clasifican de tipo ADN o ARN en función del mecanismo de propagación²³⁻²⁵. Mientras que los de ADN utilizan un mecanismo de *cortar y pegar*, los de ARN, llamados retrotransposones, utilizan un ARN intermedio para moverse por el genoma mediante un mecanismo de *copiar y pegar*. Dentro de los retrotransposones podemos encontrar los que tienen un origen retroviral (retrotransposones con LTR), como la familia de retrovirus endógenos (ERV)^{15, 26} que se subdividen en clases en función de la similitud de sus secuencias entre especies, y los que no tienen un origen retroviral, que son los más abundantes. Estos últimos se dividen en dos subtipos basados en su tamaño y estructura: los LINE y los SINE. Los SINE, a su vez, se subclasifican en función de su estructura o secuencia en los Alu, los SVA (elemento repetitivo evolutivamente muy joven formado por fragmentos de otros elementos, en concreto *SINE-VNTR-Alu*²⁷) y los MIR, que parecen proceder de antiguos ARNt²⁸ que mantienen en la actualidad una clara función reguladora²⁹. Sin embargo, la mayoría de los transposones del genoma han perdido su capacidad de salto, menos del 0,05 % retienen esta capacidad³⁰. Además, aunque las células somáticas eucariotas supriman continuamente su

actividad para garantizar la integridad del genoma³¹⁻³³, los transposones siguen proliferando en prácticamente todas las especies conocidas de eucariotas³⁴.

Por todo lo mostrado anteriormente, el trabajo que presentamos en el estudio de Arroyo y colaboradores⁶⁹ se ha centrado en profundizar en el conocimiento de la expresión de secuencias genómicas en el cáncer de pulmón, como fuente de nuevos biomarcadores que pudieran facilitar el diagnóstico y el pronóstico de esta enfermedad. Para ello, nuestro trabajo⁶⁹ estudia la expresión de secuencias genómicas de tejido tumoral y sano en cáncer de pulmón a tres niveles: (I) a nivel de secuencias repetitivas, (II) a nivel de la influencia de la expresión de esas secuencias en su entorno más cercano, y por último (III) a nivel de expresión de genes que codifican proteínas.

BIOMARCADORES DERIVADOS DE SECUENCIAS REPETITIVAS

Como ya se ha comentado, a la hora de identificar nuevos biomarcadores, gran parte del esfuerzo se ha centrado en la parte codificante de proteínas del genoma, debido a la enorme complejidad del estudio de las zonas no codificantes, en especial aquellas que albergan transposones. Esta falta de interés ha hecho que su abordaje en profundidad se convierta en un nuevo foco de conocimiento. El capítulo seis del nuestro trabajo⁶⁹ indicamos que tras comparar la expresión diferencial de las secuencias repetitivas en adenocarcinoma (LAC), cáncer microcítico (SCLC) y cáncer escamocelular (SqCC), se demuestra que no se produce una desregulación global de las secuencias repetitivas, incluidos los transposones, como se había propuesto originalmente³⁵, sino que los cambios están muy controlados, dato que se encuentra en consonancia con lo descrito por Clayton *et al.*³⁶. Es por ello que la activación o represión de ciertos elementos repetitivos se pueden utilizar de biomarcadores.

De hecho, en ese mismo capítulo demostramos que en una cohorte local de LAC la sobreexpresión de los elementos repetitivos HERVL18-int, LTR54B y MER65-int, y la represión del LTR18B podrían servir de distintivo para la filiación de este tipo histológico de tumor^{38,39}. Este estudio esperanzador cuenta con el problema del escaso número de pacientes incorporados, además de estar centrado en un solo tipo histológico de cáncer de pulmón.

