



## POTENCIAL USO BIOMARCADOR DE LOS RETROTRANSPOSONES EN EL ADENOCARCINOMA DE PULMÓN

M. Arroyo Varela<sup>1</sup>, R. Bautista Moreno<sup>2</sup>, R. Larrosa Jiménez<sup>2,3</sup>, J.L. de la Cruz Ríos<sup>1</sup>, M.A. Cobo Dols<sup>4</sup>, M. Gonzalo Claros Díaz<sup>5</sup>.

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

<sup>2</sup>Plataforma Andaluza de Bioinformática-SCBI, Universidad de Málaga.

<sup>3</sup>Dpto. de Arquitecturas de Computadores, Universidad de Málaga.

<sup>4</sup>U.G.C. de Oncología, Hospital Regional Universitario de Málaga.

<sup>5</sup>Dpto. de Biología Molecular y Bioquímica, Universidad de Málaga.

Proyecto financiado con Beca Fundación Neumosur 12/2015.

### Resumen:

**Hipótesis y objetivo:** partiendo de la hipótesis de que la reprogramación de los transposones en el cáncer nos puede orientar sobre su desarrollo, en este trabajo se pretende determinar qué transposones podrían servir de biomarcadores con futuros propósitos diagnósticos y pronósticos.

**Material y métodos:** a partir de muestras congeladas de biopsias de adenocarcinoma de pulmón, se ha secuenciado ARN total del tejido tumoral y sano adyacente de ocho pacientes intervenidos en el Hospital Regional de Málaga. Se han analizado con un flujo de trabajo bioinformático específico que cuantifica y proporciona la expresión diferencial de los transposones cuando se compara el tejido sano y tumoral de cada paciente.

**Resultados:** en este trabajo prospectivo, el nivel de transposición global no cambia entre el pulmón sano y el adenocarcinoma. Se han identificado siete transposones con expresión diferencial significativa: cinco se sobreexpresan en las células del adenocarcinoma y los otros dos se sobreexpresan en las células sanas del pulmón. Todos los que son de la clase de retrovirus endógenos humanos (HERV) tienen un gran potencial como biomarcador al reprogramarse de la misma forma en todos los pacientes.

**Conclusión:** el nivel de transposición en el cáncer de pulmón no está desregulado, sino reprogramado, y los transposones afectados por la reprogramación podrían considerarse biomarcadores en potencia.

**Palabras clave:** ultrasecuenciación, RNA-seq, transposones, elementos móviles, cáncer de pulmón, adenocarcinoma.

### POTENTIAL USE OF RETROTRANSPOSONS AS A BIOMARKER IN LUNG ADENOCARCINOMA

#### Abstract:

**Hypothesis and objective:** transposon reprogramming can be related to cancer development. The aim of this study is to determine the role of transposons in lung cancer and evaluate the possible role of transposon as a diagnostic and prognostic biomarkers in lung cancer.

**Material and methods:** total RNA from lung adenocarcinoma was sequenced. We analyzed RNA from tumor and adjacent healthy tissue from eight patients. By using a specific software the differential expression of the transposons in healthy and tumor tissue was analyzed in each patient.

**Results:** this prospective study shows that in our population, the overall transposition level does not change between the healthy and lung adenocarcinoma tissue. Seven transposons with significant differential expression have been found. Five were upregulated in adenocarcinoma cells, and the other two were upregulated in healthy lung cells. Those that belong to the human endogenous retroviruses (HERV) class show a high biomarker potential since they are reprogrammed in the same way in all patients.

**Conclusion:** the transposition level in lung cancer is not deregulated but reprogrammed. Transposons affected by the reprogramming may be considered as potential biomarkers.