El aumento de la cohorte de LAC en el estudio, además de la inclusión de pacientes de SCLC nos ha permitido identificar a 15 transposones específicos de LAC, un número que se multiplica casi por cinco en el SCLC, alcanzado la cifra de 71. De esos resultados se deriva que el elemento HERVL18-int se encuentra sobreexpresado en LAC y en SCLC, aunque con mayor

variabilidad entre los pacientes, corroborándose esta sobreexpresión en los pacientes locales con LAC. Llama poderosamente la atención que sea una familia muy concreta de transposones, la HERV, la que se transcriba de forma significativa en ambos tipos histológicos de cáncer. De estos resultados también se desprende que los transposones AluYg6 y LTR18B se encuentran reprimidos en todos los pacientes de LAC y SCLC, que HERVK11D-Int se encuentra sobreexpresado en SCLC y no en LAC, y que UCON88 permanece sobreexpresado en SCLC y no presenta expresión en LAC³⁹. Dado que además se conoce que muchos de estos transposones aparecen con expresión diferencial en otros cánceres, queda clara su capacidad como biomarcadores diagnósticos específicos. Todo lo anterior indica que los cambios son repetitivos y específicos de la enfermedad (y no espurios ni dependientes del paciente), y que este análisis se puede extender a otros cánceres o a otras enfermedades, lo que abre un nuevo camino para la búsqueda de biomarcadores específicos que no sean solo genes codificantes de proteínas.

INFLUENCIA DEL ENTORNO GENÓMICO

Como se ha comentado anteriormente, el 45% del genoma está formado por secuencias repetitivas, y algunas de ellas se expresan de forma específica en cáncer de pulmón, postulándose algunos de estos transposones como posibles biomarcadores. Sin embargo, la expresión dirigida de algunas copias de estos transposones específicos podría jugar un papel importante en el desarrollo del cáncer de pulmón, influyendo en la expresión de los genes de su entorno, como ya se ha demostrado con algunos elementos repetitivos LTR, los cuales pueden actuar como amplificadores transcripcionales^{40, 41}. Por tanto, cabe preguntarse si la expresión diferencial observada en algunos transposones altera la expresión de los genes de su entorno, o si es la expresión de los genes la que hace que los transposones cercanos cambien su expresión. Encontrar una respuesta no es fácil, por un lado debido a la dificultad de cuantificar la expresión de las secuencias repetitivas, y por otro lado a que determinar quién es la causa y quién el efecto se hace complicado. En este sentido, en el capítulo siete de nuestro trabajo⁶⁹ se desarrolla un método de análisis que hemos denominado NearTrans⁴², con el cual se ha podido corroborar la existencia de una relación mutua entre la expresión de transposones activos (o las secuencias repetitivas de origen transposónico remanentes) y la expresión de los genes adyacentes.

Los análisis mostrados en ese trabajo han permitido identificar 16 copias distintas de transposones en LAC

cuya expresión se correlaciona con los genes de su entorno más cercano, ambos elementos expresados de forma diferencial; este número se eleva a 36 copias distintas en SCLC. Vuelve a ser curioso que la mayoría de los transposones se encuentren reprimidos, tanto en LAC como en SCLC, y que además pertenezcan a la familia de los HERV, cuando ya ha sido descrito el efecto adverso de la sobreexpresión de los HERV en las células cancerosas⁴³. Se sabe que en las células sanas, la expresión de HERV se suele prevenir mediante la metilación del ADN y otros mecanismos de control epigenético. Además, se están acumulando datos de que ciertas proteínas de los HERV tienen propiedades oncogénicas. Sin embargo, también se ha descrito que los ácidos nucleicos derivados de la expresión de los ERV exógenos activan la respuesta inmunitaria innata que lleva a la célula a la apoptosis⁴⁴,⁴⁵, lo que nos sugiere que la represión general de los transposones de tipo HERV podría ser un mecanismo de defensa de la célula tumoral para eludir la respuesta del sistema inmunitario. Por tanto, la importancia de la expresión de determinadas copias en el desarrollo de la neoplasia está aún por determinar.

La principal causa del desarrollo de la neoplasia en SCLC es el tabaco⁴⁶, siendo un 98% de los casos fumadores, lo que hace sospechar que los cambios epigenéticos producidos por el tabaco⁴⁷ en el entorno genómico donde se encuentran posicionadas las copias específicas de transposones podría alterar la expresión de copias concretas. Además, el hecho de que el SCLC tenga su origen en el tabaco podría hacer que se expresaran de forma diferencial más elementos en él que en el LAC, que es un cáncer cuya inducción no está tan relacionada con este hábito⁴⁸.

Nuestros resultados indican que a pesar de las diferencias comentadas, la comparación de los resultados obtenidos en LAC y SCLC arrojan tres parejas transposón-gen comunes a ambos tipos histológicos de cáncer: *AluY-LONRF3*, *L1PA3-AC007743.1*, y *LTR10A-ATF7IP2*, en las que ambos componentes están reprimidos. El que coincidan tan sólo tres parejas hace pensar en comportamientos genómicos muy dispares entre los dos tipos histológicos de cáncer pulmón, algo esperable si tenemos en cuenta la influencia del tabaquismo en SCLC⁴⁸.

Todo lo anterior puede derivar en una relación directa entre la expresión de transposones y genes cercanos, ya que siempre que están a una distancia inferior a 40.000 pares de bases ambos elementos de la pareja se expresan en el mismo sentido. Hay una excepción en el caso de una secuencia centromérica, lo que es difícil de interpretar porque no hay acuerdo sobre el significado de la expresión en los centrómeros debido a la posición especial que ocupan dentro del genoma⁴⁹

. Además, se ha publicado que el centrómero CENP-A está sobreexpresado en un 47,3% de los casos de LAC analizados en trabajo publicado por Wu *et al.*⁵⁰, por lo que se convierte en el primer biomarcador de uso pronóstico en LAC descrito en la bibliografía: cuanto más se expresa, peor es el pronóstico. Seguramente, los estudios de metilación del genoma arrojen luz en el futuro sobre el tema y permitan descubrir nuevos biomarcadores.

De estos resultados se puede proponer que la expresión diferencial de transposones puede servir para descubrir no solo nuevos transposones, si no también genes que tengan relación con el cáncer, e incluso proponer su uso como biomarcadores «universales», o para usar el conocimiento de la relación entre transposones y genes cercanos para establecer dianas terapéuticas, a partir de su influencia en las redes metabólicas. En definitiva, estos resultados apoyan la utilidad de este tipo de estudios para el descubrimiento de posibles biomarcadores específicos así como la incorporación de información sobre la biología del cáncer, y en concreto del cáncer pulmón.

BIOMARCADORES GÉNICOS

Aunque existe muy poca información sobre la importancia de los transposones y otras secuencias repetitivas, tanto a nivel de expresión como de su funcionamiento en el desarrollo del cáncer, los perfiles de expresión génica de muchos tipos de cáncer se conocen con mucho más detalle⁵¹⁻⁵³. Se sabe que no todos los cambios adversos surgen en las secuencias codificantes del gen, sino que muchos de ellos afectan a la secuencia de las moléculas reguladoras. También se sabe que muchos de estos cambios alteran la expresión de uno o varios genes, que se están empezando a utilizar como biomarcadores clínicos en muchos estudios de cáncer⁵⁴. Está descrita la posibilidad de dividir el LAC en subtipos en función de la expresión de ciertos genes, que les conferirían características biológicas, patológicas y clínicas muy diferentes, gracias al estudio de los perfiles de expresión en una cohorte de 509 individuos⁵⁵. En otro estudio más reciente se demuestra que el perfil molecular del cáncer de pulmón de SCLC en el noreste de Brasil difiere del de las poblaciones en otras regiones⁵⁶, lo que nos indica la importancia de la variabilidad poblacional en la identificación de biomarcadores.