**Keywords:** high-throughput sequencing, RNA-seq, transposons, mobile elements, lung cancer, adenocarcinoma

Recibido: 16.07.2017. Aceptado: 03.08.2018

Dra. Macarena Arroyo Varela  
[macarrojo@uma.es](mailto:macarrojo@uma.es)

## INTRODUCCIÓN

El cáncer es una de las principales causas de muerte en la actualidad a pesar de los avances en el tratamiento quimioterápico, radioterápico y quirúrgico. Entre ellos, el cáncer de pulmón continúa siendo la causa más frecuente de muerte por cáncer en el mundo, con una tasa de supervivencia a los cinco años de aproximadamente el 16%<sup>1</sup>. Por desgracia, muchos pacientes se diagnostican en estadios avanzados de la enfermedad, lo que limita enormemente las opciones terapéuticas<sup>2</sup>. Los pacientes se suelen tratar con primeras y segundas líneas de quimioterapia y únicamente una pequeña proporción de ellos (5 - 7%) son candidatos para tratamientos con dianas terapéuticas. No se ha conseguido eliminar la resistencia a estos tratamientos, por lo que hay que seguir buscando nuevas opciones terapéuticas, a ser posible, basadas en nuevos mecanismos implicados en el desarrollo del cáncer.

Los transposones, también denominados elementos móviles, son secuencias de ADN capaces de moverse de un lado a otro del genoma de manera autosuficiente. Como consecuencia de su actividad, provocan cambios en el genoma, como la alteración de la cantidad de ADN del genoma, la sobreexpresión de un gen cercano, la interrupción de genes o incluso traslocaciones. Se piensa que son la base de la evolución del genoma humano<sup>3, 4</sup>, y se sabe que muchas enfermedades, entre ellas muchos tipos de cáncer, son consecuencia del movimiento de los transposones<sup>5</sup>.

En los humanos, al menos el 45% del genoma lo constituyen derivados de elementos móviles<sup>6, 7</sup> (Figura 1). Un 3 % del genoma humano son transposones de ADN cuya cantidad no aumenta, ya que saltan por un mecanismo de «cortar y pegar». El resto son transposones de ARN que sí intervienen de forma activa en el aumento de tamaño de los genomas al utilizar un mecanismo de «copiar y pegar» con la ayuda de una retrotranscriptasa y un ARN intermedio. Los transposones de ARN, también denominados retrotransposones, pueden ser de origen vírico (atestiguado por la presencia de repeticiones terminales largas denominadas LTR, como la familia de retrovirus endógenos de humano [HERV]<sup>6, 8</sup>), o no (sin secuencias LTR, como son los SINE y los LINE<sup>9</sup>). Los de tipo LTR y los LINE codifican la retrotranscriptasa necesaria para su transposición, mientras que los demás elementos móviles no son autónomos (carecen de esta retrotranscriptasa) y su movimiento depende de los dos anteriores<sup>10</sup>.

Se venía aceptando que la expresión de los retrotransposones debería ser más elevada en los tejidos tumorales, en comparación con el tejido normal<sup>11</sup>, dado que se ha encontrado una elevada expresión de las proteínas codificadas

por los LINE-1 en muchos cánceres<sup>9</sup>. Se ha comprobado que muchos tumores presentan las alteraciones e inestabilidades genómicas típicas de la retrotransposición descontrolada. Estos datos sugieren que los tumores ofrecen un entorno altamente permisivo para la retrotransposición, sobre todo en presencia de determinadas mutaciones «conductoras»<sup>9</sup>. Es pues evidente, que merece la pena estudiar más a fondo qué ocurre con la transposición en el tejido tumoral, por si la transposición pudiera, por ejemplo, explicar la progresión de la enfermedad tras un tratamiento quimioterápico o una intervención quirúrgica.

Sin embargo, en nuestro laboratorio contamos con indicios de que la transposición en el cáncer de pulmón no está en absoluto desregulada, sino simplemente reprogramada, y creemos que esta reprogramación nos puede ayudar a descubrir secuencias transposónicas que podrían servir de biomarcadores. Por eso, este trabajo empezará con la valoración del nivel de transposición global cuando la célula pulmonar pasa a ser cancerosa, y valorar si hay transposones con expresión diferencial entre el tejido pulmonar sano y el adenocarcinoma de pulmón que podrían servir de futuros biomarcadores de la enfermedad.