Por otro lado, dada la alta variabilidad provocada por la neoplasia, así como la variabilidad debida a la heterogeneidad de la población de células cancerosas, el que se haya conseguido encontrar un perfil de expresión común y estable en secuencias repetitivas para el cáncer de pulmón, como se ha descrito anteriormente, es un resultado esperanzador. En este sentido, en el capítulo

ocho de nuestro trabajo⁶⁹ se han conseguido identificar 9 posibles genes biomarcadores a partir de cohortes pequeñas y locales de LAC, SqCC y SCLC, seis de ellos (*CAPN8-2*, *IYD*, *MSLN*, *MUC1*, *SLC44A4* y *TMC5*) tienen capacidad diagnóstica demostrada in situ en cohorte grande para LAC, tres para SCLC (*ADYCAP1*, *BMX* y *TUBA1A*) y ninguno para SqCC. Llama la atención el mayor número de posibles biomarcadores identificados para LAC, aún teniendo el menor número de pacientes, lo que denota un sesgo genómico, e independiente del ambiente. Al combinar estos resultados con su posible capacidad pronóstica de supervivencia⁵⁷ se ha encontrado una relación entre la expresión de los genes y la supervivencia de los pacientes⁵⁷, lo que demuestra que la sobreexpresión de los genes *TMC5*, *CAPN8-2* y *MUC1* puede utilizarse tanto con valor diagnóstico como pronóstico de LAC, por lo que son unos buenos candidatos para ser trasladados a la práctica clínica después de las respectivas validaciones. En ese trabajo no hemos podido medir el valor pronóstico de los genes específicos de SCLC ya que en las bases de datos públicas no se encuentran almacenados datos derivados de este tipo histológico de cáncer, algo que merma las posibilidades de sacar conclusiones en ese sentido. Todo lo anterior indica que existen biomarcadores que podrían servir no sólo para diagnosticar el cáncer de pulmón, sino también para clasificar la familia histológica concreta, gracias a la elevada diferencia de expresión en algunos de los genes. Pero en vez de intentar adaptar los nuevos métodos para seguir usando una clasificación tradicional, se deberían usar para establecer nuevas clasificaciones, que fueran de un nivel superior a la existente. Además esto permitiría un tratamiento personalizado basado en el tipo del cáncer, y no en el órgano. Aunque esta es una idea antigua⁵⁸, está sin resolver, a pesar de lo mucho que se ha investigado en ello⁵⁹⁻⁶².

En cualquiera de los casos es imprescindible la comprobación de la fiabilidad de estos biomarcadores por otro tipo de técnicas. La validación de estos elementos requiere estudios en poblaciones de mayor tamaño, lo que supone un aumento del coste y la complejidad del proceso si se pretenden ultrasecuenciar dichas muestras. Para sistematizar y simplificar costes y análisis suele recurrirse a la validación por RT-PCR que necesita genes de referencia cuya expresión sea constitutiva. En la bibliografía se pueden hallar muchos genes de referencia específicos de tejido, incluido el pulmón⁶³, pero no puede descartarse que el estado patológico del paciente los altere^{64,65}.

Esta alta variabilidad se ha corroborado en el capítulo nueve de nuestro trabajo^{66,69} cuando se ha intentado identificar genes de referencia de pulmón independientes de la neoplasia a partir de muestras de RNA-seq de LAC y

SCLC. Basándose el nivel de expresión y en el coeficiente de variación de la expresión, un buen gen de referencia será el que tenga un alto nivel de expresión constitutiva entre los tejidos sanos y tumorales de todos los pacientes. En este sentido, hemos sido capaces de identificar para SCLC el gen *RPL14* que codifica una proteína ya utilizada como gen de referencia en tejido normal y tumoral de riñón⁶⁷. Este mismo proceso define también como posible candidato a gen de referencia *ATP5F1A-202*, una proteína mitocondrial no descrita aún en la bibliografía con esta utilidad. Además se ha identificado como otro posible gen de referencia al gen *UBA52*, que codifica las proteínas ribosómicas que ya habían sido identificadas como gen de referencia en cáncer de mama⁶⁸, así como en vejiga o testículo, a través de metanálisis de microarray⁶⁷.

Con respecto al LAC, y en función de nuestros resultados proponemos el gen *HNRNPA1P54*, además de un conjunto de 10 genes ya descritos como de mantenimiento (*housekeeping*), de forma que se sabe que realizan funciones básicas para la célula, y que por tanto su expresión debe permanecer estrechamente regulada.

De todo ello se deduce que la elección de los genes de referencia depende tanto del tejido como de la enfermedad, por lo que conviene disponer de una buena batería de genes de referencia por cada tipo histológico de cáncer de pulmón para poder realizar las validaciones de forma más segura. A cambio, el método propuesto es útil y cómodo para determinar genes de referencia a partir de datos de RNA-seq, con lo que se pueden investigar las bases de datos de expresión de acceso libre para ofrecer distintas listas de genes de referencia para distintas situaciones, tanto de cáncer de pulmón como de cualquier otra enfermedad o situación fisiológica.

CONCLUSIONES

Conseguir el diagnóstico precoz del cáncer de pulmón vendrá dado por el conocimiento genético de esta enfermedad, permitiendo encontrar cambios en mutaciones o expresión de secuencias genómicas que nos lleve a diferenciar, o a tratar, de forma diferente distintos tipos de cáncer. De los estudios derivados de nuestro trabajo⁶⁹ se puede concluir que no existe una desregulación inespecífica de la expresión de los elementos repetitivos cuando las células pulmonares sufren una transformación neoplásica. De la misma forma se ha demostrado que la clase de transposones ERV y las secuencias repetitivas derivadas de ellas parecen cambiar más su expresión que la familia LINE, al menos en cánceres de pulmón. Las secuencias repetitivas HERVL18-int, AluYg6 y LTR18B se postulan como biomarcadores comunes tanto de LAC como de SCLC, el primero de ellos por sobreexpresión y los otros dos por represión. La secuencia repetitiva HERVK11D-Int se sobreexpresa en LAC y no en

SCLC, mientras que UCON88 permanece sobreexpresado en SCLC y no presenta expresión en LAC, evidenciando su capacidad como biomarcadores diagnósticos específicos. A nivel de la influencia en su entorno genómico de la expresión de las secuencias repetitivas se puede concluir que las parejas de transposones y genes cercanos presentan en general un cambio de expresión en el mismo sentido y de la misma magnitud, y ha permitido descubrir nuevos genes que podrían estar implicados en el cáncer de pulmón. Las parejas encontradas en común entre SCLC y LAC han sido *AluY-LONRF3*, *L1PA3-AC007743.1*, y *LTR10A-ATF7IP2*, estando reprimidos en el tejido tumoral en todos los casos. Al focalizar los resultados en la expresión génica, se observa que en el LAC la sobreexpresión de los genes *TMC5*, *CAPN8-2* y *MUC1* puede utilizarse tanto con valor diagnóstico como pronóstico, los genes *ADCYAP1*, *TUBA1A* e *IYD* pueden utilizarse solo como biomarcadores de pronóstico, mientras que los genes *MSLN* y *SLC44A4* podrían usarse como biomarcadores complementarios de diagnóstico. Mientras en SCLC, los genes *ADCYAP1*, *BMX* y *TUBA1A* se proponen con valor diagnóstico.