## MATERIAL Y MÉTODOS

### Selección de muestras y secuenciación

Hemos escogido el adenocarcinoma como «cáncer de pulmón modelo» porque es el cáncer de pulmón más frecuente tanto en la población fumadora como en la no fumadora. Tras la necesaria aprobación por el Comité Ético del Hospital Regional de Málaga, se han solicitado al BioBanco del IBIMA los pacientes intervenidos por videotoracoscopia o toracotomía de adenocarcinoma pulmonar en el Hospital Regional Universitario de Málaga cuyas muestras se hayan conservado a -80 °C en una solución de RNAlater™ (ThermoFischer Scientific) para mantener la integridad del ARN. Los especímenes incluidos en el estudio cumplían las siguientes características: 1) que fueran muestra procedentes de un adenocarcinoma de pulmón estadio IIA-B o IIIA; 2) que la intervención se hubiera realizado entre 2010 y 2011 para conocer su evolución; 3) que el paciente no hubiera recibido ni quimioterapia ni radioterapia antes de la intervención; 4) que el paciente no hubiera tenido nunca otra neoplasia; 5) que hubiera material suficiente de adenocarcinoma y de tejido sano para extraer el ARN de ambos. Se extrajo ARN total del tejido sano y tumoral de cada uno de los pacientes en el Biobanco y se secuenciaron en el NextSeq 550 de Illumina® que la Universidad de Málaga oferta en el Servicio de Ultrasecuenciación

del Centro de Supercomputación y Bioinformática (SCBI), siguiendo el protocolo estándar marcado por el fabricante para un experimento de ultrasecuenciación de ARN (RNA-seq) para obtener lecturas pareadas de  $2 \times 100$  nt.

### Análisis de datos

Las muestras de RNA-seq se han procesado con un flujo de trabajo específico (Arroyo et al, 2018, enviado a publicar) para la limpieza de las lecturas de ultrasecuenciación con nuestro algoritmo *SeqTrimNext*<sup>13</sup>, mapeo sobre la versión hg38 del genoma humano con el mapeador *Bowtie*<sup>14</sup>, cuantificación de la expresión de los transposones basada en el algoritmo de *RepEnrich*<sup>12</sup> y la ayuda de las herramientas de *SAMtools*<sup>15</sup>, y el cálculo de la expresión global, así como la diferencial, de los transposones con la librería de *R edgeR*<sup>16</sup>. Tras filtrar los resultados para retirar del análisis los transposones que se expresaban poco (menos de 10 cuentas por millón en al menos 2/3 de las muestras), quedaron finalmente 1267 transposones que aparecieron en 4 713 583 posiciones del genoma. En la expresión diferencial se tuvo en cuenta que los valores de expresión estaban emparejados (tejido sano y tumoral de cada paciente). Se consideró que el cambio de expresión de un transposón era significativo cuando cumplía dos condiciones: que el cambio de expresión al pasar de normal a tumoral fuera de al menos el doble, y que superara el filtrado de falsos negativos para las pruebas múltiples basado en el FDR (*false discovery rate*).

## RESULTADOS

### Las lecturas fueron de gran calidad

Los criterios de selección de las biopsias almacenadas en el Biobanco fueron muy rigurosos y había muy pocos casos que cumplieran todas las condiciones. Finalmente, se seleccionaron 8 muestras totalmente anonimizadas, cuyos detalles sobre el origen, tratamiento y evolución se detallan en la Tabla 1. De cada biopsia se obtuvieron dos muestras, una de la parte sana del pulmón (N en la Tabla 2) y otra de la parte diagnosticada como adenocarcinoma (T en la Tabla 2). El número de lecturas brutas obtenidas de la ultrasecuenciación del ARN por cada muestra, así como el número de lecturas que resultaron útiles para el estudio se ofrecen en la Tabla 2. Conviene observar que el porcentaje de lecturas útiles oscila entre el 79 % y el 87 %, lo que indica que la secuenciación fue de gran calidad y que el ARN a secuenciar estaba íntegro.

Por otra parte, el porcentaje de lecturas que se alinean con el genoma humano de referencia estuvo entre el 86,50 % y el 97,09 % (Tabla 2), lo que también demostró que las lecturas útiles son de humano y se podrán utilizar para los análisis de expresión. Solo existe una muestra con una calidad inferior a las demás, la ID 2 tipo T, en la que tanto el número de lecturas útiles, como el porcentaje de lectura mapeadas fue muy inferior a la media de las muestras. Por tanto, junto con su pareja, se ha descartado para el estudio (ID 2 y tipo N).

### La transposición sigue regulada en el adenocarcinoma

Se ha calculado el cambio de expresión global en el tejido sano con respecto al tumoral en cada uno de los pacientes de la Tabla 1. Como puede verse en la Figura 2, la expresión diferencial normalizada ( $\log_2 FC$ ) oscila de forma casi simétrica entre 0,539 y  $-0,573$ , con la media en 0,043 y la mediana en  $-0,013$ . La presencia de puntos individualizados en el diagrama de cajas presentado en la Figura 2 sugiere que hay transposones que podrían haber sufrido una reprogramación específica de su expresión, bien por activación o bien por represión en el adenocarcinoma de pulmón. Por tanto, conviene explorar si existe aleatoriedad o no entre los transposones que presentan una expresión diferencial significativa entre las células de pulmón sano y el tumoral.

### Hay 7 transposones con expresión diferencial

El análisis estadístico de la expresión diferencial, cuyo resultado se representa en la Figura 3, muestra que la gran mayoría de los transposones no cambia significativamente su expresión entre el pulmón sano y el adenocarcinoma. En cambio, aparecen seis elementos móviles (ALR\_Alpha, HERVL18-int, LTR18B, LTR54B, MER136 y MER65-int) claramente diferentes que muestran un cambio de expresión estadísticamente significativo entre el tejido tumoral (adenocarcinoma de pulmón) y el pulmón sano en los siete pacientes analizados. MSR1 también presenta expresión diferencial, pero su comportamiento está muy cerca de los que se han considerado invariables.

Por su naturaleza, cuatro transposones (MER65-int, LTR18B, HERVL18-int, LTR54B) son de tipo HERV, mientras que dos (ALR\_Alpha y MSR1) son de tipo satélite, y el último es un transposón de ADN (MER136). Con respecto a su comportamiento, cinco se expresan con claridad en el adenocarcinoma (y están reprimidos en el pulmón normal), mientras que dos (MSR1 y LTR18B) se reprimen específicamente en el adenocarcinoma, a pesar de que se estaban expresando en el tejido

normal. Por tanto, la reprogramación es totalmente específica e incluye tanto activación como represión de transposones.

### Cuatro transposones presentan potencial biomarcador

Para poder utilizar estos transposones como biomarcadores diagnósticos o predictivos, es indispensable que su comportamiento sea constante en todos los pacientes. Para comprobarlo, analizamos el cambio de expresión de cada uno de los 7 transposones con expresión diferencial en el conjunto de pacientes con la idea de seleccionar los que no incluyeran el valor  $\log_2 FC = 0$  entre la caja y los «bigotes» en una representación de tipo *boxplot* como la de la Figura 4. Los que incluían el cero no se podían considerar como buenos biomarcadores, puesto que la probabilidad de que un paciente tuviera este transposón con un cambio

de expresión de signo opuesto al esperado sería demasiado alta para una aplicación clínica. Por este motivo, se descartaron ALR\_Alpha y MER136 y MSR1.