En definitiva, los biomarcadores propuestos en nuestro trabajo⁶⁹ ha sido base de esta revisión y podrían proponerse para mejorar la clasificación actual del cáncer de pulmón, pasando así de una clasificación histológica a una funcional, que definiera directamente un tratamiento personalizado, y permitiera un mejor conocimiento de la progresión y comportamiento de la enfermedad. Se abre así la puerta a la validación experimental de estos biomarcadores, sobre todo los que tienen valor pronóstico y diagnóstico para LAC. Si se validan no solo genes, sino también elementos repetitivos y transposones, se conseguiría dotar de robustez al diagnóstico para la práctica clínica. Puesto que los resultados mostrados tienen la precaución de escoger los biomarcadores por su comportamiento homogéneo en todos los pacientes (y no en una mayoría), no se descarta que éstos puedan aplicarse a corto plazo a las bronoscopias, e incluso a la biopsia líquida. Al incluir también genes de referencia, no sería necesario obtener muestras de tejido pulmonar sano, ya que se podría hacer una determinación directa basada en los genes de referencia para determinar los niveles de expresión. Con todo esto, no solo se estaría apoyando los diagnósticos anatomopatológicos, sino que se podría tener una idea del pronóstico del paciente. Y quién sabe si también facilitar el diagnóstico precoz del cáncer de pulmón y la implantación de screenings preventivos en las poblaciones de riesgo que reduzcan la mortalidad⁷⁰. En pocas palabras, los biomarcadores permitirían no sólo una más rápida, sencilla y precisa clasificación, sino también el desarrollo de nuevos tratamientos.

BIBLIOGRAFÍA

1. Venables JP. Aberrant and alternative splicing in cancer. *Cancer Res.* 2004; 64(21): 7.647–7.654.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144(5):646–674.
3. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000; 100(1): 57–70.
4. Bray F, Ferlay J, Soerjomataram I et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68(6): 394–424.
5. Walser T, Cui X, Yanagawa J et al. Smoking and lung cancer: the role of inflammation. *Proc Am Thorac Soc.* 2008; 5(8): 811–815.
6. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. International agency for research on cancer iarc monographs on the evaluation of carcinogenic risks to humans. Personal habits and indoor combustions. vol. 100; 2012.
7. Hubaux R, Becker-Santos DD, Enfield KS et al. Arsenic, asbestos and radon: emerging players in lung tumorigenesis. *Environ Health.* 2012; 11(1): 89.
8. Ruiz RB, Hernández PS. Diet and cancer: risk factors and epidemiological evidence. *Maturitas.* 2014; 77(3): 202–208.
9. James CR, Quinn JE, Mullan PB et al. BRCA1, a potential predictive biomarker in the treatment of breast cancer. *The oncologist.* 2007; 12(2): 142–150.
10. Henry NL, Hayes DF. Cancer biomarkers. *Mol Oncol.* 2012; 6(2): 140–146.
11. Gasparri R, Romano R, Sedda G et al. Diagnostic biomarkers for lung cancer prevention. *J Breath Res.* 2018 feb; 12(2): 027111.
12. Seijo LM, Peled N, Ajona D et al. Biomarkers in lung cancer screening: Achievements, promises and challenges. *J Thorac Oncol.* 2018.
13. Garrido P, Conde E, de Castro J et al. Updated guidelines for predictive biomarker testing in advanced non-small-cell lung cancer: a National Consensus of the Spanish Society of Pathology and the Spanish Society of Medical Oncology. *Clin Transl Oncol.* 2019 Oct.
14. Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. *Cell.* 2017 Feb; 168(4): 629–643.
15. Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6.822): 860–921.
16. Lu YQ, Lu KH. Advancements in next-generation sequencing for diagnosis and treatment of non-small-cell lung cancer. *Chronic Dis Transl Med.* 2017; 3(1): 1.
17. Ohno S. So much “junk” DNA in our genome. *Brookhaven Symp Biol.* 1972; 23: 366–70.