El resto de los transposones (entre los que están todos los HERV) sí cumplían las características para ser un potencial biomarcador (Figura 4): la caja y los bigotes de HERVL18-int, LTR54B y MER65-int no cruzaron el valor cero, por lo que se sobreexpresaron en todos los adenocarcinomas analizados. Por su parte, la caja y los bigotes de LTR18B tampoco cruzo el cero, estando reprimido en todos los adenocarcinomas de pulmón analizados. Se propone, por tanto, que la sobreexpresión de HERVL18-int, LTR54B y MER65-int, junto a la represión de LTR18B serían una firma típica del adenocarcinoma de pulmón en la cohorte de pacientes del Hospital Regional de Málaga.

**Tabla 1:** Información sobre las biopsias del Biobanco seleccionadas para este estudio

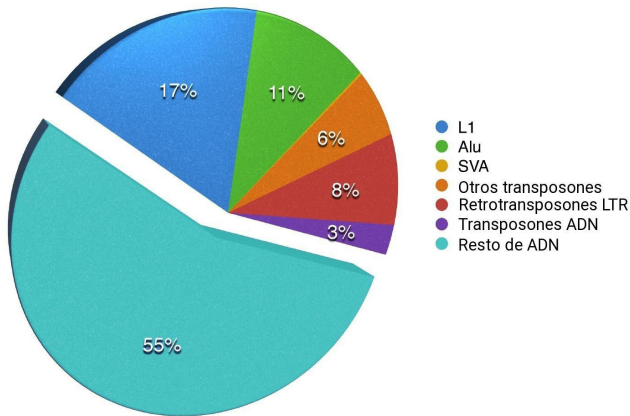
ID	Edad	Sexo	Estadio	Sup.	Abordaje quirúrgico	Tipo de resección	Derivación a Oncología
1	58	Hombre	IIIA	Sí	TPD	LSD	Sí
2	70	Hombre	IIA	Sí	TPD	LM	Sí
3	65	Mujer	IIA	No	VDT	LSI	Sí
4	71	Hombre	IIIA	No	TPI	LII	Sí
5	70	Hombre	IIA	No	VDT	LSD	Sí
6	64	Hombre	IIIA	No	TPD	LM	Sí
7	63	Hombre	IIIA	No	TPA	LID	Sí
8	58	Hombre	IIA	No	TPI	LII	Sí

Abreviaturas usadas: ID: identificador de la biopsia. Sup: supervivencia a la escritura del artículo. VDT: videotoracosopia. TPI: toracotomía posterolateral izquierda. TPD: toracotomía posterolateral derecha. TPA: toracotomía posterior amiotónica. LSI: lobectomía superior izquierda. LII: lobectomía inferior izquierda. LM: lobectomía media. LID: lobectomía inferior derecha. LSD: lobectomía superior derecha.

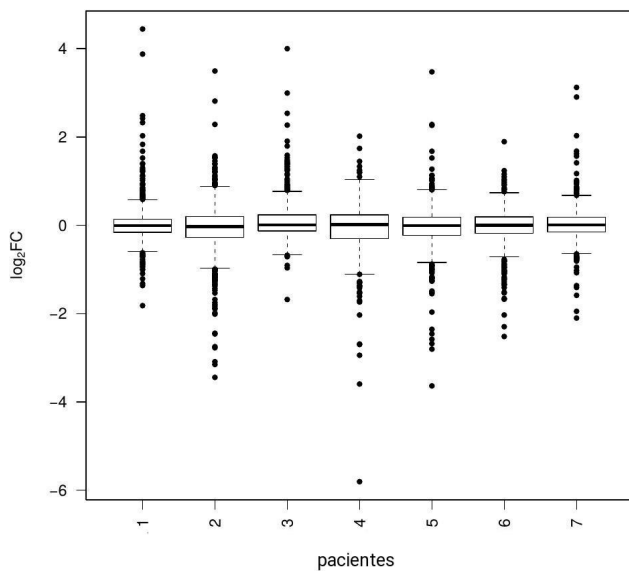
ID	Tipo	Lect. brutas	Lect. útiles	Mapeadas (%)
1	T	222386780	210722264	96,21
1	N	131717912	123396996	97,01
2	T	66757802	63134335	86,50
2	N	108876982	101760493	97,00
3	T	109767878	103743903	97,02
3	N	101584630	96177771	95,54
4	N	125835936	117685290	96,57
4	T	109997648	102890314	95,28
5	T	113174710	106143254	96,54
5	N	111701644	103993702	96,10
6	T	124076878	116085011	96,70