18. Dawkins R. *The Selfish Gene*. 30th Anniversary Edition—with a new Introduction by the Author. 1976; p. 1–13.
19. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*. 1997; 13(8): 335–340.
20. Kazazian HH. Mobile elements and disease. *Current Opinion in Genetics and Development*. 1998; 8(3): 343–350.
21. Lee E, Iskow R, Yang L et al. Landscape of Somatic Retrotransposition in Human Cancers. *Science (New York, NY)*. 2012; 967(6.097): 967–71.
22. Hancks DC, Kazazian HH. Active human retrotransposons: Variation and disease. *Current Opinion in Genetics and Development*. 2012;22(3):191–203.
23. Wicker T, Sabot F, Hua-Van A et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007; 8(12): 973.
24. Rebollo R, Romanish MT, Mager DL. Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. *Annu Rev Genet*. 2011; 46(1): 120913153128008.
25. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet*. 2008; 9(5): 411–2; author reply 414.
26. Hayward A. Origin of the retroviruses: when, where, and how? *Current Opinion in Virology*. 2017; 25: 23–27.
27. Hancks DC, Kazazian Jr HH. SVA retrotransposons: evolution and genetic instability. In: *Seminars in cancer biology*. vol. 20. Elsevier; 2010. p. 234–245.
28. Jurka J, Zietkiewicz E, Labuda D. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. *Nucleic Acids Res*. 1995 01; 23(1): 170–175.
29. Thornburg BG, Gotea V, MakaŃowski W. Transposable elements as a significant source of transcription regulating signals. *Gene*. 2006; 365: 104–110.
30. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends Genet*. 2007; 23(4): 183–191.
31. Czech B, Munafò M, Ciabrelli F et al. piRNA-guided genome defense: from biogenesis to silencing. *Annu Rev Genet*. 2018; 52: 131–157.
32. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol*. 2008; 9(1): 22.
33. Bamezai S, Rawat VP, Buske C. Concise review: The Piwi-piRNA axis: pivotal beyond transposon silencing. *Stem Cells*. 2012; 30(12): 2.603–2.611.
34. Blumenstiel JP. Birth, School, Work, Death, and Resurrection: The Life Stages and Dynamics of Transposable Element Proliferation. *Genes*. 2019; 10(5): 336.
35. Shalgi R, Pilpel Y, Oren M. Repression of transposable-elements - a microRNA anti-cancer defense mechanism? *Trends Genet*. 2010; 26(6): 253–259.
36. Clayton EA, Wang L, Rishishwar L et al. Patterns of Transposable Element Expression and Insertion in Cancer. *Frontiers in Molecular Biosciences*. 2016; 3.
37. Kong Y, Rose CM, Cass AA et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nature communications*. 2019; 10(1): 1–14.
38. Arroyo M, Bautista R, Larrosa R et al. Potencial uso biomarcador de los retrotransposones en el adenocarcinoma de pulmón. *Revista Española de Patología Torácica*. 2018; 30(4): 224–230.
39. Arroyo M, Bautista R, Larrosa R et al. Biomarker potential of repetitive-element transcriptome in lung cancer. *PeerJ*. 2019; 7: e8277.
40. Thompson PJ, Macfarlan TS, Lorincz MC. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol Cell*. 2016; 62(5): 766–776.
41. Maruggi G, Porcellini S, Facchini G et al. Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. *Mol Ther*. 2009; 17(5): 851–856.
42. Larrosa R, Arroyo M, Bautista R et al. NearTrans can identify correlated expression changes between retrotransposons and surrounding genes in human cancer. In: Rojas I, Ortuño F, editors. *Lecture Notes in Bioinformatics*. vol. 10.813. IWBBIO18. Springer; 2018. p. 373–382.
43. Bannert N, Hofmann H, Block A et al. HERVs New Role in Cancer: From Accused Perpetrators to Cheerful Protectors. *Front Microbiol*. 2018; 9: 178.
44. Chiappinelli KB, Strissel PL, Desrichard A et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell*. 2017 04; 169(2): 361.
45. Roulois D, Loo Yau H, Singhania R et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell*. 2015 Aug; 162(5): 961–73.
46. Varghese AM, Zakowski MF, Helena AY et al. Small-cell lung cancers in patients who never smoked cigarettes. *J Thorac Oncol*. 2014; 9(6): 892–896.
47. Ambatipudi S, Cuenin C, Hernandez-Vargas H et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics*. 2016; 8(5): 599–618.
48. Gazdar AF, Zhou C. Lung cancer in never-smokers: a different disease. In: *IASLC Thoracic Oncology*. Elsevier; 2018. p. 23–29.
49. Mizuno H, Kawahara Y, Wu J et al. Asymmetric distribution of gene expression in the centromeric region of rice chromosome 5. *Front Plant Sci*. 2011 ;2:16.