6	N	107635734	100443840	95,96
7	N	93663614	89151278	96,34
7	T	121221114	110742683	95,01
8	T	93690798	86191840	97,09
8	N	103100030	97261908	96,82

**Figura 1:** Composición del genoma humano. Los transposones forman un 45 % del genoma, de los que el 42 % son retrotransposones (L1+Alu+SVA+Otros+LTR) y el resto (3 %), transposones de ADN.

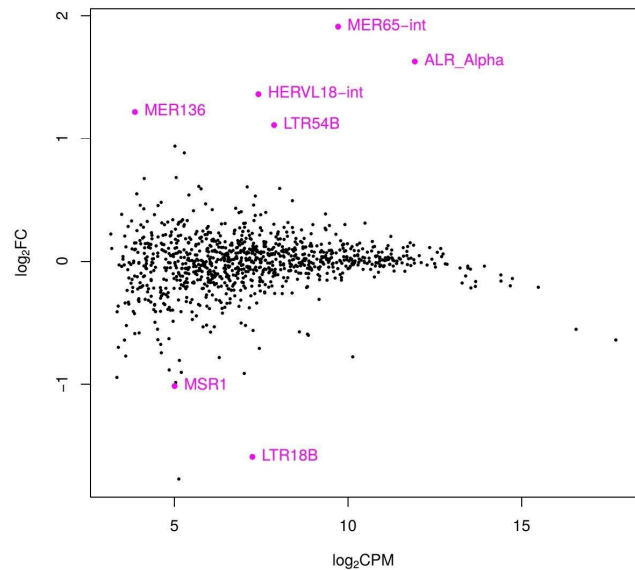


**Figura 2:** Distribución del cambio de expresión de la transposición por paciente. Se ha calculado la expresión diferencial normalizada (eje de ordenadas) para los 1267 transposones entre el tejido tumoral y normal, agrupados por paciente (eje de abscisas). log<sub>2</sub>FC: logaritmo en base 2 del cambio de expresión normalizado entre tejido tumoral y normal; los valores positivos indican expresión en tejido tumoral y represión en el tejido normal, mientras que los valores negativos indican represión en el tejido tumoral y expresión en el tejido normal.

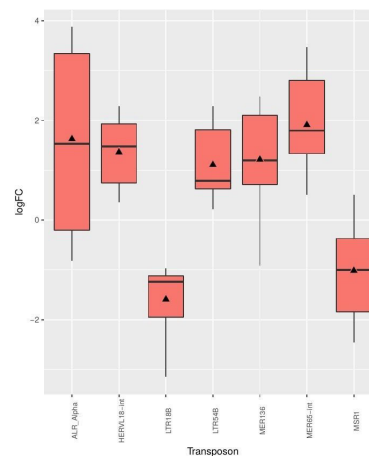


**Figura 3:** Distribución del cambio expresión de los 1267 transposones. Cada punto corresponde al análisis estadístico de la expresión diferencial de cada transposón entre los 7 pacientes cuando se compara en el mismo paciente su tejido sano y el tumoral. En el eje de abscisas se representa la intensidad de la expresión (log<sub>2</sub>CPM: logaritmo en base 2 del valor de expresión normalizada del transposón), de manera

que cuanto más a la derecha aparece el punto, más se expresa el transposón. En el eje de ordenadas se representa el cambio de expresión (log<sub>2</sub>FC: logaritmo en base 2 del cambio de expresión normalizada entre tejido tumoral y normal), por lo que cuanto más hacia arriba o hacia abajo, mayor es el cambio de expresión. Se destacan en otro color y con su nombre los transposones con cuyo cambio de expresión se considera estadísticamente significativo, o sea, los transposones con expresión diferencial entre tejido de pulmón tumoral y sano.



**Figura 4:** Distribución del cambio de expresión entre los pacientes presentada por los transposones con expresión diferencial de la figura 3. El cambio de expresión normalizada (log<sub>2</sub>FC, eje de ordenadas) de cada transposón (eje de abscisas) se ha representado en forma de gráfico de cajas y bigotes para detectar cuáles son los que aparecen en algún paciente con una expresión en sentido opuesto al esperado. El triángulo superpuesto representa la expresión diferencial promedio representada en la Figura 3. Dentro de la caja se encuentra el 50% de las muestras, y si se incluyen los bigotes, quedan incluidas el 99% de las muestras. Que no haya ningún punto fuera de los bigotes significa que no hay ninguna muestra discordante.



## DISCUSIÓN

El reducido tamaño de la cohorte (8 pacientes) hace que los resultados obtenidos tengan una validez principalmente prospectiva. Afortunadamente, la gran calidad de las muestras, reflejado en el alto porcentaje de lecturas útiles tras el procesamiento y el alto porcentaje de mapeo (Tabla 2), indica que el ARN total de partida aislado de las muestras por el Biobanco conservaba su integridad. Esto no solo confirma que las condiciones de conservación de muestras del Biobanco de Andalucía son apropiadas para este tipo de estudios, sino que las muestras aportan un número de lecturas por muestra adecuado para el experimento, a pesar de que se hayan tenido que descartar las de ID 2 (Tabla 1).

El hecho de que la mediana de la transposición global esté en  $-0,013$  (valor que no se distingue estadísticamente de 0) en la Figura 2 indica que los transposones no se desregulan en el adenocarcinoma con respecto al pulmón normal. Por tanto, estos resultados no respaldan el modelo de Shalgi y colaboradores<sup>11</sup>, sino que queda claramente demostrado que la regulación de los transposones no se pierde de forma global cuando la célula pulmonar se convierte en cancerosa. Es más, los datos presentados apoyan la hipótesis de que la célula tumoral sigue manteniendo bajo un estrecho control a los transposones, tal y como también aparecen en una reciente publicación<sup>17</sup> donde se estudian cómo se comportan los transposones en diferentes tipos de cáncer.

En las Figuras 3 y 4 se ve claramente que seis transposones (ALR\_Alpha, HERVL18-int, LTR18B, LTR54B, MER136 y MER65-int) tienen una expresión diferencial claramente distinta a todos los demás. Lo que llama la atención es que haya cuatro transposones de tipo HERV (claramente ligados a otros tipos de cáncer<sup>18</sup>), pero que no aparezca ningún LINE1, que es lo que se ha venido describiendo para otros muchos cánceres<sup>12, 17</sup>. Esto parece sugerir que la expresión diferencial de los transposones podría ser dependiente del tipo de tumor, y que no solo los LINE provocan cambios durante la conversión neoplásica de las células. Este comportamiento específico es el que sugiere fuertemente que las secuencias de los transposones podrían servir de biomarcadores diagnósticos y, quizá, pronósticos (suponiendo que cuantos más transposones se reprogramen, más probable es que la célula sufra reorganizaciones que redunden en la aparición de resistencias<sup>11</sup>).