50. Wu Q, Qian YM, Zhao XL et al. Expression and prognostic significance of centromere protein A in human lung adenocarcinoma. *Lung Cancer*. 2012; 77(2): 407–414.
51. Li M, Qiu M, Xu Y et al. Differentially expressed protein-coding genes and long noncoding RNA in early-stage lung cancer. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2015.
52. D'Angelo G, Di Rienzo T, Ojetti V. Microarray analysis in gastric cancer: a review. *World journal of gastroenterology : WJG*. 2014 Sep; 20(34): 11.972–11.976.
53. Seo JS, Ju YS, Lee WC et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res*. 2012; 22(11): 2.109–2.119.
54. Martínez-Ledesma E, Verhaak RGW, Treviño V. Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. *Sci Rep*. 2015 Jul; 5: 11.966.
55. Fukui T, Shaykhiev R, Agosto-Perez F et al. Lung adenocarcinoma subtypes based on expression of human airway basal cell genes. *Eur Respir J*. 2013 Nov; 42(5): 1.332–44.
56. Oliveira ACdSMd, Silva AVAd, Alves M et al. Molecular profile of non-small cell lung cancer in northeastern Brazil. *J Bras Pneumol*. 2019 Jun; 45(3): e20180181.
57. Gyórfy B, Surowiak P, Budczies J et al. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One*. 2013; 8(12): e82241.
58. Golub TR, Slonim DK, Tamayo P et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999; 286(5.439): 531–537.
59. Ayyad SM, Saleh AI, Labib LM. Gene expression cancer classification using modified K-Nearest Neighbors technique. *Biosystems*. 2019; 176: 41–51.
60. Xu J, Wu P, Chen Y et al. A Novel Deep Flexible Neural Forest Model for Classification of Cancer Subtypes Based on Gene Expression Data. *IEEE Access*. 2019; 7: 22.086–22.095.
61. Hussain F, Saeed U, Muhammad G et al. Classifying Cancer Patients Based on DNA Sequences Using Machine Learning. *Journal of Medical Imaging and Health Informatics*. 2019; 9(3): 436–443.
62. Tarek S, Elwahab RA, Shoman M. Gene expression based cancer classification. *Egyptian Informatics Journal*. 2017; 18(3): 151 – 159.
63. Warrington JA, Nair A, Mahadevappa M et al. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics*. 2000 Apr; 2(3): 143–147.
64. Rubie C, Kempf K, Hans J et al. Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes*. 2005 Apr; 19(2): 101–9.
65. Kozera B, Rapacz M. Reference genes in real-time PCR. *J Appl Genet*. 2013; 54(4): 391–406.
66. Carmona R, Arroyo M, Jiménez-Quesada MJ et al. Automated identification of reference genes based on RNA-seq data. *Biomed Eng Online*. 2017; 16(1): 65.
67. Cheng WC, Chang CW, Chen CR et al. Identification of reference genes across physiological states for qRT-PCR through microarray meta-analysis. *PLoS One*. 2011; 6(2): e17347.
68. Popovici V, Goldstein DR, Antonov J et al. Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinf*. 2009; 10(1): 42.
69. Arroyo M. Biomarcadores de cáncer de pulmón basados en la expresión de genes, transposones y secuencias repetitivas [Ph.D. thesis]. Universidad de Málaga; 2019. Available from: <https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1807110>.
70. Kanodra NM, Silvestri GA, Tanner NT. Screening and early detection efforts in lung cancer. *Cancer*. 2015; 121(9): 1.347–1.356.