De los 7 transposones con expresión diferencial, solo 4 parecen tener un potencial claro como biomarcadores de adenocarcinoma (Figura 4). La firma consistiría en la sobreexpresión

de HERVL18-int, LTR54B y MER65-int, y la represión de LTR18B. Si a esto añadimos que se acaba de describir que la represión del HERV-K en el cáncer de páncreas hace disminuir el crecimiento celular y tumoral<sup>18</sup>, podemos concluir que estos cuatro HERV con expresión diferencial tienen potencial biomarcador claro para el adenocarcinoma de pulmón e incluso para la evolución de la enfermedad. Resulta esperanzador que LTR18B y HERVL18-int también tengan potencial biomarcador en otra cohorte de 50 pacientes con adenocarcinoma de pulmón obtenidos de las bases de datos públicas (Arroyo et al, enviado para publicación). Esto supone un gran hallazgo que abriría la puerta al diseño de biomarcadores específicos de adenocarcinoma de pulmón no solo con genes y ARN no codificantes como hasta ahora, sino también con secuencias transposónicas, como se ha empezado a sugerir recientemente<sup>19</sup>.

## AGRADECIMIENTOS

Quisiéramos agradecer a la Plataforma Bioinformática de Andalucía y al Servicio de Ultrasecuenciación, ambos de la Universidad de Málaga, el apoyo que han prestado para el desarrollo de este proyecto. Un agradecimiento especial a Neumosur por su apoyo, ya que, sin la beca 12/2015, esta investigación no se podría haber desarrollado.

## BIBLIOGRAFÍA

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016; 66(1): 7–30.
2. Stewart BW, Wild CP, editors. *World Cancer Report 2014*. Lyon, France: IARC Publications; 2014.
3. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. *Science.* 2016; 351(6274): aac7247.
4. Vogt J, Bengesser K, Claes KBM et al. SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome biology.* 2014; 15(6): R80.
5. Hancks DC, Kazazian HH. Active human retrotransposons: Variation and disease. *Curr Opin Genet Dev.* 2012; 22(3): 191-203.
6. Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409(6822): 860–921.

7. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nature reviews Genetics*. 2009; 10(10): 691–703.
8. Hayward A. Origin of the retroviruses: when, where, and how? *Current Opinion in Virology*. 2017; 25: 23–27.
9. Sciamanna I, Gualtieri A, Piazza PV et al. Regulatory roles of LINE-1-encoded reverse transcriptase in cancer onset and progression. *Oncotarget*. 2014; 5(18): 8039–8051.
10. Sciamanna I, Landriscina M, Pittoggi C et al. Inhibition of endogenous reverse transcriptase antagonizes human tumor growth. *Oncogene*. 2005; 24(24): 3923–31.
11. Shalgi R, Pilpel Y, Oren M. Repression of transposable elements - a microRNA anti-cancer defense mechanism? *Trends Genet*. 2010; 26(6): 253–9.
12. Criscione SW, Zhang Y, Thompson W et al. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC genomics*. 2014; 15: 583
13. Falgueras J, Lara AJ, Fernandez-Pozo N et al. SeqTrim: a high-throughput pipeline for preprocessing any type of sequence reads. *BMC Bioinformatics*. 2010; 11(1): 38.
14. Langmead B, Trapnell C, Pop M and Salzberg SL. Ultrafast and memory- efficient alignment of short ADN sequences to the human genome. *Genome Biology*. 2009; 10(3): R25.
15. Li H, Handsaker B, Wysoker A et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16): 2078–2079.
16. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1): 139–40.
17. Clayton EA, Wang L, Rishishwar L et al. Patterns of Transposable Element Expression and Insertion in Cancer. *Frontiers in Molecular Biosciences*. 2016; 3: 76.
18. Li M, Radvanyi LG, Yin B et al. Down-regulation of human endogenous retrovirus type K (HERV-K) viral env RNA in pancreatic cancer cells decreases cell proliferation and tumor growth. *Clinical Cancer Research*. 2017; en prensa (DOI: 10.1158/1078-0432.CCR-17-0001).
19. Apostolou P, Toloudi M, Chatziioannou M et al. Involvement of retrotransposon L1 in stemness and cellular plasticity. *Cell Commun Adhes*. 2015; 22(1): 1–7.